Essays on Industrial Organization and Public Economics

By

Igor Cerasa

B.S., Università Commerciale Luigi Bocconi, 2013
M.S., Università Commerciale Luigi Bocconi, 2016
M.A., Brown University, 2018

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in the Department of Economics at Brown University

Providence, Rhode Island
May, 2022

This dissertation by Igor Cerasa is accepted in its present form
by the Department of Economics as satisfying the
dissertation requirements for the degree of Doctor of Philosophy.

Date _____ _____

Emily Oster, Advisor

Recommended to the Graduate Council

Date _____ _____

John Friedman, Reader

Date _____ _____

Jesse M. Shapiro, Reader

Approved by the Graduate Council

Date _____ _____

Andrew G. Campbell, Dean of the Graduate School

Curriculum Vitae

Igor Cerasa is originally from Italy. In 2013, he received his Bachelor of Science, cum laude, in International Economics, Management, and Finance, from Bocconi University in Milan, Italy.

In 2016, he earned a Master of Science in Economic and Social Sciences, cum laude, from the same university. At Bocconi University, he was recipient of the Bocconi Graduate Merit Award, and Visiting Student at the Innocenzo Gasparini Institute for Economics Research.

He began his doctoral studies in Economics at Brown University in 2017, and received a Master of Art in Economics in 2018. During his stay at Brown University he was recipient of the Ermenegildo Zegna Founder's Scholarship, and pre-doctoral trainee at the Population Studies and Training Center.

Preface

This dissertation is composed of two self contained chapters, exploring topics in Industrial Organization and Public Economics.

The first chapter, a joint project with Tommaso Coen, investigates the role of pre-purchase information in markets with experience goods, products whose quality is unknown ex-ante. In particular, we are interested in how sellers can influence the perceived quality by strategically censoring or disclosing information. We study this issue in the context of online wine retail, where expert and crowd based reviews are very common. We created a unique dataset of expert ratings combining both an independent source and a major online retailer. Our results document that the seller is likely to censor the lowest scores. To understand if censorship can have an impact on comsumers' welfare, we turn our attention to the demand side of the market, and study whether consumers are aware of these strategic interactions. We design and implement on online discrete choice experiment, combined with an information treatment, to estimate the impact on expert scores on consumers' choices, and test whether they are aware of censorship. Our results show that consumers care about expert ratings, as wines with higher scores are more likely to be selected than comparable wines without. The information treatment did not stimulate a change in consumers behavior with respect to expert ratings. We propose three alternative explanations: (1) that consumers are naive, in the sense that they always act as if they observe ratings from a complete source, (2) that consumers are sophisticated and are not surprised to learn about sellers' behavior, or (3) that participants to the experiment

did not read or understand our information treatments, or that they did not care about it.

In the second chapter, which is individual work, I study the political and economics determinants of local budget formation, more specifically the role played by traffic fines. Using data from Italian municipalities, I provide evidence of a revenue and an electoral motive for municipal governments to impose tickets. I exploit an exceptional fiscal consolidation that happened in Italy, where the central government reduced transfers to municipalities above the threshold of 5,000 inhabitants, but not on those below. A Regression Discontinuity approach shows that municipalities affected by the policies reacted by increasing revenues from traffic fines, to cover roughly 10 percent of their budget reduction. Interestingly, no beneficial effect was observed when looking at road accidents. Using variation across municipalities in local election timing, I also provide evidence that traffic is enforced relatively less strictly in electoral years, inducing an increase in road accidents and injured people.

Acknowledgments

I would like to express my sincerest gratitude to my advisors, Emily Oster and Jesse Shapiro, for their valuable guidance and precious suggestions. I would also like to thank John Friedman for his insightful comments, and Angelica Spertini for her help and passionate dedication.

I am extremely grateful to Marco, Nicola, Valeria, and Tommaso for their generous support in completing this dissertation.

I would like to thank my classmates and colleagues for making this long and challenging journey much more enjoyable. A special thanks to Chien-Tzu, Cosimo, and Juan Pedro for being always present whenever I needed them. I deeply thank Viola, for everything she has done for me in the past five years.

The most heartfelt thanks goes to my parents, my sister, and my family for their patience, understanding, and their constant loving support throughout these many years spent abroad, and away from them.

Finally, I want to dedicate this work to *nonno* Tonino, hoping he is proud of me.

Contents

List of Tables

List of Figures

Chapter 1

Censorship and Experience Goods. Evidence from Online Wine Retail

1.1 Introduction

The[1] share of consumption devoted to experience goods has markedly increased over time (Goldman, Marchessou and Teichner, 2017). Experience goods are products and services whose quality is unknown before consumption, and include books, movies, restaurants, wines and other such goods. When making a purchase decision, individuals often rely on guidance from pre-purchase information, which conveys a signal of quality (Reimers and Waldfogel, 2021). Given the relevance that quality information has in the context of experience goods, it is critical to understand what signal is delivered to consumers, and what purpose that signal serves. In particular, we are interested in how signals of quality are communicated by sellers, who typically have deeper knowledge of a market than consumers. For example, sellers may find it profitable to manipulate quality signals, censor them completely, or even find it optimal to disclose full information. A strictly related question is whether consumers are aware of the strategic incentives faced by sellers. If consumers consider the seller's strategy when making a purchase decision, the effect of strategic incentives on consumer welfare is ex-ante ambiguous.

We therefore ask the following questions. First, do sellers censor quality information presented to consumers? Second, what are consumers' attitudes towards pre-purchase

[1]This Chapter is joint work with Tommaso Coen (*Brown University*).

information when it is directly provided by sellers versus independent sources? Are consumers aware of sellers' strategic incentives or are they acting naively? Finally, what are the welfare implications for consumers when pre-purchase information is censored?

This paper addresses these questions by studying the impact of expert scores in the online retail market for wine. The online marketplace for wine has several interesting features that make it particularly suitable for investigating our questions. First, wine is a clear example of an experience good, and quality information is especially desirable for consumers. Second, the wine market has a developed system of independent and reliable expert opinions that are published in specialized magazines and journals which pre-dates the internet era. Over the past several decades, digitization has made it even easier to collect and disseminate independent evaluations from diverse sources (Moreno-Ternero, 2020). In addition to expert ratings, crowd-based reviews have emerged as a popular and powerful complement to traditional expert scores. Third, sellers in this market are likely to have better access to information than are buyers, increasing the likelihood that they engage in strategic behavior (Miller, Genc and Driscoll, 2007).[2] Finally, we collect data from both a seller and an independent aggregator of expert ratings, allowing us to exploit the different incentives faced by parties when deciding how to disclose scores.[3] We are therefore in a position to precisely describe the seller's behavior and document the extent of information manipulation.

Our analysis proceeds in two stages. First, we investigate whether there is strategic selection of product reviews by retailers. We construct a novel dataset that includes expert

[2]Under stringent conditions, markets with asymmetric information can reach the so-called *unraveling* equilibrium (Grossman, 1981), where every available information is reported, except that pertaining to the lowest type. For this to happen, stringent assumptions need to be met: namely that consumers are aware that the information exists (Hirshleifer, Lim and Teoh, 2004), are able to understand it (Fishman and Hagerty, 2003), and can think strategically about the consequences of censorship (Brown, Camerer and Lovallo, 2012). We suspect this does not apply to the market we study in this project, where we are therefore likely to observe partial censorship, and the potential for policy interventions, such as mandatory or third party disclosure.

[3]We make the assumption, that we deem sensible for our analysis, that the seller provides quality information with the goal of maximizing profits from wine sales. On the other hand, we assume that for the independent website the goal itself is to maximize the quality of the information provided.

wine reviews from two different sources, covering more than 100,000 distinct wines. The first source is a leading wine retailer's online store, and the other is the comprehensive database of a leading wine search engine. These data allow us to compare the distribution of reviews across websites and test whether the retailer censors reviews that assign lower scores to a wine. Second, we consider whether censorship by the retailer can impact consumers' choices. At the same time, we address important data limitations. For example, ratings are likely to be endogenous since a seller can set the price of a wine based on its rating. An even larger concern is that it is impossible to observe in the real world score values that have been censored, making it difficult to access detailed and unbiased data on consumers' behavior from the field.

We tackle this issue by designing and implementing an online random choice experiment to collect data on consumer behavior. In the experiment, we ask consumers to choose wines from a virtual shelf, showing them randomized information about the prices, attributes, expert scores and crowd-based ratings of each wine.[4] To evaluate how censorship impacts consumer choices, we randomly assign consumers to control and treatment groups. Individuals receive an information treatment[5] that varies based on their treatment group assignment. We define two main information provisions: in the first case we warn respondents that the displayed scores are manipulated by the seller, while in the second case we inform respondents that expert ratings come from an impartial and complete source, in addition to providing them a warning that sellers can engage in score manipulation. Individuals in the control group receive no additional information, in an attempt to mimic a real-world environment. Respondents are also randomized with re-

[4]In randomizing expert scores, we selected both the number and the values of ratings available for that wine. This includes the possibility that the wine is unrated. A deep investigation on why, in the wine market, some wines have scores and others don't is beyond the scope of this paper. We identify two possible reasons: either the wine was never rated by an expert, or it was evaluated, but the score was not released (e.g. some experts do not disclose scores below 80 points). We find it plausible that the former applies to most wines in the market, given the disproportion between the number of rating sources and the amount of wine being produced.

[5]See Haaland, Roth and Wohlfart (2022) for a comprehensive review of how this methodology has been applied in the field of economics.

spect to the expert scores they observe in the discrete choice experiment. While some respondents observe scores from a complete score distribution, others are shown ratings from a censored distribution, akin to what happens in a real store. Comparing behavior across treatment groups allows us to investigate whether consumers are aware of the seller's strategic incentives and if they react, and to evaluate the possible consequences for consumer welfare.

We obtain three sets of results. First, we document that the seller engages in extensive censorship of the lowest scores, totally omitting values of 85 points or less on a 50 to 100 point scale. In addition, the probability of a rating being reported is significantly higher if the value is equal to or above 90 points, relative to scores between 85 and 89 points. We also find evidence that the seller engages in rating shopping (Skreta and Veldkamp, 2009), a behavior in which the seller releases only the highest score for a given wine when more than one are available. Second, our experiment demonstrates that expert scores play an important role in consumer choices. Estimates from a linear probability model show that wines with a single score of 85 points are about 68 percent more likely to be chosen than are comparable wines with no ratings. We estimate that the price premium consumers would be willing to pay for a wine rated 85 points is 7.5 dollars, or 30 percent of the average wine price. The result is even more striking for wines rated 90 points, where the price premium is as high as 15 dollars, or 60 percent of the average price. Third, we document that our information treatment did not stimulate a statistically meaningful change in consumers' behavior. In particular, we show that respondents in the control group act similarly to individuals who are made aware of censorship. Likewise, we do not find sufficient evidence to reject the hypothesis that consumers who do not receive an information treatment evaluate expert scores differently than do consumers who have been informed of seller manipulation.

We discuss three alternative explanations consistent with these findings. First, consumers may be naive, in that they do not anticipate censorship when making a purchase

decision. It is also possible that their naivete is persistent, and that their behavior does not change even when they are explicitly warned about the presence of censorship. Consequently, strategic sellers can extract a greater surplus from consumers by manipulating and censoring expert scores than by disclosing all scores. A second explanation is that consumers are sophisticated enough that they are not surprised to learn of sellers' behavior, or more broadly that they are not affected by censorship. A third possibility is that participants in our experiment did not read, understand or act on our information treatment.

Our work contributes to several strands of literature. First, a number of studies evaluate the impact of professional reviews on sales in different markets like movies (Reinstein and Snyder, 2005), books (Sorensen, 2007; Garthwaite, 2014; Reimers and Waldfogel, 2021), wines (Hilger, Rafert and Villas-Boas, 2011; Thrane, 2019; Bonnet, Hilger and Villas-Boas, 2020; Cheng and Reyes, 2020; Villas-Boas, Bonnet and Hilger, 2021), showing that negative reviews can sometime have no effect on sales (Friberg and Gronqvist, 2012) or even increase sales (Berger, Sorensen and Rasmussen, 2010). The relevance of crowd-based reviews has been similarly investigated in the book (Chevalier and Mayzlin, 2006), restaurant (Luca, 2016), and movie industries (Duan, Gu and Whinston, 2008). More generally, quality disclosure has important applications in a variety of fields, including education, finance, healthcare, consumer goods, and services (see Dranove and Jin, 2010 for a review). We provide a meaningful contribution by studying the interaction between information provision and sellers incentives, suggesting a link between these two strands of literature, and analyzing the two phenomena in a single context. So far, the literature has overlooked the importance of how information is delivered to consumers, with the notable exception of a few papers focusing on strategic reporting in movies (Reinstein and Snyder, 2005), school rankings (Luca and Smith, 2015), and restaurants (Dai and Luca, 2020). We add to the literature on review manipulation, which has primarily focused on inauthentic reviews (Luca and Zervas, 2016), with a unique analysis of censorship (see

Hauser, 2020, for a theoretical discussion). From a methodological standpoint, our strategy combines information treatment (Haaland, Roth and Wohlfart, 2022) and stated choice methods (Louviere, Hensher and Swait, 2000), contributing to the literature that studies the impact of information in an experimental setting as in Senecal and Nantel (2004). Finally, our paper belongs to the wine economics literature, and relates most strongly to the vast body of literature interested in understanding the meaning and impact of expert scores. In particular, several papers (Hilger, Rafert and Villas-Boas, 2011; Bonnet, Hilger and Villas-Boas, 2020; Villas-Boas, Bonnet and Hilger, 2021) have studied the willingness to pay for expert scores in traditional retail settings, Gokcekus and Gokcekus (2019) show that consumers tend to mentally cluster wines above and below 90 points rather than considering scores in a continuous way, and Moreno-Ternero (2020) documents a recent increase in the number of critics. Most closely to us, Thrane (2019) and Cheng and Reyes (2020) study the relationship between expert and crowd-based ratings in experimental settings, but ignore the impact of strategic disclosure. Our extension of their analysis is a crial step in investigating the potential welfare effects of censorship.

The remainder of this paper is divided as follows: Section 1.2 introduces the wine market, and the information available to consumers. Section 1.3 describes how we collected and analyzed expert ratings from the seller's website and a comprehensive database. In Section 1.4 we present our survey experiment and discuss its implementation. Section 1.5 highlights our empirical strategy. We discuss the main results in Section 1.6, and Section 1.7 concludes.

1.2 Market Background

The online retail market for wines is a setting with several features that lend themselves to examining the role of expert and crowd-based reviews.

First, the United States is the leading wine market in the world, with an overall size

of $74.1 billion in 2019.[6] In the same year, the online retail component reached $2.6 billion, with a steady growth expected to continue and increase in the future (Nesin, 2019). Consumers can access wine products in the online market through a number of channels: online stores, wine clubs, subscriptions, and auctions. In our analysis we focus on the most traditional online shopping experience, where individuals purchase wines by choosing from a list of options, and have them delivered to their place of residence. In addition, we restrict our study to bottles that cost less than 250 dollars, whose sales make for more than 95 percent of the market value[7], in order to abstract from considerations for luxury goods.

Second, wines represents a clear example of an experience good, and are characterized by a great deal of asymmetric information on quality between producers, sellers and buyers. As a consequence, consumers need to rely on observable product attributes to infer their quality, including prices (Rao, 2005; Shiv, Carmon and Ariely, 2005), prizes won, and results from blind tastings (Hilger, Rafert and Villas-Boas, 2011; Bonnet, Hilger and Villas-Boas, 2020; Villas-Boas, Bonnet and Hilger, 2021). Evidence also shows that quality signals are positively associated to the prices that producers and sellers can charge (Ali, Lecocq and Visser, 2010; Paroissien and Visser, 2020).

Third, the wine sector has a well established system of independent expert opinions in place, published in specialized magazines, journals and websites. More recently, crowd-based reviews have emerged as a complement to traditional ratings. While the focus of this paper is on expert score censorship, we think it is important to include crowd-based scores in our analysis, since they are very likely to mitigate the impact of expert ratings when consumers make purchase decisions. Section 1.2.1 contains additional details on wine ratings, as well as other attributes that consumers take into consideration.

[6] See data from the Wine Institute (link accessed March 3, 2022).
[7] See statistics based on Nielsen data, as reported by Statista.com (url accessed April 19th, 2022).

1.2.1 Pre-Purchase Information

Expert Ratings. The judgment of experts matters in many types of markets, and the wine sector is no exception. Indeed, a vast body of literature has shown that opinion leaders can be influential in affecting the price (Roberts and Reagans, 2007; Ali, Lecocq and Visser, 2010; Dubois and Nauges, 2010; Masset, Weisskopf and Cossutta, 2015; Oczkowski and Pawsey, 2019) and popularity of wine (Friberg and Gronqvist, 2012; Villas-Boas, Bonnet and Hilger, 2021). Expert reviews are usually published by specialized magazines, but their ratings, most commonly based on a 100-point scale, are widely available to the public and are often advertised by sellers or directly by the producers, in the form of a sticker on the bottle. While wine quality is hard to define, most experts provide a guide to interpret their ratings, as we show in Appendix Table 1.A.1 for three of the most popular sources: Wine Spectator, the Wine Advocate, and Wine Enthusiast. Wines with scores above 80 points are generally considered at least good, while scores of 90 or more denote superb or outstanding products. In online retail, expert scores are reported in the wine description, and associated to a label indicating its source. Importantly, their inclusion can be decided by the sellers, without having to commit to a specific policy. As a consequence, they can decide whether to publish any score at all, some of them, or all information that is available.

Traditionally, expert ratings were limited to premium wines, which is the focus of much of the existing literature on the economics of wine. However, the increase in online publications and dedicated websites has generated an increase in the supply of expert sources; that is publications where more than one expert is involved in the rating. It is also common for winemakers to submit samples of their wines to experts for review in order to advertise and promote their product. In some cases (e.g. Wine Enthusiast), a score below 80 points is not reported by the graders, making it almost impossible for the public to discern wines that are not recommended from those that were never tested and

rated.

Crowd-Based Reviews. Crowd-based ratings differ from expert ratings along many dimensions: first, they are provided by individuals, who often remain anonymous, and who can leave reviews on a number of different platforms. Their reliability rests on the number of consumers who have tasted a specific wine, rather than on the identity of the reviewer. In addition, crowd-based ratings are often collected by independent sources that have little interest in manipulating those scores. Having been generated in a decentralized way, crowd-based scores can cover a wider pool of wines than can specialized publications.

These ratings generally fall on a 5-star scale, reporting the average score (rounded to the nearest half-star) over all individual contributions. The quality signal is therefore a combination of the average score and the sample size, and may differ from the information provided by experts. Discrepancies between scores reflect differences in preferences, which is well-documented in the literature (Goldstein et al., 2008).

Additional Wine Attributes. In principle, many observable characteristics of wines can be taken as a signal of quality. As we mentioned before, price is an obvious proxy (Roberts and Reagans, 2007; Goldstein et al., 2008), as are grape variety and regional appellations (Schamel and Anderson, 2003), winery reputation (Landon and Smith, 2012), bottle shape and label (Sáenz-Navajas et al., 2013), and even bottle closures (Bekkerman and Brester, 2019).

1.3 Wine Ratings Data

In this section we document that sellers can act strategically in disclosing or censoring expert scores. With this goal in mind, we build a unique dataset combining information from two main sources: a leading online wine retailer in the Unites States, and a comprehensive database of expert ratings. In addition, we complement this information with

crowd-based reviews from a popular mobile wine application.

Online repositories have inherently different incentives than do sellers, as they try to maximize the quality of information provided to users both in terms of the coverage and precision of the available information. On the other hand, we can safely speculate that sellers report ratings with the ultimate goal of maximizing their profits. As a consequence, the distribution of reported scores may differ between the two sources for the same sample of wines. Comparing ratings that are reported to those which are omitted can provide useful insight as to whether censorship exists, and on what type of strategic decisions are made by the seller. Importantly, our analysis does not investigate the seller's decision on which wines should appear in the store. This could potentially be an important factor in traditional brick and mortar retail venues, when shelf space is finite. The online retailer we study has the explicitly stated goal of providing the most extensive variety of wines in the world, covering all price ranges. We are therefore confident that the individual score of each wine is likely to play a negligible role in the decision on whether or not it is sold.

1.3.1 Sample and Data Collection

We collect data from three independent sources over the period of December 2020 to May 2021: a comprehensive database of wine ratings, a major online wine retailer, and a mobile wine app. Appendix Section 1.B.1 describes in detail how the collection process was performed.

Wine Seller. From the seller's website, we identified approximately 340,000 wines listed on their virtual shelf, including those that are no longer available for purchase (for example past vintages and sold out items). We use this set of products as the initial sample, and we subsequently attempt to match each observation with additional sources. For each wine we collect the relevant attributes, including name, winery, appellation, vintage, color, variety, country of origin, price, and expert ratings. More than 30 percent of

the wines are linked to at least one expert rating, obtained from seven different publications.

Comprehensive Score Database. We searched for every wine found on the seller's website in the comprehensive score database, and link any matches to the original observation. To maximize data quality, we only allow for perfect matches based on name, winery and vintage. Appendix Table 1.B.7 provides summary statistics from the process, and shows that close to 30 percent of the wines were correctly paired. We are able to collect the list of expert scores, truncated to five elements, i.e. we are only able to observe a random sample of five ratings whenever there are six or more.

Wine App. To complement the seller's dataset with a more comprehensive coverage of crowd-based reviews, we gather data from Vivino, a popular mobile wine app. Vivino is used by more than 55 million users and contains almost 15 million wines rated on a 5-star scale, allowing for half-star increments[8]. Overall, we are able to collect data from each of the three sources for roughly 20 percent of the initial sample. While we do not actively use the Vivino data in our censorship analysis, we rely on the data in designing our discrete choice experiment, as crowd-based scores represent a potentially important factor in consumers' demand for wines.

1.3.2 Analysis

The main goal of the analysis in this section is be to compare scores from the sellers to those from the independent database. We therefore restrict the sample to wines that were available in both. In addition, we only keep wines made after the year 2010, and with a price below 250 dollars, in order to limit the impact of outliers. Finally, we only include expert scores from sources reported in both datasets. This leaves us with 36,052 unique

[8] At the time of writing, a higher level of disaggregation is available for Android users.

wines and 60,515 scores, as described in Table 1.1 containing the descriptive statistics for our sample.

The average price of wines with at least one score is only slightly above the unrestricted mean by roughly six dollars, suggesting that even if expensive wines are reviewed more frequently, selection does not seem to play an important role. Appendix Figure 1.A.1 plots the distribution of expert scores by price, clearly showing that most of our ratings come from wines costing less than 100 dollars, with half of the distribution falling below 50 dollars.

Empirical Strategy. In the interest of understanding the determinants of expert scores disclosure from the seller, we estimate the following linear probability model

$$Disclosed_{ijv} = \beta_j + \beta_v + f(score_{ijv}) + \beta^x \cdot X_i + \epsilon_{ijv} \tag{1.1}$$

where the variable *Disclosed* is a dummy that takes a value of one if the score for wine i, from expert j and vintage v is reported by the seller. We include expert and vintage fixed effects, to control for systematic differences in the reviewers and year specific factors that may affect quality. Our goal is to evaluate the impact of expert score values in a flexible way, to allow for non-linear effects and potential discontinuities. Individual wine attributes are included as additional controls; namely price, wine type, varietal, and country of origin.

Results. In Figure 1.1 we plot, for each score value, the fraction of expert ratings reported by the seller on their website. It is quite evident from the raw data that higher scores are more likely to be disclosed, with a strong discontinuity around the 90 point threshold. Another interesting feature of the graph is that, notwithstanding how high the score is, only about three quarters of the available ratings are displayed, suggesting that

there may be reasons other than censorship why a score is not reported.[9] To quantify the magnitude of this apparent censorship, we estimate equation 1.1 and present results in Table 1.2.

Columns (1) and (2) show that results are robust to the inclusion of a full set of controls. We include expert scores as a categorical variable, pooling scores below 89 points in the baseline category. First, it should be noted that scores of less than 89 points are unlikely to be reported, as the average is only 2 percent. A score of 89 is about 20 percentage points more likely to be displayed on the website, a tenfold increase. Wines above 90 points are even more likely to be reported, close to 60 percent of the time.

In column (3) we investigate the possibility that a score is reported if it is the highest available for a specific wine. This is only true for ratings equal to 89 points, for which the probability of disclosure goes up by an additional 13.3 percentage points if no better score is available for that wine.

Appendix Table 1.A.2 provides the same type of analysis, using only sources whose ratings are reported at least 30 percent of the time. The results are fairly similar, and if anything they show more persuasive evidence of the existence of rate shopping, which applies also for scores below 89 points.

Robustness. Our data is limited in the sense that we only have information from a single, although large, seller. We are therefore concerned that our results may be driven by chance. In particular we are interested in understanding if the same score distribution could emerge if sellers were not acting strategically. In Appendix 1.B.2 we describe and perform two simulation exercises that try to answer the following questions: (1) how

[9] One possibility is that, at the time we performed data collection, the seller was not aware of the existence of the score, or that some experts are reported less often than others. To mitigate the issue, we restrict our analysis to experts that the seller explicitly declares to source from on their website. However, we noted that some experts, while mentioned on the website, are clearly underrepresented in the seller's dataset. For instance expert scores from Vinous are reported only 12 percent of the time. Removing these outliers (namely, sources with less than 30 percent reporting rate) does not change the interpretation of our results. In Appendix Figure 1.A.2 we replicate Figure 1.1 without outliers. The effect of censorship appears, if anything, more pronounced than what we observe in the full sample.

would the distribution of disclosed scores look, if the seller were choosing at random from the comprehensive database, and (2) what would be the distribution of reported ratings if the seller were to report only the highest score available for each wine.

In Appendix Figure 1.B.5 we compare the actual score distribution to the simulated one, showing that it is unlikely that the seller is randomly choosing what ratings they disclose. Similarly, Appendix Figure 1.B.6 provides strong evidence that the seller is reporting only the maximum values available, compared to a simulated scenario in which ratings are randomly censored.

1.3.3 Discussion

We have provided arguably convincing evidence the the seller is likely to strategically disclose expert ratings, especially when they can engage in rating shopping (Skreta and Veldkamp, 2009); that is whenever they have more than one value to choose from. The net reduction in information delivered to consumers can affect their welfare if they care about expert ratings, for example by shifting consumption towards higher rated and more expensive wine. In the remainder of the paper, we investigate this possibility by describing and implementing a survey design. Our goal is to evaluate whether consumers take expert scores into consideration when choosing a wine, and whether they are aware of potential censorship by the seller.

1.4 Online Survey Experiment

To assess if the pattern of censorship described in the previous Section can be relevant for consumers' welfare, we designed an online survey to understand how individuals make decisions when purchasing a bottle wine. In addition, we are interested in investigating whether consumers are aware of censorship by the seller, and if it plays a role in their choices. We asked respondents to perform a discrete choice experiment in a simulated

retail environment, following the vast literature on stated choice methods (see Louviere, Hensher and Swait, 2000, for a review), and combined it with an information provision treatment, drawing from the emerging literature on this specific experimental method (Haaland, Roth and Wohlfart, 2022).

1.4.1 Sample and Data Collection

We sampled individuals from the population of American residents, aged 21 or older, who had purchased wine at least once in the month preceding the date of the interview. The survey was run in February 2022, with a sample size of 1,173 individuals. We relied on *Respondi*, a panel access company, to distribute the survey through their mailing list. By clicking on a link, respondents were redirected to our experimental platform where they could access and accept the consent form for our research study. They were assured that no identifiable data would be collected, and that *Respondi* had not shared with us any personal information. They were also informed of their rights as research subjects, in particular of their right to interrupt the survey at any moment (see Appendix 1.C.1 for a transcript of the consent form). The survey company was also involved in the initial screening and remuneration of the respondents, although we performed additional screening at the beginning of the questionnaire. To maximize the response rate, we did not put any hard quotas on additional individual characteristics. Compensation for successfully completing the survey, set by *Respondi*, was determined based on the respondent's stated preferences, and their recruitment channel. Compensation could be either cash or reward points on loyalty programs with partner companies. To ensure the quality of responses, we screened out individuals who were too fast in reading the instructions, or in answering multiple choice questions. Appendix Table 1.C.9 provides a summary of our sample. The median completion time was around 17 minutes. Appendix Table 1.C.10 reports descriptive statistics on respondents demographics, grouped by treatment assignment. Our sample is comparable to the survey run by Velikova et al. (2021) in providing

a representative sample of wine consumers. In particular, it should be noted that wine consumers have in general higher levels of education than the average population. In our sample, roughly 62 percent of the respondents hold a college degree or higher, compared to the US average of 37.9 percent (U.S. Census Bureau, 2021). Consistently with this finding, women are slightly over represented than men (56.61 percent of respondents identified as women).

1.4.2 Survey Design

The survey was designed in *oTree* (Chen, Schonger and Wickens, 2016), and is composed of several parts. Initially, respondents undergo an additional screening block where we make sure they were correctly selected by the panel company in relation to wine consumption, age and location. Having successfully passed the screening process, they are directed to our consent form and, subsequently, to the initial instructions. The main part of the survey, a discrete choice experiment, is divided into two rounds of 15 questions. At the beginning of each section, a randomized information treatment is provided. Figure 1.2 shows the initial instruction page, containing a screenshot of the virtual market in which respondents make their choices. The survey ends with a background questionnaire. The full protocol is available in Appendix 1.C.

Discrete Choice. The main task required in our survey is to perform a discrete choice experiment in a simulated online wine retail environment. Each respondent is presented with a series of virtual shelves, each made up of five wines representing popular wine categories from countries widely available in the United States. We require individuals to imagine a real online store experience, and to select a wine from the list. We always allow for the possibility that no purchase is made as our alternative option, and ask each respondent to complete two rounds of 15 questions each, for a total of 30 choices. Appendix Figure 1.C.7 shows a screenshot of the virtual market displayed in the survey.

Wine Characteristics. The wines shown in the virtual markets are characterized by several attributes. We include information usually available on online wine stores, specifically the name of the wine, year, the origin, its price, as well as expert and crowd-based scores for the wine. Each market[10] is generated in three steps: first, we determine the color of the wine (red or white). Second, we randomly select 5 wines of that color from a pool of 200 wines, priced between 10 and 40 dollars, that we randomly generated from the seller's dataset. Third, we randomly assign expert and crowd-based scores to each wine, drawing from a distribution that mimics what we observe in the complete dataset for expert scores. Specifically, each wine can be assigned anywhere between zero and three expert ratings, with values ranging from 80 to 95 points, and a crowd-based average between 2.5 and 4.5, with sample size between 4 and 1,010. The exact generating process and the underlying score distributions are described in Appendix 1.C.3.

Background Information. The final part of the survey contains a series of background questions to gather individual characteristics and information on the respondents consumption habits. We asked them to provide information on their gender, education, city, income, and employment status. We also ask about how often they purchase wine, where they buy it, and what attributes they take into account when choosing which wine to purchase. The complete questionnaire is available in Appendix 1.C.1.

1.4.3 Information Treatment

A major component of our experiment is the information provision treatment, designed to investigate consumers awareness of censorship and its potential impact on demand and welfare. At the beginning of the survey, individuals are randomly assigned into six groups, each made of two treatment blocks (one per round), with probabilities described in Figure 1.4. They are all presented with an initial instruction page (Figure 1.2) explaining the goal

[10] We denote by *market* the menu of wines that pertain to a single individual choice. Over the course of the survey, each respondent makes 30 choices from 30 different markets.

of the experiment, the tasks they have to perform, and a background description of what wine ratings are. Before the beginning of each round, respondents are treated with an information provision that differs based on the group they are assigned to. We modify the information set of our respondents in two ways: the first is to display information messages, while the other is by manipulating the expert scores they can observe. For each block, we identify five possible treatments: control, No Censoring, Warning 1, Warning 2, and Full Information.

Control. Before starting their discrete choice task, individuals in the control group only receive basic instructions from the initial page. When performing the virtual wine selection, they observe expert scores from a censored distribution, similar to what would happen on the seller's website. More precisely, we remove all expert scores below 85, and 75 percent of scores between 85 and 90 points. This scenario is our closest representation of a real world setting, although it should be noted that having randomized expert scores is an important departure from the actual purchasing experience.[11]

No Censoring. Respondents assigned to the No Censoring group receive only the initial instructions, and are identical to the control individuals in this respect. However, when performing the virtual wine selection, they observe expert scores from a complete distribution, that is a distribution where we report every rating available, as described in Appendix 1.C.3. This scenario is slightly different from the real world, because respondents are able to observe ratings usually censored in the actual online store. It has the important benefit of allowing us to estimate the impact of lower expert scores on consumers' demand.

[11] In particular, some sophisticated consumers may already know the actual scores obtained by some wines, and be confused if they observe different values. We believe the only a small share of our respondents belongs to this category, and that choosing from a large pool of wines can mitigate the impact of this effect in our experiment.

Warning 1. Individuals in the Warning 1 treatment group are provided with an additional information message after the initial instructions. This additional message informs them of our findings from Section 1.B. In particular, we make them aware that sellers can selectively disclose expert ratings, and that values below 85 points are never reported, values between 85 and 89 points are rarely reported, and scores above 90 points are almost always reported. The text of the message is displayed in Figure 1.3a. After receiving this piece of information, they move on to the discrete choice task, where they observe expert scores from a complete distribution, similar to the No Censoring group.[12]

Warning 2. Subjects assigned to the Warning 2 treatment group observe the same information message as those in the Warning 1 group, in which they are warned against censorship by the seller. On the other hand, this treatment group differs with respect to the Warning 1 treatment group as individuals cannot observe censored ratings in the subsequent discrete choice task, similar to what happens in the control group.

Full Information. Individuals in the Full Information group receive two messages: in the first message, we inform them of our findings on censorship, similar to the messages received by the Warning 1 and Warning 2 groups. We then explicitly state that in the experiment task we display scores from every available source (see Figure 1.3b for the exact message).

In our empirical analysis we estimate the impact of expert scores on consumers' demand, for individuals before they receive our information provision, that is in the control and No Censoring groups.[13] We subsequently compare how the effect varies depending on treatment, in order to understand if respondents react to censorship and take into ac-

[12] We are aware that consumers may feel confused by the fact that, in the discrete choice exercise, they will end up observing score values we claim are always censored. This is however the only option we have to estimate consumers' demand for those score values after they receive the information treatment. We introduce the Warning 2 treatment to mitigate this issue and provide respondents with a more realistic experience.

[13] By *information provision* we intend the messages shown at the beginning of each round. When referring to consumers, the attribute *uninformed* used in this context is to be interpreted as "consumers who have not received an information message".

count this censorship when considering which wine to purchase. We are interested in three types of comparisons: first, we compare uninformed consumers from the No Censoring group to those belonging to Full Information. Differences in behavior between the two groups would demonstrate that individuals are not aware of censorship, and can therefore benefit from additional information. Second, we compare respondents in the control and Warning 2 groups, as differences in behavior would demonstrate that they are not aware of censorship, and that they react to it when properly informed. Third, we compare subjects from the Warning 2 and Full Information groups, where a change in behavior would signal that censorship affects consumers decisions and can thus have an impact on their welfare. Section 1.5.2 describes in greater detail our hypotheses and the empirical strategy we implement to test them.

1.5 Empirical Strategy

Having presented the structure of the survey, we now describe our empirical approach to identify the impact of expert scores on consumers' choices, and to understand if they are aware of censorship, taking it into consideration when making a purchase decision.

1.5.1 Impact of Wine Scores

We look at the impact of expert wine scores by estimating the following linear probability model pooling together the control and No Censoring individuals, that is individuals who have not received any information message.[14]

[14] In the control group we display scores from a distribution that censors the lowest values, while in the No Censoring arm we show ratings from the full, uncensored, distribution. In principle, we might expect individuals to update their behavior based on how frequently each score appears, or to change their opinion on wines without ratings. In Appendix Figure 1.A.3 and Appendix Table 1.A.3 we compare the behavior of individuals in the control and No Censoring groups, to understand if their behavior differs depending on whether scores come from a full or censored distribution. We do not find convincing evidence that the two groups treat expert scores in a different way, and we therefore conduct our analysis pooling the two groups to improve the precision of our estimates.

$$
\begin{aligned}
Selected_{vim} = & \beta_m + price_v + \\
& f(expert\_score_{vm}) + \beta^{EN} \cdot N\_expert\_scores_{vm} + \\
& f(crowd\_score_{vm}) + \beta^{CS} \cdot N\_crowd\_scores_{vm} + \\
& \beta^x \cdot X_i + \beta^v \cdot V_v + \epsilon_{vim}
\end{aligned} \tag{1.2}
$$

Selected is a dummy variable that takes a value of 1 if the wine v is selected by individual i in market m, and 0 otherwise. A market is defined as the group of five wines available to the individual in each virtual shelf, and it is added to control for the quality of alternative options. We also include the price of wine in the regression, and allow for a non linear effect of expert and crowd-based score values. In addition, we include the interaction between the two terms to allow for complementarities between expert and crowd-based sources. Wine specific controls include varietal, country of origin, and vintage. Demographic controls include age, gender, education, self reported income, self reported effort and a variable that measures wine app usage. A detailed definition of the questions used to generate demographic controls is available in Appendix Section 1.C.1.

1.5.2 Assessing Consumers' Behavior

The second topic we address with our analysis is whether consumers are aware of seller censorship and how they address this censorship when making purchases. We identify consumers' attitudes by analyzing their behavior in the different treatment branches of our experiment.

No Censoring vs Full Information. The first question we are interested in answering is whether consumers who have not received our information treatment behave differently from individuals who know that they are observing all the available expert scores. Namely, these consumers know that every available score is reported. We can test this

hypothesis by estimating the following linear model:

$$
\begin{aligned}
Selected_{vim} = & \beta_m + price_v + \beta_1 \cdot full\_information_{im} + \\
& f(expert\_score_{vm}) + \boldsymbol{\beta_2} \cdot full\_information_{im} \cdot f(expert\_score_{vm}) + \\
& \beta^{EN} \cdot N\_expert\_scores_{vm} + \\
& f(crowd\_score_{vm}) + \beta^{CS} \cdot N\_crowd\_scores_{vm} + \\
& \beta^{v} \cdot V_v + \epsilon_{vim}
\end{aligned}
\tag{1.3}
$$

where, as before, *Selected* is an indicator variable that equals 1 if the wine is selected, and 0 otherwise. The variable *full_information* is a dummy variable that takes a value of one if the respondent has been subject to the Full Information treatment, and zero if the respondent is in the No Censoring treatment group. We are interested in the sign and magnitude of the elements of vector $\boldsymbol{\beta_2}$, which describes the impact that our information message regarding the distribution of expert scores has on consumers' demand for wine.

Failing to reject the hypothesis that $\boldsymbol{\beta_2 = 0}$ can have different interpretations: (1) that uninformed consumers' demand is not affected by censorship, as their behavior does not change when operating in a market with full information. Alternatively, (2) that they are already aware of censorship, or can infer it by looking at the scores displayed, and can therefore take it into account when making a selection, or (3) that uninformed consumers are naive, meaning that they treat expert scores as if they were drawn from a complete, uncensored, distribution.

Control vs Warning. A second comparison we make is whether uninformed consumers behave differently than consumers who have received information on censorship. In order to test this hypothesis we compare individuals in the control and Warning 2 groups and estimate a model similar to equation 1.3:

$$
\begin{aligned}
Selected_{vim} = & \beta_m + price_v + \beta_1 \cdot warning2_{im} + \\
& f(expert\_score_{vm}) + \boldsymbol{\beta_2} \cdot warning2_{im} \cdot f(expert\_score_{vm}) + \\
& \beta^{EN} \cdot N\_expert\_scores_{vm} + \\
& f(crowd\_score_{vm}) + \beta^{CS} \cdot N\_crowd\_scores_{vm} + \\
& \beta^{v} \cdot V_v + \epsilon_{vim}
\end{aligned} \tag{1.4}
$$

where $warning2$ takes a value of 1 if the individual is in the the Warning 2 treatment group, and 0 if the individual is in the control group. We are again interested in the coefficient vector $\boldsymbol{\beta_2}$ that shows how the relevance of expert scores changes in the different treatment groups.

Rejecting the null hypothesis that $\boldsymbol{\beta_2 = 0}$ would demonstrate that consumers can benefit from information[15] by changing their behavior, providing evidence that uninformed consumers are not completely anticipating the seller's strategy, and that in turn censorship may impact their welfare. On the other hand, failing to reject the null hypothesis can have different interpretations: (1) that consumers are already aware of censorship, therefore providing them with additional information does not affect their demand, (2) that consumers' demand is not affected by censorship, hence their behavior would not change, or (3) that consumers are persistently naive, in the sense that they do not update their behavior even after they are made aware of the censorship.

Full Information vs Warning. Finally, we compare consumers behavior in the Warning 2 and Full Information groups of our experiment. This will allow us to evaluate how consumers who have been informed about censorship operate in two scenarios, one where

[15] In this context, information is to be interpreted as our warning message.

censorship is practiced, and one in which it is not. We estimate the following equation:

$$\begin{aligned} Selected_{vim} =& \beta_m + price_v + \beta_1 \cdot full\_information_{im} + \\ & f(expert\_score_{vm}) + \boldsymbol{\beta_2} \cdot full\_information_{im} \cdot f(expert\_score_{vm}) + \\ & \beta^{EN} \cdot N\_expert\_scores_{vm} + \\ & f(crowd\_score_{vm}) + \beta^{CS} \cdot N\_crowd\_scores_{vm} + \\ & \beta^{v} \cdot V_v + \epsilon_{vim} \end{aligned} \tag{1.5}$$

where *full_information* is a dummy that takes a value equal to 1 if the individual belongs to the Full Information treatment group, and 0 if the individual belongs to the Warning 2 group. Once more, we are interested in the the sign and magnitude of the coefficient vector $\boldsymbol{\beta_2}$, which indicates how censorship affects consumers demand. The differences between the two groups are both in the distribution of expert scores shown to them, and in the fact that individuals in the Full Information group are aware of what distribution the scores come from. Rejecting the null hypothesis that $\boldsymbol{\beta_2} = 0$ would provide evidence that censorship does affect consumers' decisions and can thus have an impact on their welfare. On the other hand, failing to reject the null hypothesis can be interpreted in two ways: (1) that censorship does not have an impact on consumers' demand, or (2) that consumers are persistently naive, in the sense that they do not update their behavior when censorship is practiced and they are made aware of it.

1.6 Results

We now present evidence of the impact of wine scores on the probability of purchasing a wine, as well as evidence of the impact of our information treatment. Table 1.3 contains descriptive statistics on the virtual wines shown to consumers. Overall, our sample consists of 175,950 wine options priced slightly below 24 dollars on average, almost evenly

divided between reds and whites, from France or the United States.

1.6.1 Impact of Wine Scores

We start our analysis trying to understand the impact that quality signals have on the probability of choosing a wine. In Figure 1.5a we plot the proportion of wines selected as a function of the expert ratings for the control and the No Censoring groups together. We report in blue a measure based on the maximum score available, while in red we perform the analysis for the subset of wines with a single rating. The data shows a clearly increasing relationship between expert score values and the probability that the wine is selected by the consumer. Moreover, wines that score less than 85 points do not seem to be more likely to be chosen than do wines with no score. In Figure 1.5b we depict the same relationship using crowd-based scores. Wines with a higher average score are more likely to be chosen than wines with lower values, and the effect is stronger for larger sample sizes throughout the full distribution of score values.

Given that both expert and crowd-based scores, prices and wine attributes were randomized, we are confident that these simple descriptive graphs are robust to the inclusion of controls. To provide more precise quantitative results, we estimate equation 1.2 for individuals in the control and the No Censoring groups. Consistently with Figure 1.5a, we use the highest scores available for each wine as our preferred measure, including them as a categorical variable, while controlling for crowd-based ratings and the interaction between the two. Table 1.4 shows our results, in two different specifications that produce similar estimates. One estimate contains a market fixed effect to take into account the quality of alternatives, while the other estimate contains demographic characteristics of the individual. Standard errors are always clustered at the individual level.

We are interested in the size and sign of the coefficients on expert scores, as well as the coefficient relative to the numbers of scores available. Since each wine attribute has been independently randomized, we can recover the causal effect of expert scores on the

probability that a wine is selected, while keeping other factors constant. In particular, we report estimates for wines having a crowd-based average rating of 3.5[16], and omit from the table the interaction terms, which are not statistically different from zero.

To measure the impact of a single expert rating, we need to combine the coefficient relative to its value, to the coefficient that captures the effect of the number of scores available. In Figure 1.6, panel (a) we compute the linear combination of these two coefficients and report the point estimate for different score values and numbers of ratings, together with a 95 percent confidence interval. Wines with one score equal to 85 points are 5.5 percentage points more likely to be chosen than wines with no ratings, a roughly 68 percent increase with respect to a comparable wine with no expert ratings and a crowd-based score of 3.5. Wines that score even a single 90-point are more than twice as likely to be selected as similar unrated wines with a crowd-based score of 3.5.

Given that the price has been randomized independently of scores, we can exploit the estimated price coefficient from Table 1.4 to look at the impact of expert scores in terms of willingness to pay. In Figure 1.6b we compute, for each score value, the increase in price that would make that wine as likely to be selected as a comparable wine with no scores. Based on our estimates from column (1), a wine with a single rating of 85 points is as likely to be selected as a similar wine that costs about 7.5 dollars less, a 30 percent reduction in price. Similarly, a single score of 90 points is "worth" roughly 15 dollars, a more than 60 percent decrease in price.

These results confirm that expert scores have a substantial impact on consumers' demand, and that this effect is present and economically meaningful not only for scores that are usually reported, but also for values that are often censored, namely values between 85 and 89 points. To better understand the welfare implications of our findings it is important to evaluate how censorship itself affects the way consumers make decisions.

[16]We select this baseline value for two reasons: first, the average score of a wine on Vivino, the most widely used wine rating app, is 3.6 as reported on their website (accessed April 13th, 2022); second, the sign of the regression coefficients relative to crowd-based scores are symmetric for averages below and above 3.5.

External Validity. A natural question that arises when dealing with simulated purchases is whether the results are externally valid. Individuals operating in a virtual market may behave differently from the real world, in particular, they may be less sensitive to prices given that they do not have to spend real money. A back-of-the-envelope computation of the price elasticity gives a value of -1.2[17], which is well within the range found in the literature using real data for the United States.[18]

Heterogeneity. Being confident that expert scores do in fact play a role in consumers' demand, it is plausible to imagine that consumers may simply select the wine with the largest available score, ignoring other attributes like price or crowd ratings. To understand if this is the case, we estimate equation 1.2 allowing for heterogeneous effects depending on whether the expert score reported is the highest available in the market. The results, shown in Appendix Table 1.A.4, do not seem to support this hypothesis. The coefficient associated to the dummy variable labeled *Scores is highest in the market* measures the increase in probability that a wine with no expert score is chosen, whenever that score is the highest in the market. It refers to the particular case in which all the wines in a specific market have no score, and it is not statistically different from zero. The point estimate is also fairly noisy given that only a small fraction of markets are generated with no expert ratings. The interaction terms, which show the additional impact of each score whenever it is the largest available are not significantly different from zero. This suggests that individuals take into account expert scores in conjunction with other attributes, rater than merely selecting the highest score.

[17] To compute the elasticity we used the price coefficient estimated in column (1) of Table 1.4, the average probability that a wine is selected and the average wine price in the control and No Censoring treatments. This gives $\varepsilon_p \approx -\frac{0.00732}{0.145} \cdot 23.86 \approx -1.2$.

[18] A recent paper by Bonnet, Hilger and Villas-Boas (2020) reports a value of -2.671. A more comprehensive study by Davis, Ahmadi-Esfahani and Iranzo (2008) which uses random utility models to compute elasticities for different regions finds an average of -1.80 for the United States in 2003, with a standard deviation of 1.354. Their estimates range from -0.74 for basic wines to -25.63 for ultra premium wines.

1.6.2 Assessing Consumers' Behavior

Having confirmed in Section 1.6.1 that expert ratings have a substantial effect on the demand of uninformed consumers, we turn our attention to their interaction with censorship.

In Figure 1.7 we present a descriptive graph similar in spirit to Figure 1.5, in which we plot, for each score, the fraction of wines selected by our respondents, disaggregating the data by treatment status. More specifically, in Figure 1.7a we compare individuals in the No Censoring and Full Information groups, showing that they behave in a remarkably similar way. The same is true in Figure 1.7b, where we focus on the Full Information and Warning (1 and 2) groups. Qualitatively, respondents who have been warned about censorship and experience it (Warning 2) seem to place more value on higher scores, although the difference in means is not statistically significant.[19] To formally test our hypotheses, we estimate the empirical models presented in Section 1.5.2 and show results in Table 1.5.

Column (1) contains estimates for equation 1.3 to test for changes in behavior between the No Censoring and the Full Information groups. The difference between the two is that individuals belonging to the latter have been informed that they are observing scores from an independent source with no censorship. We are mostly interested in the sign and magnitude of the interaction terms, as these coefficients reflect differences in behavior between the two treatments. We are comforted to find that baseline coefficients are very similar to Table 1.4, and we do not find convincing evidence that the treated respondents evaluate expert ratings in a different way than individuals in the No Censoring group. The individual interaction terms are small in magnitude, and despite two of them being

[19] In Appendix Figure 1.A.4 we replicate Figure 1.7b and substitute scores in the Warning 2 treatment with the original randomized scores, that is before we censored them using the algorithm described in Section 1.4.2. It can be seen that wines that score below 90 points have a similar probability of being chosen as wines with no score. This is reasonable to expect, given that due to censorship these wines often appear identical when it comes to their expert score. Namely, they will appear with no rating on the shelf because the score has been censored.

significantly different from zero (88 and 91 points), we can only marginally reject null hypothesis that they are jointly equal to zero.[20]

The pattern displayed in column (1) is very similar to what we also observe in column (2), where we compare the control and Warning 2 treatment groups. None of the interaction terms are statistically different from zero, with the exception of the coefficient relative to a score of 91 points, that follows a pattern similar to what we described in column (1). We cannot reject the null hypothesis that the coefficients are jointly equal to zero, and that individuals behave in the same way before and after receiving our information treatment.[21]

Finally, in column (3) we show estimates for equation 1.5, looking at the Full Information and Warning 2 treatment groups. We observe no major difference in behavior, with two coefficients being marginally significant, 89 and 93 points. Again, the estimates seem to show the pattern already observed in columns (1) and (2); in particular the coefficient relative to an 89 point score is smaller than expected in the Warning 2 group, if compared to the coefficients relative to scores of 85 and 90 points. We can only marginally reject the null hypothesis that they are jointly equal to zero. We note however a small but significant decrease in the selection of wines with no scores in the Full Information setting.[22]

In Appendix Table 1.A.6 we perform the same analysis presented in Table 1.5, substituting individual score values with a pooled variable that takes three values: no score, maximum score below 90 points, maximum score equal or above 90 points. The results are fairly similar to our original analysis, and the precision of our estimates does not improve despite adding additional restrictions to the model.

[20] We are willing to attribute this statistically significant effects to spurious correlation, noting in particular that the coefficient relative to a score of 91 points in the No Censoring group is surprisingly lower than the coefficient relative to 90 points. Adding the interaction term would put it right in between the magnitude of the 90 and 93 points coefficients. The same is true for the 88 points coefficient, lower than the 85 points estimate.

[21] In Appendix Table 1.A.5 we perform the same analysis of column (2) comparing the No Censoring and Warning 1 treatments, obtaining similar results.

[22] A possible explanation is that, given that the Warning 2 scenario censors many ratings, it creates a larger number of markets with no scores at all.

1.6.3 Discussion

In the preceding discussion, we analyze the impact of the treatment effects administered in our survey experiment. The main takeaway of our results is that we are not able to stimulate a statistically meaningful reaction from consumers with respect to expert scores censorship through information messages. Several factors may be responsible for our findings, for which we discuss potential causes and their implications below.

Our first concern is that respondents did not read or understand our information treatment, that they were not concerned about it, or simply that the treatment was not strong enough. We cannot rule out any of these possibilities, although we took steps to make these events less likely.

We made efforts to design our treatment in a clear, informative, and visually salient way following the standards of the emerging information provision literature (Haaland, Roth and Wohlfart, 2022), and tried to ensure the quality of our data. In particular, we screened out respondents who were not spending enough time on the instruction and treatment pages. As reported in Appendix Table 1.C.9, close to 35 percent of eligible respondents were screened out for rushing through instructions; moreover, more than 75 percent of the subjects who completed the survey self-reported having put "a lot of effort" into responding, and more than 95 percent put at least "some effort" into their responses. Finally, close to 20 percent of the respondents decided to leave an optional comment at the end of the survey, with an average length of about 60 characters. In addition, our results are consistent with an established literature that documents how consumers can be inattentive when making decisions (Eyster and Rabin, 2005; Brown, Camerer and Lovallo, 2012, 2013; Luca, 2016; Dai and Luca, 2020), and may neglect information disclosure (Hirshleifer, Lim and Teoh, 2004; Smirnov and Starkov, 2021).

As the analysis in Section 1.6.1 demonstrates, expert scores play an important role in consumer demand, by increasing the value of a wine in the eyes of the consumers. In Sec-

tion 1.6.2, we failed to reject the null hypotheses that consumers behave differently when provided information about censorship, as highlighted in Section 1.5.2. More specifically, we found that uninformed consumers are similar to individuals who are aware that scores come from a full distribution. Consistent with the literature, this result may provide evidence that consumers are in fact naive, in the sense that they do not take censorship into consideration when making a purchase. Another possible interpretation that we cannot rule out is the opposite: that consumers are not naive and are aware of censorship, and so their behavior was not affected by our information disclosure. Finally, it as also plausible that our respondents did not read or care about our information message.

We also did not witness any differential behavior when comparing the Control and Full Information groups to the Warning 2 group. As previously mentioned, we can interpret this result in different ways. First, we can take this as evidence that consumers are not only naive, but also persistent in their naivete, such that they do not react to plausible information treatments. If this were the case, a strategic seller might extract additional surplus by engaging in selective disclosure and censorship. Understanding their pricing strategy would be critical in accurately measuring the associated welfare costs for consumers. Alternatively, we can interpret the same result as evidence that consumers are already aware of censorship, or that the censored information is irrelevant for them, such that disclosing censorship does not affect their preferences. Finally, we still cannot exclude that respondents did not read, understand or take into consideration the information message we provided.

1.7 Conclusion

Wine is a notorious example of an experience good, whose quality is difficult to ascertain before purchase and consumption. Consumers can benefit from the availability of pre-purchase information, including expert and crowd-based ratings that are made available by the sellers on their online retail platforms. In this paper we study sellers' and con-

sumers' attitudes towards experts ratings, and the ways in which censorship can affect said attitudes.

First, we construct a unique dataset of expert ratings that contains data from a major wine retailer and an independent source, and use it to show that the seller engages in substantial censorship of information by omitting scores below a certain threshold and publishing most scores above 90 points.

Given the seller's behavior, it becomes relevant to understand whether consumers do in fact value the information provided, and whether they are aware that it can be strategically censored. To do so, we design and implement an online survey, combining an information treatment and a discrete choice experiment. Respondents were asked multiple times to select a wine from a virtual shelf. Individuals in the treatment group were provided with either a warning message, informing them that sellers censor lower scores (Warning 1 and Warning 2), or a message stating that the source of the scores was complete (Full Information). In addition, respondents observed wine scores from different distributions based on their treatment assignment: some (Control and Warning 2) saw scores from a censored distribution, while others (Warning 1, No Censoring and Full Information) saw scores from a complete distribution.

Our results show that consumers do care about expert ratings, since wines with higher scores are more likely to be selected than are comparable wines with lower scores. Our information treatment did not stimulate a significant change in consumers' behavior with respect to expert ratings. We discuss three possible interpretations for our findings: first, it is possible that consumers are naive in the sense that they always act as if they observe ratings from a complete distribution. Alternatively, consumers might be sophisticated enough that they are not surprised to learn about the seller's behavior. Finally, we cannot the possibility that participants in our the experiment did not read or understand our information treatments, or that they were not concerned about the treatment.

References

Ali, Héla Hadj, Sébastien Lecocq, and Michael Visser. 2010. "The impact of gurus: Parker grades and en primeur wine prices." *Journal of wine economics*, 5(1): 22–39.

Bekkerman, Anton, and Gary W Brester. 2019. "Don't Judge a Wine by Its Closure: Price Premiums for Corks in the US Wine Market." *Journal of Wine Economics*, 14(1): 3–25.

Berger, Jonah, Alan T. Sorensen, and Scott J. Rasmussen. 2010. "Positive Effects of Negative Publicity: When Negative Reviews Increase Sales." *Marketing Science*, 29(5): 815–827.

Bonnet, Celine, James Hilger, and Sofia B. Villas-Boas. 2020. "Reduced Form Evidence on Belief Updating under Asymmetric Information. Consumers' Response to Wine Expert Opinions." *European Review of Agricultural Economics*, 47(5): 1668–1696.

Brown, Alexander L., Colin F. Camerer, and Dan Lovallo. 2012. "To Review or Not to Review? Limited Strategic Thinking at the Movie Box Office." *American Economic Journal: Microeconomics*, 4(2): 1–26.

Brown, Alexander L., Colin F. Camerer, and Dan Lovallo. 2013. "Estimating Structural Models of Equilibrium and Cognitive Hierarchy Thinking in the Field: The Case of Withheld Movie Critic Reviews." *Management Science*, 59(3): 733–747.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "Otree–an Open-source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance*, 9: 88–97.

Cheng, Christopher, and Martin Reyes. 2020. "Impact of Crowd-Sourced Wine Ratings on Purchasing Behavior in a Retail Environment." American Associationof Wine Economists Working Paper No. 252.

Chevalier, Judith A., and Dina Mayzlin. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research*, 43(3): 345–354.

Dai, Weijia, and Michael Luca. 2020. "Digitizing Disclosure: The Case of Restaurant Hygiene Scores." *American Economic Journal: Microeconomics*, 12(2): 41–59.

Davis, Timothy R., Fredoun Z. Ahmadi-Esfahani, and Susana Iranzo. 2008. "Demand under product differentiation: an empirical analysis of the US wine market." *Australian Journal of Agricultural and Resource Economics*, 52(4): 401–417.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature*, 48(4): 935–63.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston. 2008. "Do online reviews matter? An empirical investigation of panel data." *Decision Support Systems*, 45(4): 1007–1016. Information Technology and Systems in the Internet-Era.

Dubois, Pierre, and Celine Nauges. 2010. "Identifying the Effect of Unobserved Quality and Expert Reviews in the Pricing of Experience Goods: Empirical Application on Bordeaux Wine." *International Journal of Industrial Organization*, 28(3): 205–212.

Eyster, Erik, and Matthew Rabin. 2005. "Cursed equilibrium." *Econometrica*, 73(5): 1623–1672.

Fishman, Michael J., and Kathleen M. Hagerty. 2003. "Mandatory Versus Voluntary Disclosure in Markets with Informed and Uninformed Customers." *The Journal of Law, Economics, and Organization*, 19(1): 45–63.

Friberg, Richard, and Erik Gronqvist. 2012. "Do Expert Reviews Affect the Demand for Wine?" *American Economic Journal: Applied Economics*, 4(1): 193–211.

Garthwaite, Craig L. 2014. "Demand Spillovers, Combative Advertising, and Celebrity Endorsements." *American Economic Journal: Applied Economics*, 6(2): 76–104.

Gokcekus, Omer, and Samin Gokcekus. 2019. "Empirical evidence of lumping and splitting: Expert ratings effect on wine prices." *Wine Economics and Policy*, 8(2): 171–179.

Goldman, Dan, Sophie Marchessou, and Warren Teichner. 2017. "Cashing in on the US experience economy." McKinsey Insights.

Goldstein, Robin, Johan Almenberg, Anna Dreber, John W Emerson, Alexis Herschkowitsch, and Jacob Katz. 2008. "Do more expensive wines taste better? Evidence from a large sample of blind tastings." *Journal of Wine Economics*, 3(1): 1–9.

Grossman, Sanford J. 1981. "The informational role of warranties and private disclosure about product quality." *The Journal of Law and Economics*, 24(3): 461–483.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2022. "Designing information provision experiments." *Journal of Economic Literature (forthcoming)*.

Hauser, Daniel N. 2020. "Censorship and Reputation." *Available at SSRN 3534133*.

Hilger, James, Greg Rafert, and Sofia Villas-Boas. 2011. "Expert Opinion and the Demand for Experience Goods: An Experimental Approach in the Retail Wine Market." *The Review of Economics and Statistics*, 93(4): 1289–1296.

Hirshleifer, D., S. Lim, and S. Teoh. 2004. "Disclosure to an Audience with Limited Attention." *Corporate Law: Law and Finance*.

Landon, Stuart, and Constance E Smith. 2012. "Quality Expectations, Reputation, and Price." In *World Scientific Reference on Handbook of the Economics of Wine: Volume 2: Reputation, Regulation, and Market Organization*. 3–31. World Scientific.

Louviere, Jordan J, David A Hensher, and Joffre D Swait. 2000. *Stated Choice Methods: Analysis and Applications*. Cambridge university press.

Luca, Michael. 2016. "Reviews, reputation, and revenue: The case of Yelp. com." Harvard Business School Working Paper 12-016.

Luca, Michael, and Georgios Zervas. 2016. "Fake it till you make it: Reputation, competition, and Yelp review fraud." *Management Science*, 62(12): 3412–3427.

Luca, Michael, and Jonathan Smith. 2015. "Strategic disclosure: The case of business school rankings." *Journal of Economic Behavior and Organization*, 112: 17 – 25.

Masset, Philippe, Jean-Philippe Weisskopf, and Mathieu Cossutta. 2015. "Wine tasters, ratings, and en primeur prices." *Journal of Wine Economics*, 10(1): 75–107.

Miller, Jon R., Ismail Genc, and Angela Driscoll. 2007. "Wine Price and Quality: In Search of a Signaling Equilibrium in 2001 California Cabernet Sauvignon." *Journal of Wine Research*, 18(1): 35–46.

Moreno-Ternero, Olivier Gergaud; Victor Ginsburgh; Juan D. 2020. "Wine Ratings." American Association of Wine Economists Working Paper No. 257.

Nesin, Bourcard. 2019. "US Online Alcohol Sales Reach Usd 2.6bn." Rabobank RaboResearch.

Oczkowski, Edward, and Nicholas Pawsey. 2019. "Community and Expert Wine Ratings and Prices." *Economic Papers: A journal of applied economics and policy*, 38(1): 27–40.

Paroissien, Emmanuel, and Michael Visser. 2020. "The Causal Impact of Medals on Wine Producers' Prices and the Gains from Participating in Contests." *American Journal of Agricultural Economics*, 102(4): 1135–1153.

Rao, Akshay R. 2005. "The quality of price as a quality cue." *Journal of marketing research*, 42(4): 401–405.

Reimers, Imke, and Joel Waldfogel. 2021. "Digitization and Pre-purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings." *American Economic Review*, 111(6): 1944–71.

Reinstein, David A., and Christopher M. Snyder. 2005. "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics." *The Journal of Industrial Economics*, 53(1): 27–51.

Roberts, Peter W., and Ray Reagans. 2007. "Critical Exposure and Price-quality Relationships for New World Wines in the US Market." *Journal of Wine Economics*, 2(1): 84–97.

Sáenz-Navajas, María-Pilar, Eva Campo, Angela Sutan, Jordi Ballester, and Dominique Valentin. 2013. "Perception of Wine Quality According to Extrinsic Cues: The Case of Burgundy Wine Consumers." *Food Quality and Preference*, 27(1): 44–53.

Schamel, Günter, and Kym Anderson. 2003. "Wine Quality and Varietal, Regional, and Winery Reputations: Hedonic Prices for Australia and New Zealand." In *World Scientific Reference on Handbook of the Economics of Wine: Volume 2: Reputation, Regulation, and Market Organization*. 33–58. World Scientific.

Senecal, Sylvain, and Jacques Nantel. 2004. "The Influence of Online Product Recommendations on Consumers' Online Choices." *Journal of retailing*, 80(2): 159–169.

Shiv, Baba, Ziv Carmon, and Dan Ariely. 2005. "Placebo effects of marketing actions: Consumers may get what they pay for." *Journal of marketing Research*, 42(4): 383–393.

Skreta, Vasiliki, and Laura Veldkamp. 2009. "Ratings shopping and asset complexity: A theory of ratings inflation." *Journal of Monetary Economics*, 56(5): 678–695. Carnegie-Rochester Conference Series on Public Policy: Distress in Credit Markets: Theory, Empirics, and Policy November 14-15, 2008.

Smirnov, Aleksei, and Egor Starkov. 2021. "Bad News Turned Good: Reversal under Censorship." *American Economic Journal: Microeconomics*, Forthcoming.

Sorensen, Alan T. 2007. "Bestseller Lists and Product Variety." *The Journal of Industrial Economics*, 55(4): 715–738.

Thrane, Christer. 2019. "Expert reviews, peer recommendations and buying red wine: experimental evidence." *Journal of Wine Research*, 30(2): 166–177.

U.S. Census Bureau. 2021. "Educational Attainment in the United States: 2021." Report CB22-TPS.02.

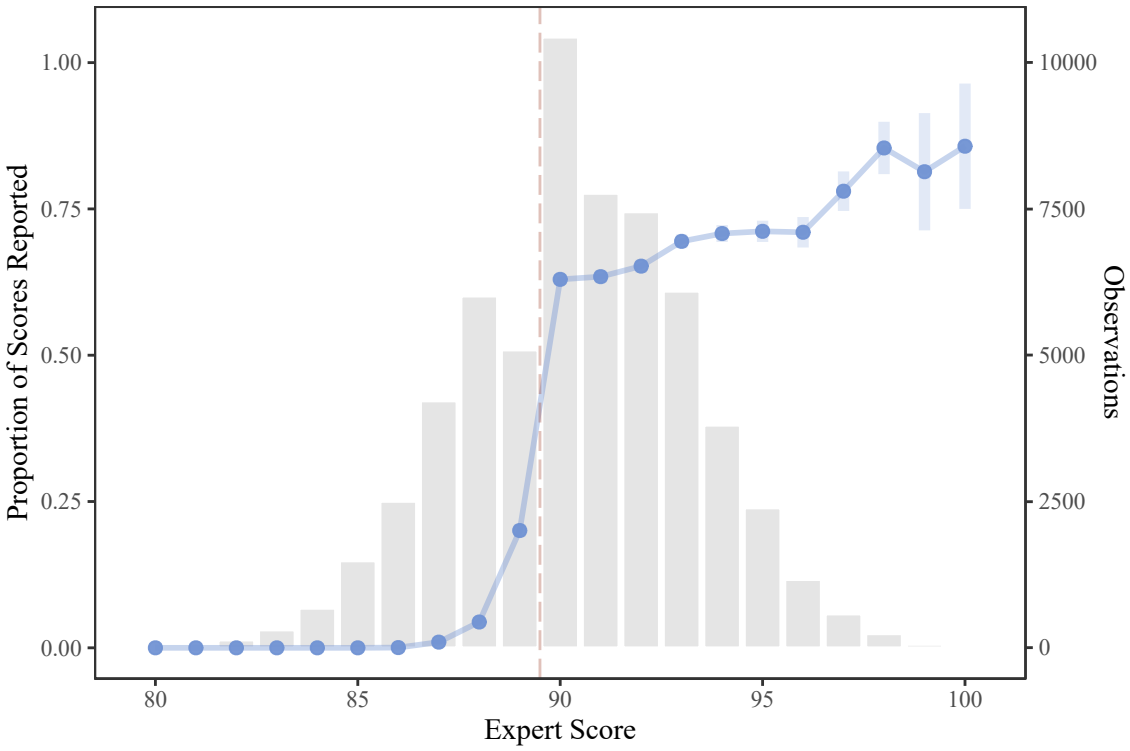
Velikova, Natalia, Oleksandra Hanchukova, Bogdan Olevskyi, and Hunter D'Camp. 2021. "The Effect of Covid-19 on U.S. Wine Consumption: Six Months After the Original Lockdown." Texas Wine Marketing Research Institute Research Report.

Villas-Boas, Sofia B., Celine Bonnet, and James Hilger. 2021. "Random Utility Models, Wine and Experts." *American Journal of Agricultural Economics*, 103(2): 663–681.

1.8 Tables and Figures

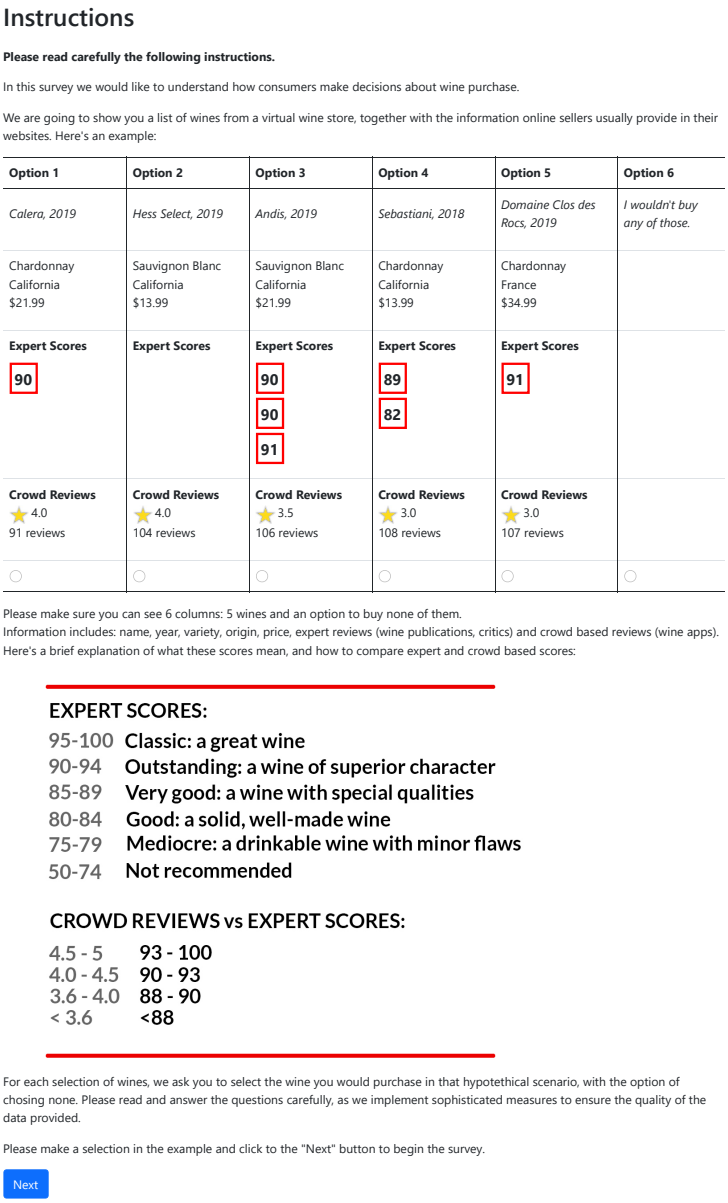
1.8.1 Figures

Figure 1.1. Expert Scores Reported by the Seller by Value



Notes: Each point indicates the fraction of expert scores reported by the seller by score value, with the 95 percent confidence interval for the mean (left axis). On the right axis, gray bars display the number of observations for each rating. For the sake of clarity, we dropped 4 observations below 80 points that are not reported by the seller.

FIGURE 1.2. BASIC INSTRUCTIONS



Instructions

Please read carefully the following instructions.

In this survey we would like to understand how consumers make decisions about wine purchase.

We are going to show you a list of wines from a virtual wine store, together with the information online sellers usually provide in their websites. Here's an example:

Option 1	Option 2	Option 3	Option 4	Option 5	Option 6
Calera, 2019	*Hess Select, 2019*	*Andis, 2019*	*Sebastiani, 2018*	*Domaine Clos des Rocs, 2019*	*I wouldn't buy any of those.*
Chardonnay California $21.99	Sauvignon Blanc California $13.99	Sauvignon Blanc California $21.99	Chardonnay California $13.99	Chardonnay France $34.99	
Expert Scores 90	**Expert Scores**	**Expert Scores** 90 90 91	**Expert Scores** 89 82	**Expert Scores** 91	
Crowd Reviews 4.0 91 reviews	**Crowd Reviews** 4.0 104 reviews	**Crowd Reviews** 3.5 106 reviews	**Crowd Reviews** 3.0 108 reviews	**Crowd Reviews** 3.0 107 reviews	
○	○	○	○	○	○

Please make sure you can see 6 columns: 5 wines and an option to buy none of them.
Information includes: name, year, variety, origin, price, expert reviews (wine publications, critics) and crowd based reviews (wine apps).
Here's a brief explanation of what these scores mean, and how to compare expert and crowd based scores:

EXPERT SCORES:

95-100	Classic: a great wine
90-94	Outstanding: a wine of superior character
85-89	Very good: a wine with special qualities
80-84	Good: a solid, well-made wine
75-79	Mediocre: a drinkable wine with minor flaws
50-74	Not recommended

CROWD REVIEWS vs EXPERT SCORES:

4.5 - 5	93 - 100
4.0 - 4.5	90 - 93
3.6 - 4.0	88 - 90
< 3.6	<88

For each selection of wines, we ask you to select the wine you would purchase in that hypotethical scenario, with the option of chosing none. Please read and answer the questions carefully, as we implement sophisticated measures to ensure the quality of the data provided.

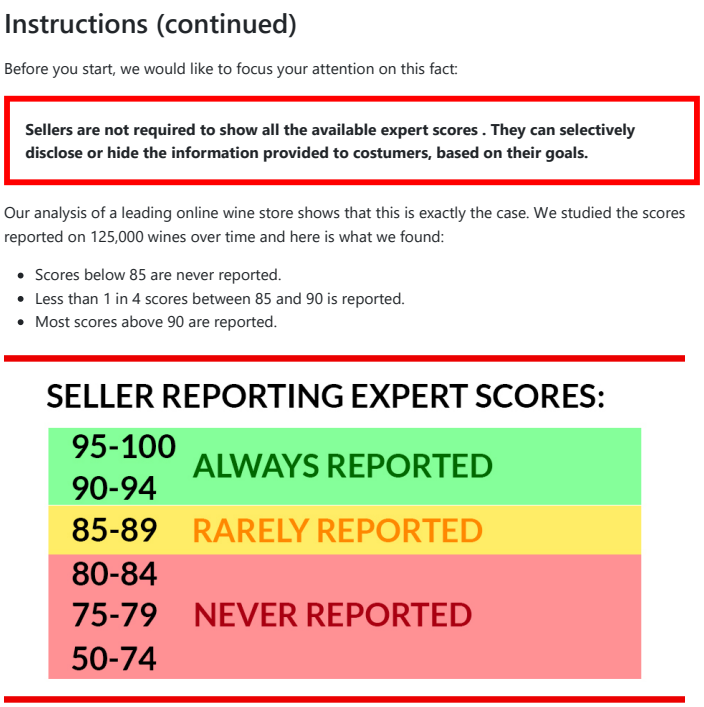
Please make a selection in the example and click to the "Next" button to begin the survey.

Next

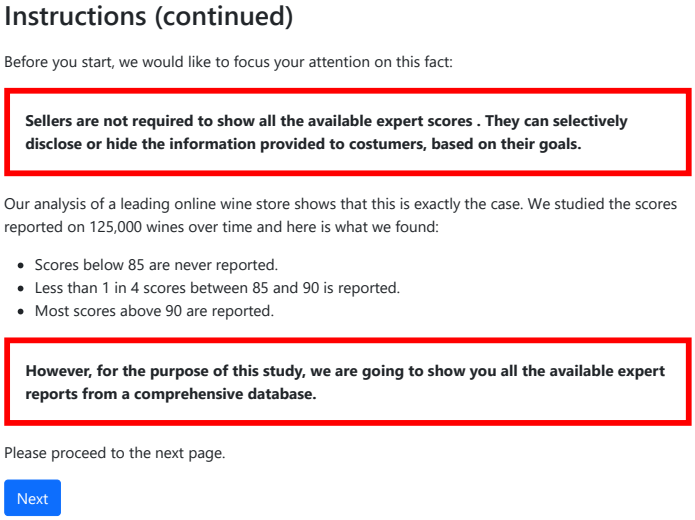
Notes: The figure displays a screenshot of the initial instruction page displayed to every respondent.

FIGURE 1.3. INFORMATION TREATMENTS

(a) Warning 1 and Warning 2 Information Message



(b) Full Information Message



Notes: The figure in panel (a) displays a screenshot of the warning message shown to the Warning 1 and Warning 2 individuals. The figure in panel (b) presents the message displayed to respondents in the Full Information treatment.

Figure 1.4. Survey Design

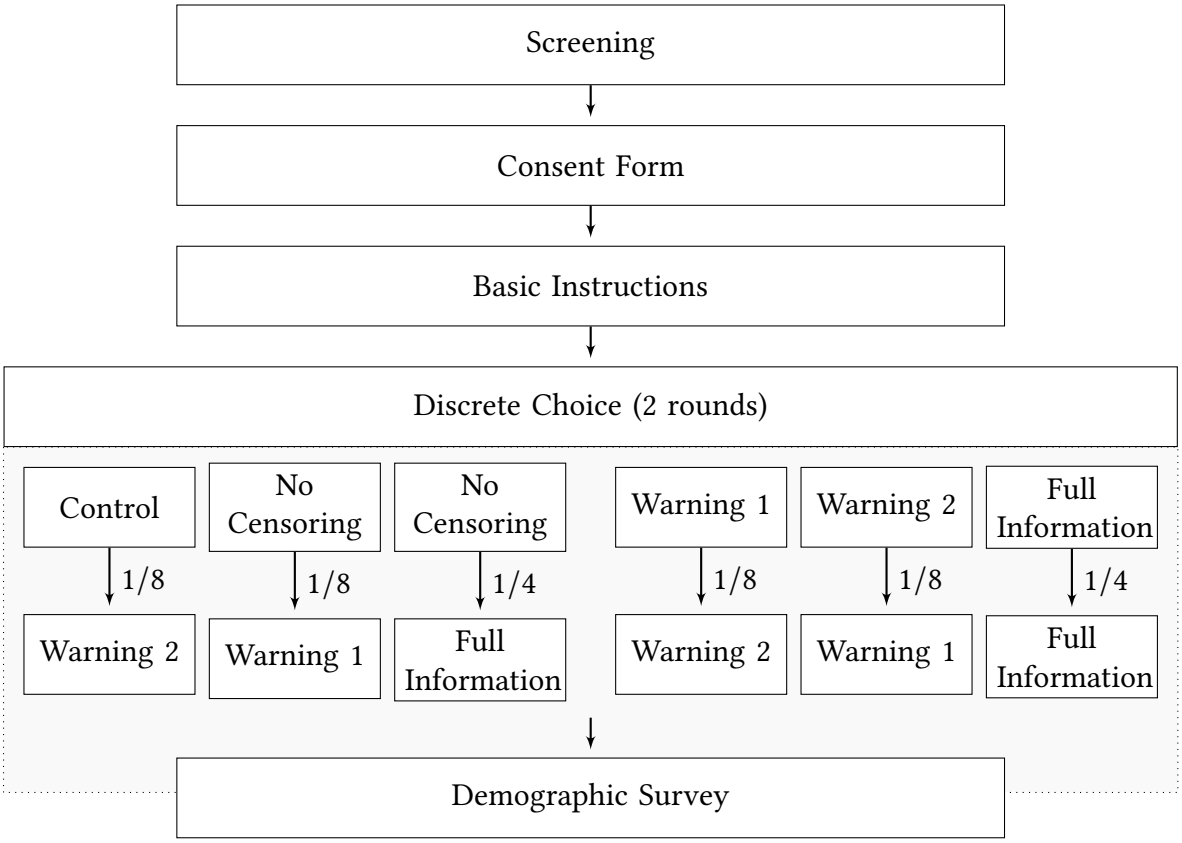
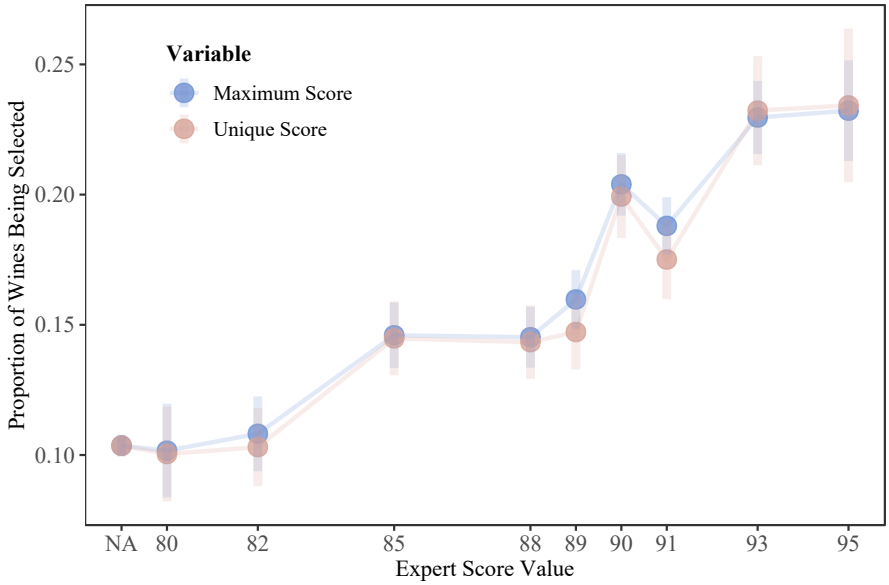
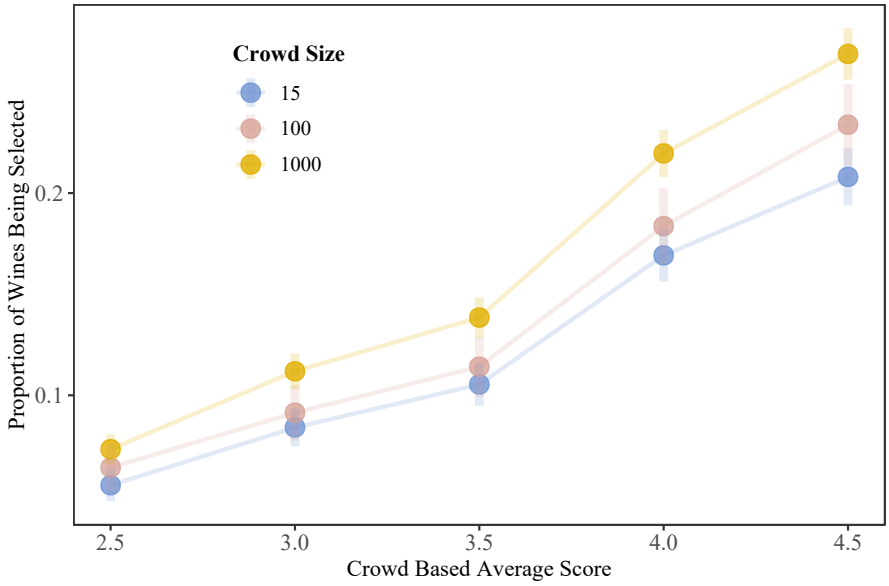


Figure 1.5. Descriptive Figures

(a) Wine Selection and Expert Scores



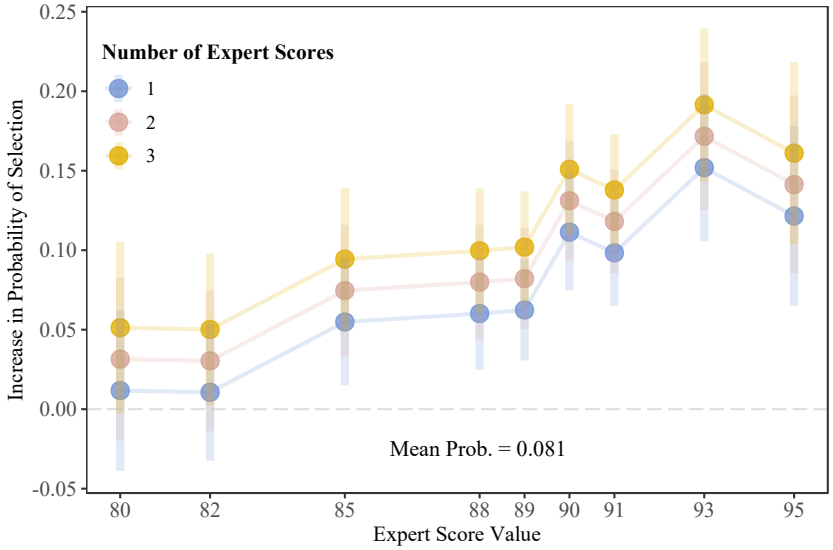
(b) Wine Selection and Crowd-based Scores



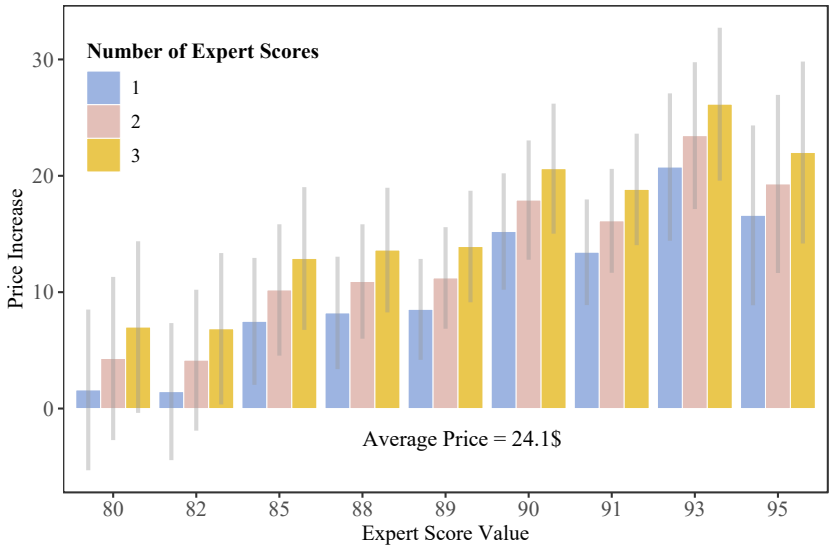
Notes: Figures show the collapsed mean of a dummy variable equal to 1 if the wine was selected for different values of expert scores (Panel a) and crowd-based scores (Panel b). Crowd sizes have been perturbed by adding random noise drawn uniformly from the set of integers in the interval [-11, 10]. Sample is restricted to individuals in the control and No Censoring groups.

Figure 1.6. Impact of Expert Scores on Demand

(a) Expert Scores and Wine Purchases



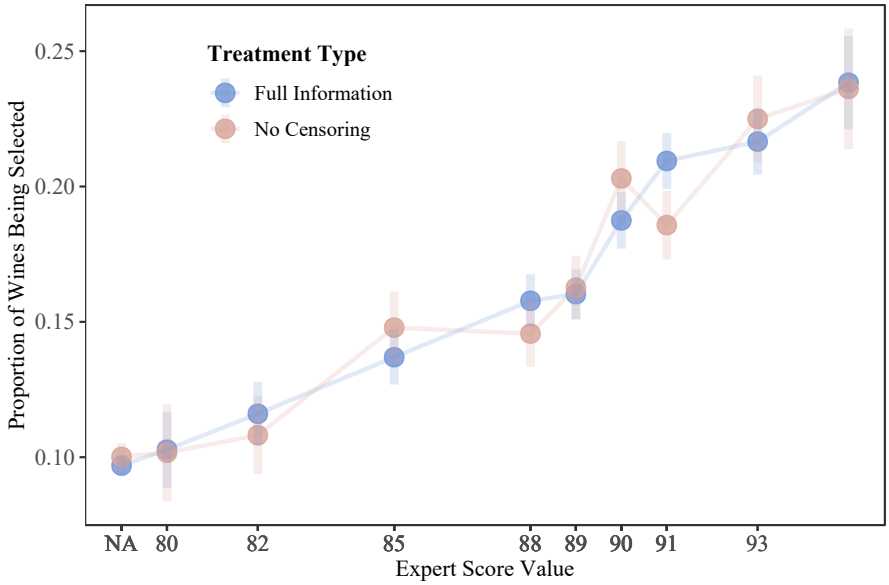
(b) Expert Scores and Price Premium



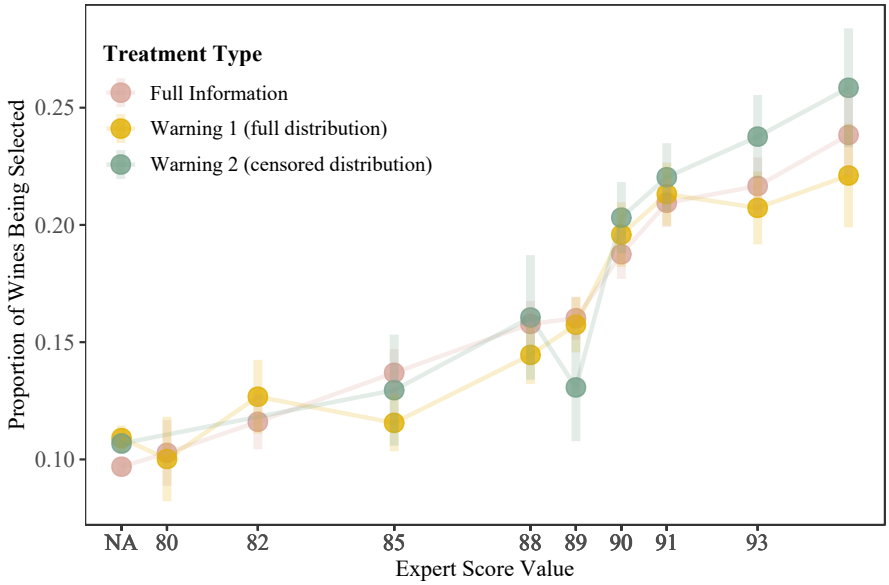
Notes: Figures are based on the estimates contained in column (1) of Table 1.4. In Panel (a) we plot the impact of expert ratings on the probability that a wine is chosen. The baseline probability reported at the bottom is computed for a wine with no expert scores available, and an average crowd-based score of 3.5. In Panel (b) we use the price coefficient from Table 1.4 to compute the average price increase required to equalize the purchase probability for a wine without scores and a comparable wine with one or more expert ratings. The baseline average price is computed for a wine with no expert scores available and an average crowd-based score of 3.5.

Figure 1.7. Wine Selection by Treatment Type

(a) No Censoring vs Full Information



(b) Full Information vs Warning (1 and 2)



Notes: Figures show the collapsed mean of a dummy variable equal to 1 if the wine was selected, for different maximum values of the expert score, by treatment assignment. Panel (a) compares the impact of expert scores in the control and Full Information settings. Panel (b) compares their effect in the control, Warning 1 and Warning 2 treatments. The algorithm that censors scores in the Warning 2 scenario is described in Section 1.4.2.

1.8.2 Tables

Table 1.1
Expert Ratings – Descriptive Statistics

Variable	N	Mean	Sd	Min	Max
Wine Data					
Price	36051	42.49351	39.68468	2.99	249.99
Max Exper Score	36052	90.62302	2.950233	70	100
Number of Scores	36052	1.678492	.8930554	1	7
Average Score	36052	90.08433	2.707763	70	100
Crowd Score Value	9429	4.06164	.4107764	2.1	5
Crowd Sample Size	36052	6.801897	28.61349	0	1305
Expert Score Data					
Reported by Seller	60513	.4606448	.4984529	0	1
Expert Score Value	60513	90.44371	2.861151	70	100

Notes: Table contains summary statistics on wines (top panel) and expert scores (bottom panel) from the independent database. The variable *Reported by Seller* takes a value equal to 1 if the score appears on the seller's website, and 0 otherwise.

Table 1.2
Expert Ratings – Regression Analysis

	(1)	(2)	(3)
Expert Scores (baseline: score is below 89 points)			
89 points	0.180**	0.205***	0.119**
	(0.0510)	(0.0252)	(0.0405)
90 points or more	0.643***	0.567***	0.587***
	(0.101)	(0.0562)	(0.0822)
Price ($100)		-0.00109	0.000707
		(0.0213)	(0.0226)
Scores is Highest Available			0.0311
			(0.0190)
Highest Score * 89 points			0.133**
			(0.0415)
Highest Score * 90 points or more			-0.0361
			(0.0413)
Mean of Dependent Variable for baseline	0.0199		
Observations	60,513	60,017	60,017
R^2	0.331	0.513	0.516
Controls	-	X	X
Expert FE	-	X	X
Vintage FE	-	X	X

Notes: Standard errors clustered at the expert source level are reported in parenthesis. Significance level: ***<0.01, **<0.05, *<0.1. Dependent variable is a dummy that takes a value of one if the score is reported by the seller. Expert rating is reported as a categorical variable, base level is a score below 89 points. Controls are added in columns (2) and (3) and include price, wine type dummy, varietal dummy, and country of origin dummy in addition to expert and vintage fixed effects. In column (3) we allow for heterogeneity based on whether a score is the highest available for a given wine.

TABLE 1.3
SURVEY EXPERIMENT – DESCRIPTIVE STATISTICS

	Treatment Assignment					
	Full Info	No Censoring - Full Info	No Censoring - Warning1	Control - Warning2	Warning 1	Warning 2
Observations	37,200	47,100	26,550	23,850	22,650	18,600
Share of Red Wines	0.51	0.49	0.50	0.51	0.50	0.51
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Share of French Wines	0.47	0.47	0.48	0.47	0.47	0.48
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Price	23.85	23.86	23.90	23.91	23.81	23.88
	(8.66)	(8.63)	(8.67)	(8.65)	(8.70)	(8.67)
N. Expert Scores	1.00	1.00	1.00	0.46	1.00	0.47
	(0.96)	(0.96)	(0.96)	(0.67)	(0.96)	(0.67)
Expert Score Avg.	87.62	87.61	87.61	90.50	87.64	90.50
	(3.46)	(3.48)	(3.47)	(2.32)	(3.46)	(2.31)
Crowd Score Avg.	3.50	3.50	3.50	3.50	3.50	3.50
	(0.71)	(0.71)	(0.71)	(0.71)	(0.71)	(0.71)
Crowd Sample Size	373	375	370	372	369	371
	(446.11)	(446.52)	(445.09)	(445.73)	(444.96)	(445.88)

Notes: Table contains the means of wine attribute variables for each treatment assignment. Standard deviations are reported in parentheses.

Table 1.4
Impact of Expert Scores – Regression Results

	(1)	(2)
Price	-0.00732***	-0.00635***
	(0.00179)	(0.00136)
Number of Expert Scores	0.0198***	0.0139***
	(0.00506)	(0.00374)
Expert Scores (baseline: no score)		
80 points	-0.00806	-0.0128
	(0.0263)	(0.0195)
82 points	-0.00912	-0.00913
	(0.0225)	(0.0168)
85 points	0.0351*	0.0271*
	(0.0209)	(0.0157)
88 points	0.0404**	0.0331**
	(0.0191)	(0.0141)
89 points	0.0426**	0.0325**
	(0.0176)	(0.0134)
90 points	0.0916***	0.0663***
	(0.0195)	(0.0147)
91 points	0.0785***	0.0646***
	(0.0186)	(0.0140)
93 points	0.132***	0.108***
	(0.0248)	(0.0185)
95 points	0.102***	0.0811***
	(0.0299)	(0.0221)
Crowd Based Scores (baseline: 3.5)		
Avg. Score below 3.5	-0.0455***	-0.0344***
	(0.00708)	(0.00514)
Avg. Score above 3.5	0.107***	0.0885***
	(0.00874)	(0.00667)
Crowd Sample Size (100 votes)	0.00439***	0.00372***
	(0.000547)	(0.000408)
Wine Controls	X	X
Demographic Controls	-	X
Market FE	X	-
Observations	48,750	48,750
R^2	0.194	0.120

Notes: Standard errors clustered at the individual level are reported in parentheses. Significance level: ***<0.01, **<0.05, *<0.1. Dependent variable is a dummy that takes a value of 1 if the wine is selected. Expert scores is represented by the maximum value available for each wine, and is reported as a categorical variable (baseline is no score available). The average crowd-based score is treated as a categorical variable (baseline score equal to 3.5). We include interaction terms between expert and crowd-based scores, coefficients (none of which are statistically significant) are omitted. Column (1) includes market fixed effect to control for the quality of alternative wines in addition to the wine specific controls: wine varietal, country of origin, vintage, and name. Column (2) includes respondents demographic controls: age, gender, education, income, self reported effort, and wine apps usage.

Table 1.5
Expert Scores and Treatment Assignment

	(1) No Censoring vs Full Information		(2) Control vs Warning 2		(3) Warning 2 vs Full Information	
Price	-0.00707***		-0.00534***		-0.00687***	
	(0.000926)		(0.00145)		(0.000940)	
Number of Expert Scores	0.0146***		0.0334***		0.0170***	
	(0.00244)		(0.00740)		(0.00286)	
Treatment Dummy	-0.00294		-0.00548		-0.0114***	
	(0.00351)		(0.00407)		(0.00392)	
Expert Scores	*Baseline Coeff.*	*Interaction Terms*	*Baseline Coeff.*	*Interaction Terms*	*Baseline Coeff.*	*Interaction Terms*
80 points	-0.00118	-0.00793	-	-	-0.0119	-
	(0.00878)	(0.0113)	-	-	(0.00774)	-
82 points	-0.000798	0.00202	-	-	-0.00119	-
	(0.00784)	(0.00980)	-	-	(0.00685)	-
85 points	0.0303***	-0.0102	-0.0158	0.00172	0.00234	0.0145
	(0.00793)	(0.00917)	(0.0192)	(0.0213)	(0.0125)	(0.0134)
88 points	0.0223***	0.0163*	-0.0189	0.0355	0.0342**	0.000761
	(0.00740)	(0.00865)	(0.0214)	(0.0246)	(0.0150)	(0.0160)
89 points	0.0356***	0.00232	-0.0147	0.000046	0.00376	0.0302**
	(0.00799)	(0.00895)	(0.0193)	(0.0211)	(0.0118)	(0.0128)
90 points	0.0684***	-0.0106	0.0474***	0.000260	0.0659***	-0.0123
	(0.00819)	(0.00914)	(0.0166)	(0.0160)	(0.00918)	(0.0107)
91 points	0.0523***	0.0217**	0.0367**	0.0247*	0.0823***	-0.0128
	(0.00888)	(0.00974)	(0.0145)	(0.0149)	(0.0105)	(0.0122)
93 points	0.0840***	-0.00205	0.0765***	0.00311	0.101***	-0.0244*
	(0.00985)	(0.0112)	(0.0187)	(0.0176)	(0.0118)	(0.0137)
95 points	0.0982***	0.00875	0.0534**	0.0366	0.112***	-0.0104
	(0.0120)	(0.0141)	(0.0249)	(0.0261)	(0.0159)	(0.0184)
Crowd Based Scores (baseline: 3.5)						
Avg. Score below 3.5	-0.0465***		-0.0398***		-0.0465***	
	(0.00295)		(0.00508)		(0.00313)	
Avg. Score above 3.5	0.0899***		0.0868***		0.0898***	
	(0.00479)		(0.00679)		(0.00472)	
Crowd Sample Size (in 100s)	0.00497***		0.00462***		0.00553***	
	(0.000349)		(0.000546)		(0.000372)	
F-stat Interaction Terms	1.658		0.828		1.903	
P-value for the F-test	0.0953		0.565		0.0661	
Observations	97,575		42,450		91,275	
R^2	0.105		0.105		0.103	

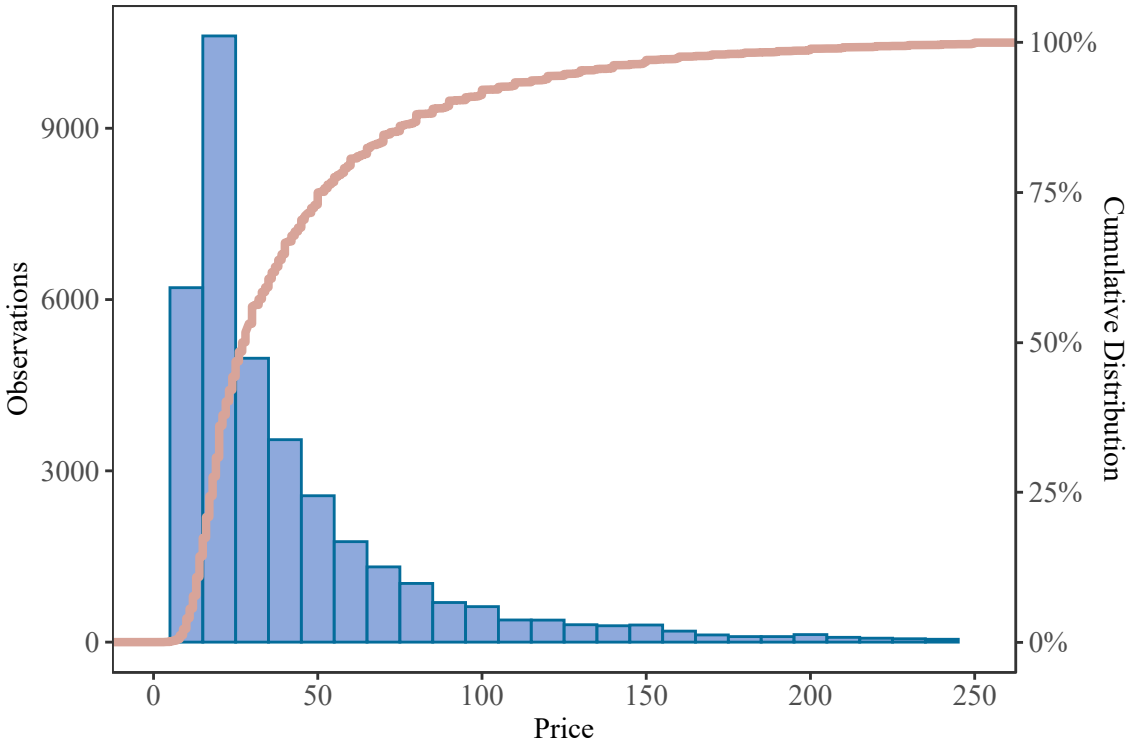
Notes: Standard errors clustered at the individual level are reported in parentheses. Significance level: ***<0.01, **<0.05, *<0.1. Dependent variable is a dummy that takes a value of 1 if the wine is selected. Expert scores is represented by the maximum value available for each wine, and is reported as a categorical variable (baseline is no score available). The average crowd-based score is treated as a categorical variable (baseline score equal to 3.5). All specifications include market fixed effects and wine characteristics: wine varietal, country of origin, vintage, and name. Column (1) reports results for equation 1.3, column (2) shows estimates for equation 1.4, and column (3) contains results for equation 1.5. We test the joint hypothesis that all the interaction terms are equal to 0. The F-statistic and associated p-value are reported at the bottom of the table.

Appendix

1.A Additional Figures and Tables

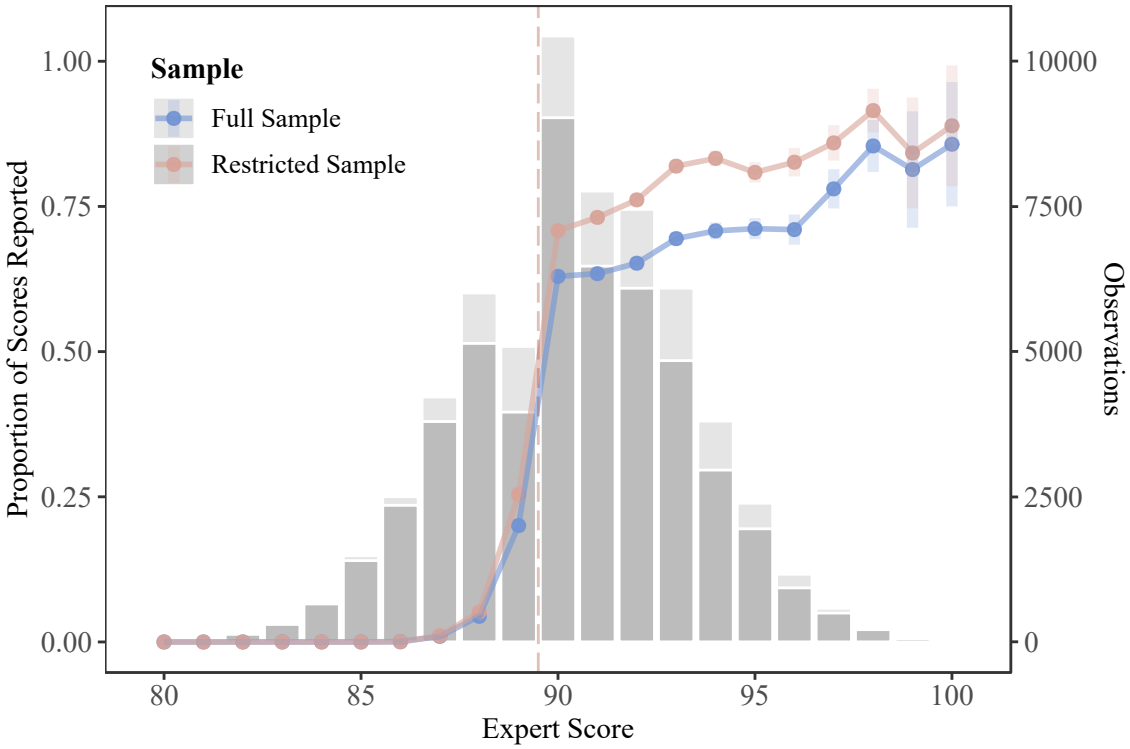
1.A.1 Additional Figures

Figure 1.A.1. Expert Scores Distribution by Price



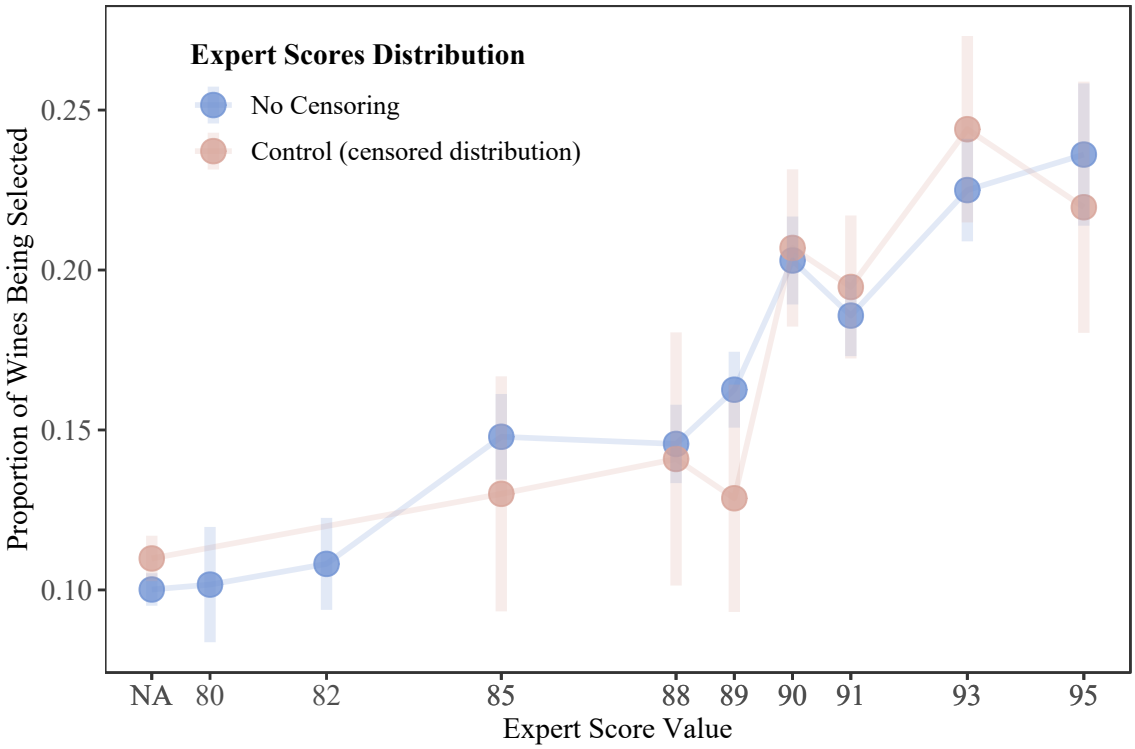
Note: The figure shows a histogram of all the available scores for wines below 250 dollars. Absolute frequency is reported on the primary axis. The cumulative distribution, in red, is plotted on the secondary axis.

Figure 1.A.2. Expert Scores Reported by the Seller – Sample Comparison



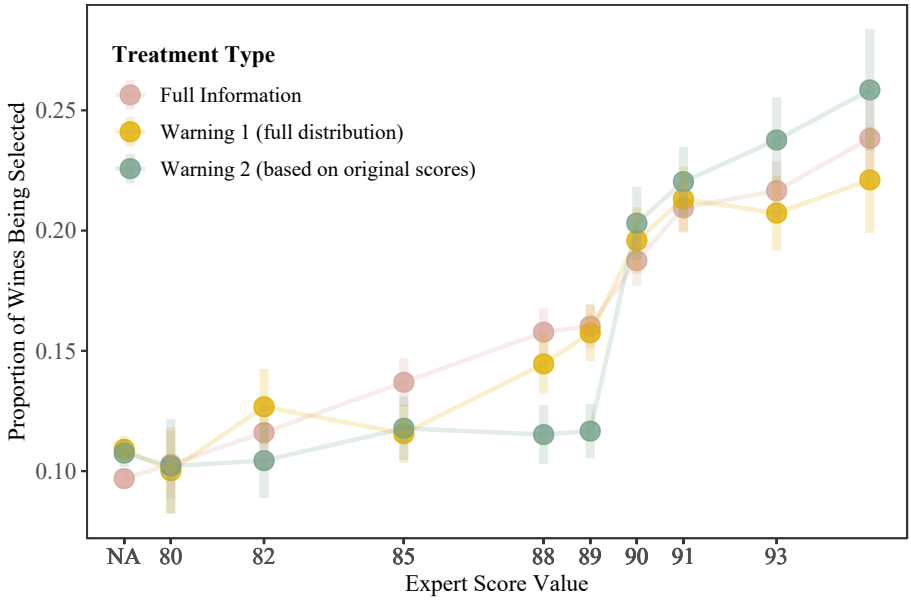
Note: Each point indicates the fraction of expert scores reported by the seller, for all the possible values, with the 95 percent confidence interval for the mean (left axis). The blue dots aggregate values over the full sample of expert sources, the red dots include only expert sources with an average reporting rate above 30 percent. On the right axis, gray bars indicate the number of scores available.

Figure 1.A.3. Expert Scores and Censorship



Note: The figure shows the collapsed mean of a dummy variable equal to 1 if a wine was selected, over maximum values of the expert scores, for the the control and No Censoring groups.

Figure 1.A.4. Wine Selection by Treatment Type – Original Scores



Note: The figure shows the collapsed mean of a dummy variable equal to 1 if a wine was selected, over maximum values of the expert scores, by treatment assignment. For the Warning 2 treatment group, we use the original maximum score available for each wine, independently on whether it was censored or not. The algorithm used to censor scores in the Warning 2 scenario is described in Section 1.4.2.

1.A.2 Additional Tables

Table 1.A.1
Description of Popular Expert Rating Scales

Score	Desciprition
Wine Spectator	
95-100	Classic; a great wine
90-94	Outstanding; superior character and style
80-89	Good to very good; wine with special qualities
70-79	Average; drinkable wine that may have minor flaws
60-69	Below average; drinkable but not recommended
50-59	Poor; undrinkable, not recommended
The Wine Advocate	
96-100	An extraordinary wine of profound and complex character displaying all the attributes expected of a classic wine of its variety. Wines of this caliber are worth a special effort to find, purchase and consume.
90-95	An outstanding wine of exceptional complexity and character. In short, these are terrific wines.
80-89	A barely above average to very good wine displaying various degrees of finesse and flavor as well as character with no noticeable flaws.
70-79	An average wine with little distinction except that it is a soundly made. In essence, a straightforward, innocuous wine.
60-69	A below average wine containing noticeable deficiencies, such as excessive acidity and/or tannin, an absence of flavor or possibly dirty aromas or flavors.
50-59	A wine deemed to be unacceptable.
Wine Enthusiast	
98-100	Classic. The pinnacle of quality
94-97	Superb. A great achievement
90-93	Excellent. Highly recommended
87-89	Very Good. Often good value; well recommended
83-86	Good. Suitable for everyday consumption; often good value
80-82	Acceptable. Can be employed in casual, less-critical circumstances

Sources. Wine Spectator scale is available here. The Wine Advocate scale is available here. The Wine Enthusiast scale is available here.
Note. Wine Enthusiast does not report scores below 80 points.

Table 1.A.2
Expert Ratings – Selected Expert Sources

	(1)	(2)	(3)
Expert Scores (baseline: score is below 89 points)			
89 points	0.231**	0.209***	0.0870
	(0.0584)	(0.0385)	(0.0410)
90 points or more	0.741***	0.628***	0.676***
	(0.0782)	(0.0238)	(0.0287)
Price ($100)		0.000909	-0.00108
		(0.0282)	(0.0279)
Scores is Highest Available			0.0419**
			(0.0127)
Highest Score * 89 points			0.184***
			(0.00533)
Highest Score * 90 points or more			-0.0733***
			(0.0149)
Mean of Dependent Variable for baseline	0.0222		
Observations	50,884	50,409	50,409
R^2	0.447	0.522	0.527
Controls	-	X	X
Expert FE	-	X	X
Vintage FE	-	X	X

Notes: Standard errors clustered at the expert source level are reported in parenthesis. Significance level: ***<0.01, **<0.05, *<0.1. Dependent variable is a dummy that takes a value of 1 if the score is reported by the seller. Expert rating is reported as a categorical variable, base level is a score below 89 points. Controls are added in columns (2) and (3) and include price, wine type dummy, varietal dummy, and country of origin dummy in addition to expert and vintage fixed effects. In column (3) we allow for heterogeneity based on on whether a score is the highest available for a given wine. The sample is restricted to expert sources from which the seller reports at least 30 percent of the scores available on the independent rating aggregator.

Table 1.A.3
No Censoring and Control Comparison – Regression Results

	(1)	(2)	
Expert Scores		*Interaction Terms*	
80 points	-0.00127	-	-
	(0.00926)	-	-
82 points	-0.000531	-	-
	(0.00827)	-	-
85 points	0.0278***	85 ponts * Censored Sample	-0.0251
	(0.00868)		(0.0195)
88 points	0.0201**	88 ponts * Censored Sample	-0.0153
	(0.00841)		(0.0215)
89 points	0.0334***	89 ponts * Censored Sample	-0.0280
	(0.00901)		(0.0185)
90 points	0.0679***	90 ponts * Censored Sample	0.000412
	(0.00913)		(0.0163)
91 points	0.0495***	91 ponts * Censored Sample	0.00925
	(0.0100)		(0.0147)
93 points	0.0815***	93 ponts * Censored Sample	0.0174
	(0.0106)		(0.0190)
95 points	0.0953***	95 ponts * Censored Sample	-0.0204
	(0.0130)		(0.0257)
Censored Sample Dummy	0.0141***		
	(0.00466)		
F-Stat Interaction Terms	1.066		
P-value for the F-test	0.384		
Observations	48,750		
R^2	0.112		

Notes: Standard errors clustered at the individual level are reported in parentheses. Significance level: ***<0.01, **<0.05, *<0.1. Table reports coefficients for the estimation of equation 1.2 with a dummy equal to 1 if the ratings come from the control group. Expert scores in the control group omit values below 85, and 75 percent of values between 85 and 89. Dependent variable is a dummy that takes a value of 1 if a wine is selected. Expert scores is represented by the maximum value available for each wine, and is reported as a categorical variable (baseline is no score available). The regression includes market fixed effects. The coefficients of control variables are omitted and include: crowd-based scores, wine varietal, country of origin, vintage, and name. We test the joint hypothesis that all the interaction terms are equal to 0. The F-statistic and associated p-value are reported at the bottom of the table.

Table 1.A.4
Impact of Expert Scores – Heterogeneity

	(1)	(2)
Price	-0.00724***	
	(0.00180)	
Number of Expert Scores	0.0193***	
	(0.00507)	
Score is highest in the market	0.0704	
	(0.0572)	
Expert Scores (baseline: no score)		*Interaction Terms*
80 points	0.000462	-0.0900
	(0.0132)	(0.102)
82 points	-0.00413	-
	(0.0117)	-
85 points	0.0301**	-0.0287
	(0.0117)	(0.0691)
88 points	0.0255**	-0.0517
	(0.0114)	(0.0628)
89 points	0.0344***	-0.0317
	(0.0120)	(0.0597)
90 points	0.0819***	-0.0467
	(0.0132)	(0.0602)
91 points	0.0640***	-0.0536
	(0.0147)	(0.0584)
93 points	0.0859***	-0.0361
	(0.0245)	(0.0640)
95 points	0.0450	-
	(0.0594)	-
Crowd Based Scores (baseline: 3.5)		
Avg. Score below 3.5	-0.0537***	
	(0.00556)	
Avg. Score above 3.5	0.113***	
	(0.00769)	
Crowd Sample Size	0.00440***	
	(0.000548)	
F-stat Interaction Terms	0.295	
P-value for the F-test	0.956	
Observations	48,750	
R^2	0.194	

Notes: Standard errors clustered at the individual level are reported in parentheses. Significance level: ***<0.01, **<0.05, *<0.1. Dependent variable is a dummy that takes a value of one if the wine is selected. Expert scores are represented by the maximum value available for each wine, and are reported as a categorical variable (baseline is no score available). The interaction term for a score of 92 is omitted because of multicollinearity. A maximum score of 95 is always the highest level in the market. The average crowd-based score is treated as categorical variable (baseline score equal to 3.5). The variable labeled "Score is highest in the market" takes a value of one if the score is the highest available among the alternative wines. Interaction terms with expert score values are reported in column (2). The specification includes market fixed effects to control for the quality of alternative wines in addition to wine specific controls: wine varietal, country of origin, vintage, and name. We test the joint hypothesis that all the interaction terms are equal to 0. The F-statistic and associated p-value are reported at the bottom of the table.

Table 1.A.5
Expert Scores and Treatment Assignment

	(1) No Censoring vs Warning 1	
Price	-0.00710***	
	(0.00103)	
Number of Expert Scores	0.0160***	
	(0.00300)	
Treatment Dummy	0.00548	
	(0.00461)	
Expert Scores	*Baseline Coeff.*	*Interaction Terms*
80 points	-0.00284	-0.0132
	(0.00909)	(0.0124)
82 points	-0.00259	0.00516
	(0.00800)	(0.0108)
85 points	0.0281***	-0.0299***
	(0.00844)	(0.0103)
88 points	0.0202**	-0.00707
	(0.00786)	(0.0100)
89 points	0.0329***	-0.0105
	(0.00855)	(0.0106)
90 points	0.0659***	-0.0124
	(0.00872)	(0.0119)
91 points	0.0495***	0.0171
	(0.00942)	(0.0117)
93 points	0.0812***	-0.0218
	(0.0103)	(0.0133)
95 points	0.0952***	-0.0209
	(0.0126)	(0.0169)
Crowd Based Scores (baseline: 3.5)		
Avg. Score below 3.5	-0.0418***	
	(0.00348)	
Avg. Score above 3.5	0.0941***	
	(0.00559)	
Crowd Sample Size (in 100s)	0.00422***	
	(0.000397)	
F-stat Interaction Terms	2.537	
P-value for the F-test	0.00728	
Observations	72,750	
R^2	0.106	

Notes: Standard errors clustered at the individual level are reported in parentheses. Significance level: ***<0.01, **<0.05, *<0.1. Dependent variable is a dummy that takes a value of 1 if the wine is selected. Expert scores is represented by the maximum value available for each wine, and is reported as a categorical variable (baseline is no score available). The average crowd-based score is treated as categorical variable (baseline score equal to 3.5). The regression includes market fixed effects and wine characteristics: wine varietal, country of origin, vintage, and name. We test the joint hypothesis that all the interaction terms are equal to 0. The F-statistic and associated p-value are reported at the bottom of the table.

Table 1.A.6
Expert Scores and Treatment Assignment

	(1) No Censoring vs Full Information		(2) Control vs Warning 2		(3) Warning 2 vs Full Information	
Price	-0.00701***		-0.00529***		-0.00683***	
	(0.000925)		(0.00145)		(0.000941)	
Number of Expert Scores	0.0183***		0.0372***		0.0212***	
	(0.00241)		(0.00737)		(0.00285)	
Treatment Dummy	-0.00294		-0.00547		-0.0114***	
	(0.00351)		(0.00406)		(0.00392)	
Expert Scores	*Baseline Coeff.*	*Interaction Terms*	*Baseline Coeff.*	*Interaction Terms*	*Baseline Coeff.*	*Interaction Terms*
Below 90 points	0.0175***	0.00234	-0.0205	0.0117	0.00839	0.00482
	(0.00509)	(0.00558)	(0.0138)	(0.0136)	(0.00851)	(0.00702)
Above or equal 90 points	0.0638***	0.00482	0.0465***	0.0139	0.0805***	-0.0172*
	(0.00690)	(0.00702)	(0.0132)	(0.0115)	(0.00887)	(0.0100)
Crowd Based Scores (baseline: 3.5)						
Avg. Score below 3.5	-0.0469***		-0.0399***		-0.0469***	
	(0.00294)		(0.00508)		(0.00313)	
Avg. Score above 3.5	0.0896***		0.0867***		0.0895***	
	(0.00478)		(0.00679)		(0.00471)	
Crowd Sample Size (in 100s)	0.00497***		0.00462***		0.00553***	
	(0.000349)		(0.000547)		(0.000372)	
F-stat Interaction Terms	0.242		0.867		2.744	
P-value for the F-test	0.785		0.422		0.0649	
Observations	97,575		42,450		91,275	
R^2	0.104		0.104		0.102	

Notes: Standard errors clustered at the individual level are reported in parentheses. Significance level: ***<0.01, **<0.05, *<0.1. Dependent variable is a dummy that takes a value of 1 if the wine is selected. Expert scores are represented by the maximum value available for each wine, and are pooled in a categorical variable that takes three values: no score available (baseline), maximum score below 90 points, maximum score equal or above 90 points. The average crowd-based score is treated as a categorical variable (baseline score equal to 3.5). All specifications include market fixed effects and wine characteristics: wine varietal, country of origin, vintage, and name. Column (1) reports results for equation 1.3, column (2) shows estimates for equation 1.4, and column (3) contains results for equation 1.5. We test the joint hypothesis that all the interaction terms are equal to 0. The F-statistic and associated p-value are reported at the bottom of the table.

1.B Wine Ratings Data

1.B.1 Dataset Construction

The wine ratings data collection was performed in three steps, using the seller's website as the starting point. First, we collected all wines available from the seller's virtual shelf, including wines listed as unavailable, obtaining 340,540 observations. Second, we searched each of those wine strings in the comprehensive database, pairing with the original product only if a unique result was returned. Multiple results suggested that one or more wine attributes were not identified correctly (e.g. wrong vintage). We were able to match 29.54 percent of the original observations. Finally, we performed a search on the Vivino app. The subset of wines for which all three sources are available is 66,434, roughly 20 percent of the original sample. Appendix Table 1.B.7 reports summary statistics of the collection process.

Table 1.B.7
Data Collection Results

Description	N	Percent
Wine from Seller's Website	340,540	
Wines matched to Scores Dataset	100,581	29.54%
Wines matched to Vivino	66,434	19.51%

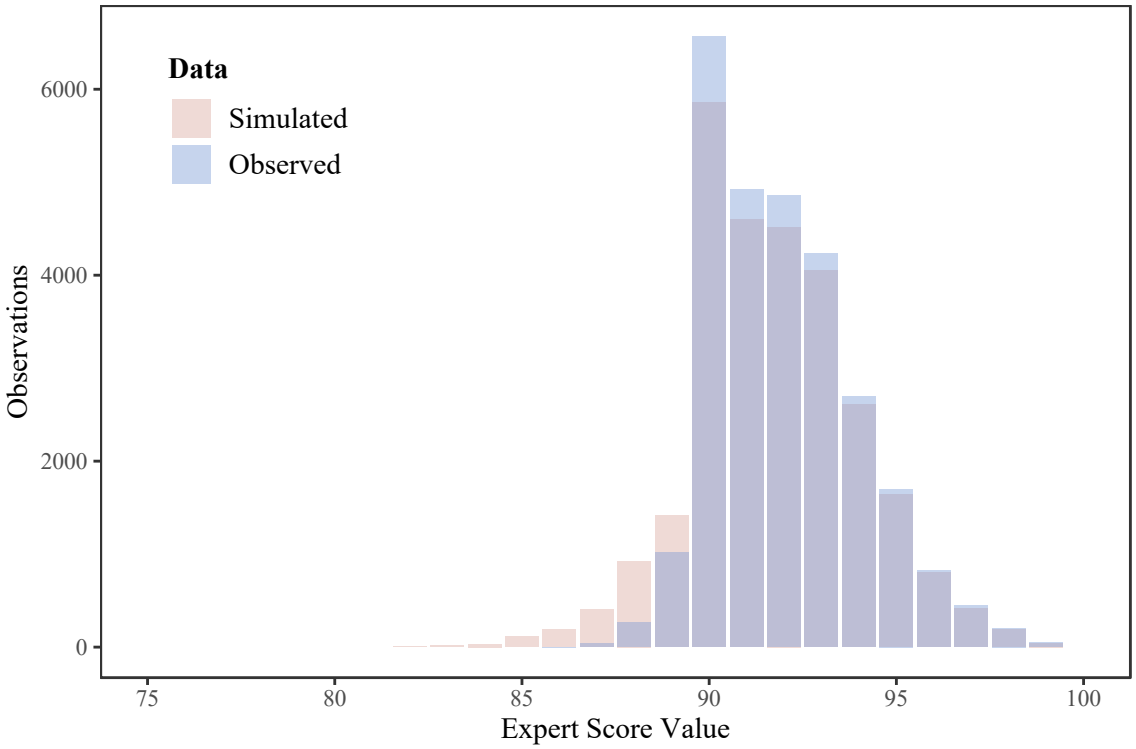
1.B.2 Simulation Exercise

With these simulations we investigate what scores are reported when no censorship is involved, to rule-out that the results we observe could be generated by chance.

Seller Choosing Randomly. For each wine w, let N_w be the number of expert ratings reported from the seller. We keep only wines with $N_w > 0$. We then draw (without replacement) N_w ratings at random from the comprehensive database.

Appendix Figure 1.B.5 plots the two distributions, showing that the observed data is clearly skewed to the right with respect to the simulated distribution.

Figure 1.B.5. Simulating a Seller that Choose Random Scores

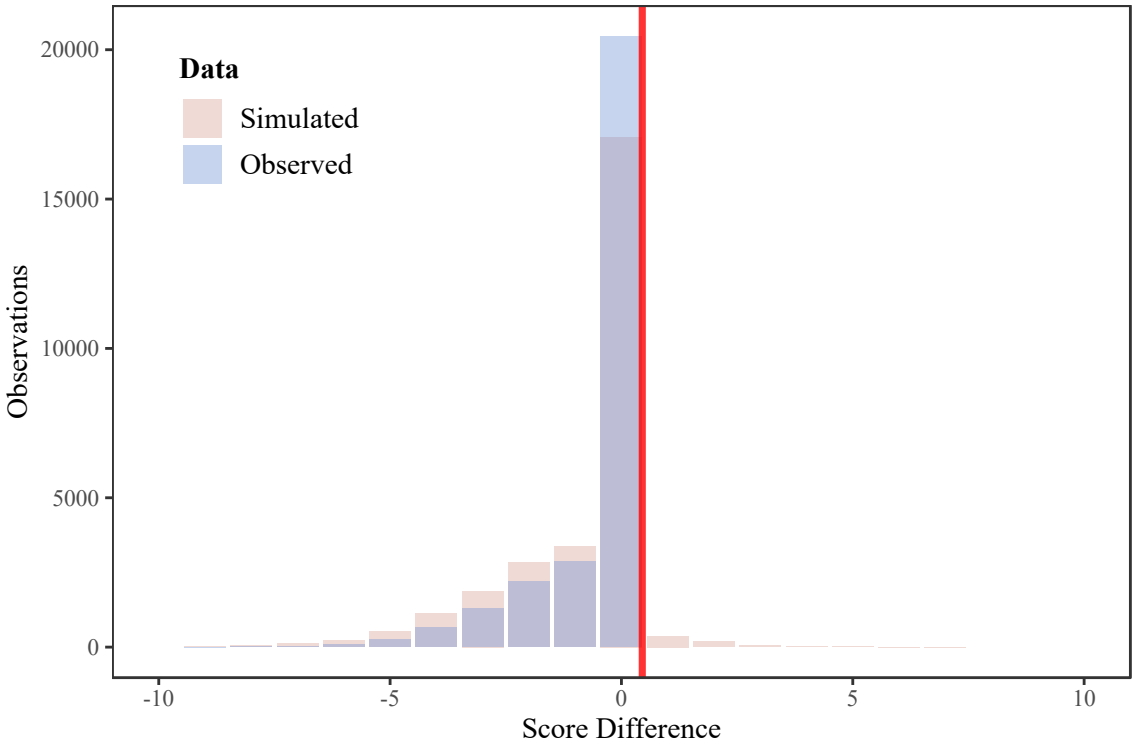


Notes: The blue bars display the score distribution observed on the seller's website. The red bars are generated by simulating a seller that chooses at random from the full sample.

Seller Reporting Maximum Scores. For each wine w, let M_w be the highest rating reported by the seller. For each expert r of wine w on the comprehensive database, let S_{rw} be the score value. Compute the difference $(S_{rw} - M_w)$, i.e. the gap between the score of expert r and the highest rating reported. If the seller only reports the best rating, the histogram $(S_{rw} - M_w)$ is right-truncated at 0 across all wines and raters.

Comparing the actual and simulated distribution in Appendix Figure 1.B.6, we can clearly observe that sellers are very likely to report only the maximum score available, compared to a simulated seller that chooses randomly from the available scores.

Figure 1.B.6. Testing if the Seller Discloses Maximum Values



Notes: The blue bars display the score distribution observed on the seller's website. The red bars are generated by simulating a seller that chooses at random from the full sample.

1.C Survey Data Description

This section contains a detailed description of the survey design.

1.C.1 Survey Questionnaire

Consent Form

You are invited to take part in a Brown University research study. Your participation is voluntary.

- RESEARCHER: Igor Cerasa, MSc (xxx)xxx-xxxx with faculty advisor John Friedman, PhD (xxx)xxx-xxxx
- PURPOSE: This study is about learning on people's attitude towards wine purchase decisions. You are being asked to be in this study because you are above the age of 21, reside in the United States of America, and purchase wine at least once per month from either on-trade or off-trade businesses.
- PROCEDURES: You will be asked to complete a brief online survey. To complete it, you will need an internet connection and a suitable desktop or mobile device. The survey will ask question about your backgrounds, and will ask you to choose wines in a virtual retail environment.
- TIME INVOLVED: The study will take about 25 minutes of your time.
- COMPENSATION: You will receive compensation from *Respondi* for your time. You will not receive money directly from Brown University or the research team. Compensation received as a client of *Respondi* is not associated with Brown University Research.
- RISKS: Some of the questions may make you feel uncomfortable. You may skip any questions you prefer not to answer or do not feel comfortable answering. This survey can be stopped at any time.
- BENEFITS: There are no direct benefits to you if you agree to be in this research study.
- CONFIDENTIALITY: To maintain confidentiality, we will assign all your data a numerical code. Your responses will not be connected to your identity. We will not collect your personal data, but only your anonymous responses. Your anonymous responses may be used for scientific purposes or shared, with other researchers for future research, teaching, or publication. In addition, any identifiable information that *Respondi* may have collected about you will not be shared with the research team. Likewise, none of the data we collect will be shared with *Respondi*. In order to protect your own privacy, we strongly recommend that you complete this survey in a private location and closing your browser window when complete.

- VOLUNTARY: You do not have to be in this study if you do not want to be. Even if you decide to be in this study, you can change your mind and stop at any time.
- CONTACT INFORMATION: If you have any questions about your participation in this study, you can call Igor Cerasa, MSc at (xxx)xxx-xxxx, or email igor_cerasa@brown.edu. You can also contact his advisor John Friedman, PhD at (xxx)xxx-xxxx or email john_friedman@brown.edu.
- YOUR RIGHTS: If you have questions about your rights as a research participant, you can contact Brown University's Human Research Protection Program at (xxx)xxx-xxxx or email them at IRB@Brown.edu.
- CONSENT TO PARTICIPATE: Clicking the next button confirms that you have read and understood the information in this document, are 21 years old or older and that you agree to volunteer as a research participant for this study. You can print a copy of this form.

Screening Questions

Please answer the following questions:

1. What type of device are you using?
   [Phone/Tablet/Mobile Device; Laptop/Desktop; Other]
2. What is your age?
   [Numerical Input]
3. What is your country of residence?
   [France; Germany; UK; Unites States of America; None of the Above]
4. How often do you perform sport activities?
   [Less than once per week; 1-2 times per week; 3 or more times per week]
5. How much wine have you purchased within the past 30 days?
   [None; 1-2 bottle; 3 or more bottles]
6. How often do you use food delivery apps (UberEats, GrubHub, etc.)?
   [Less than once per week; 1-2 times per week; 3 or more times per week]

Background Questions

1. How do you identify yourself?
   [Fluid/Genderfluid; Gender Non-Conforming/Genderqueer; Man; Non-Binary; Transgender Man/Trans Man; Transgender Woman/Trans Woman; Woman; Not Listed/Prefer to self describe; Prefer not to answer]
2. Self-described gender (if you selected "Prefer to self-describe" in the previous question)
   [String Input]

3. What is the highest school degree you earned?
   [High-school or less; Some college credits; Bachelor's degree; Master's degree or more; Prefer not to answer]

4. In which ZIP code do you live?
   [Numerical Input]

5. What is your best estimate of your annual household income?
   [Less than 20,000; $20,001 - $50,000; $50,001 - $100,000; More than $100,000; Prefer not to answer]

6. What is your occupation?
   [Employed; Self-Employed; Unemployed; Retired; Student; Other; Prefer not to answer]

7. How often do you drink wine?
   [Regularly (more than twice per week); Occasionally (more than once per month); Rarely (less than once per month); Prefer not to answer]

8. How often do you purchase wine?
   [Less than 1 bottle per week; 1-2 bottles per week; 3 or more bottles per week; Prefer not to answer]

9. When making a decision on which wine to buy, what characteristics do you usually consider? (select all that apply)
   [Alcohol Level; Brand; Country; Label; Medals; Organic Wines; Price; Ratings; Region (Appelation); Varietal; Vintage]

10. Where do you usually buy wine? (select all that apply)
    [Liquor/Wine Store; Online Store; Supermarket; Wine Club; Other]

11. How often do you consult independent reviews (magazines, wine apps, blogs) before purchasing a wine?
    [Never; Sometimes; Often; Always; Prefer not to answer]

Completion Page

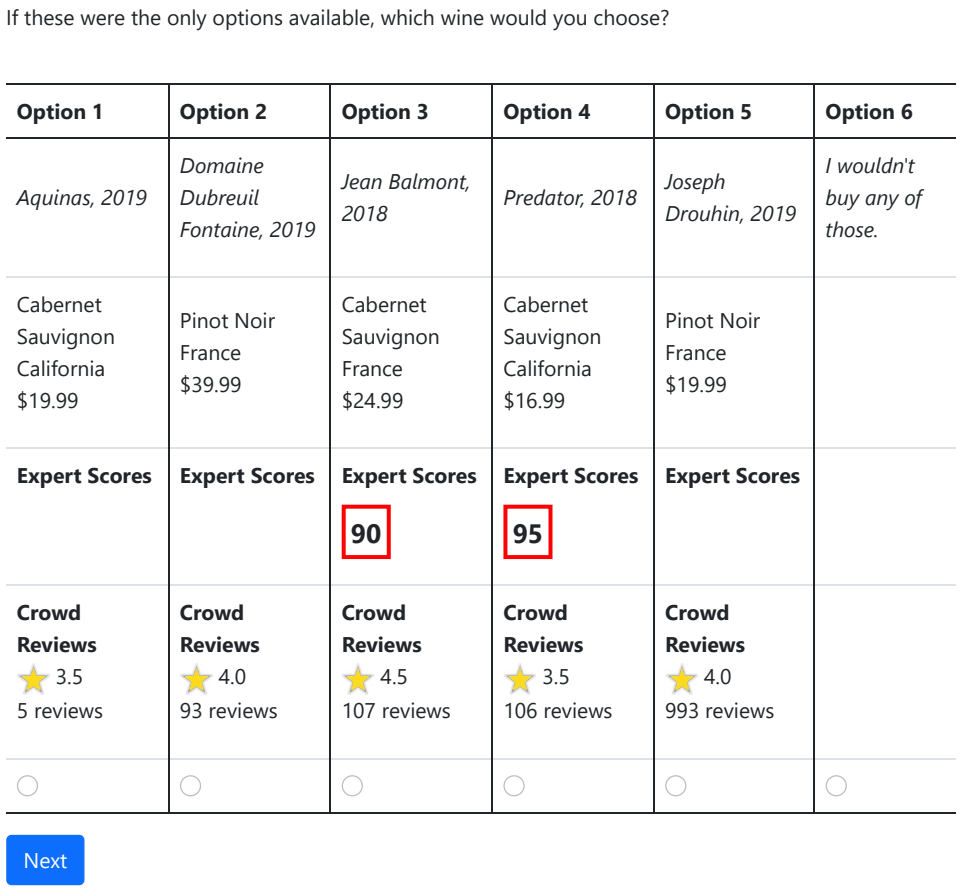
1. It is crucial for the quality of our research that we only include responses from people who devoted their full attention to this study. Otherwise, years of effort could be wasted. The answer to this question will not affect your compensation, however, please indicate how much effort you put forth in completing this survey.
   [No or almost no effort; Little Effort; Some Effort; A lot of effort; Prefer not to answer]

2. Please add any comment or suggestion you would like to make in the text-box.
   [String Input]

1.C.2 Discrete Choice Exercise

Figure 1.C.7 provides an example for the discrete choice page.

Figure 1.C.7. Discrete Choice Interface



If these were the only options available, which wine would you choose?

Option 1	Option 2	Option 3	Option 4	Option 5	Option 6
Aquinas, 2019	*Domaine Dubreuil Fontaine, 2019*	*Jean Balmont, 2018*	*Predator, 2018*	*Joseph Drouhin, 2019*	*I wouldn't buy any of those.*
Cabernet Sauvignon California $19.99	Pinot Noir France $39.99	Cabernet Sauvignon France $24.99	Cabernet Sauvignon California $16.99	Pinot Noir France $19.99	
Expert Scores	**Expert Scores**	**Expert Scores** 90	**Expert Scores** 95	**Expert Scores**	
Crowd Reviews ★ 3.5 5 reviews	**Crowd Reviews** ★ 4.0 93 reviews	**Crowd Reviews** ★ 4.5 107 reviews	**Crowd Reviews** ★ 3.5 106 reviews	**Crowd Reviews** ★ 4.0 993 reviews	
○	○	○	○	○	○

Next

Notes: The figure displays a screenshot of the virtual market in which individuals perform the discrete choice experiment.

1.C.3 Definition of Wine Attributes

In this section we describe the procedure used to generate the wines available on the virtual shelf.

Attributes are divided in two blocks: wine characteristics, and wine reviews.

Wine Characteristics. We obtained wine characteristics primarily from the scraped wine seller data, which include name, year (vintage), color, varietal, origin, price. We selected 200 wines, and for each of them we maintain this set of attributes constant throughout the experiment. The algorithm used to make the selection followed several steps:

1. We restricted the wine seller data as follows:
    - Vintage between 2016 and 2020
    - Origin is either France or California
    - Color is either red or white.
    - Variety is either Cabernet Sauvignon (red), Pinot Noir (red), Chardonnay (white), or Sauvignon Blanc (white)
    - Price is between 10.00 and 40.00 dollars.
2. We randomly selected 25 wines for each $variety \times origin$ category.
3. We randomly assigned years to be either 2018, or 2019, with equal proportions by variety.
4. We removed from the wine name any indication that may signal quality (e.g. "Gold Collection", "Special Selection", etc.).

Wine Reviews. To maximize the variation of wine reviews, we drew them randomly, in a way that resembles the distribution of the complete dataset. Table 1.C.8 shows the value sets and the probability weights used. In addition, the sample size of crowd-based reviews was perturbed by adding some noise, drawn uniformly from the set of integers in the interval $[-11, 10]$.

Table 1.C.8
Wine Reviews

Attribute	Values Set	Probability Weights
Expert Scores Number	$\{0, 1, 2, 3\}$	$[4, 4, 2, 1]$
Expert Names	$\{WS, JS, WE, RP, VI, DE, JD\}$	$[1, 1, 1, 1, 1, 1, 1]$
Expert Values	$\{80, 82, 85, 88, 89, 90, 91, 93, 95\}$	$[2, 3, 4, 4, 4, 3, 3, 2, 1]$
Crowd Based Score	$\{2.5, 3, 3.5, 4, 4.5\}$	$[1, 1, 1, 1, 1]$
Crowd Based Size	$\{15, 100, 1000\}$	$[1, 1, 1]$

Markets. Each market was obtained in three steps:

1. We randomly drew a wine color, with equal probabilities.
2. We randomly drew five wines in the chosen color category.
3. We independently assigned wine reviews to each of the five wines.

1.C.4 Survey Statistics and Quality Checks

Table 1.C.9 contains a summary of the survey completion codes. Table 1.C.10 shows a balance table of the respondents demographics.

Table 1.C.9
Survey Completion Summary

Survey Outcome	Frequency	Percent	Cum. Percentage
Completed	1,173	43	43
Screened out	697	26	69
Quality fail	470	17	86
Incomplete	389	14	100
Total	2,729	100	100

Table 1.C.10
Respondent Demographics by Treatment Assignment

	Treatment Assignment											
	No Censoring - Full Info		No Censoring - Warning1		Control - Warning2		Full Information		Warning 1		Warning 2	
Observations	314		177		159		248		151		124	
Female Share	0.61	(0.49)	0.51	(0.50)	0.47	(0.50)	0.59	(0.49)	0.56	(0.50)	0.61	(0.49)
Age	56.49	(14.92)	57.16	(15.48)	57.77	(15.39)	56.75	(14.85)	56.19	(14.63)	57.39	(14.70)
College Degree	0.64	(0.48)	0.58	(0.50)	0.58	(0.49)	0.62	(0.49)	0.68	(0.47)	0.60	(0.49)
Household Income												
Income below 20k$	0.05	(0.21)	0.05	(0.22)	0.08	(0.27)	0.04	(0.20)	0.04	(0.20)	0.06	(0.23)
Income between 20k$ and 50k$	0.22	(0.41)	0.24	(0.43)	0.25	(0.44)	0.25	(0.43)	0.25	(0.43)	0.12	(0.33)
Income between 50k$ and 100k$	0.41	(0.49)	0.40	(0.49)	0.34	(0.48)	0.38	(0.49)	0.39	(0.49)	0.50	(0.50)
Income above 100k$	0.29	(0.45)	0.30	(0.46)	0.29	(0.45)	0.30	(0.46)	0.30	(0.46)	0.30	(0.46)
Employment Status												
Employed	0.40	(0.49)	0.45	(0.50)	0.42	(0.50)	0.42	(0.49)	0.49	(0.50)	0.50	(0.50)
Self-Employed	0.10	(0.30)	0.05	(0.22)	0.10	(0.30)	0.09	(0.29)	0.09	(0.28)	0.09	(0.29)
Unemployed	0.06	(0.23)	0.05	(0.22)	0.06	(0.23)	0.05	(0.22)	0.05	(0.21)	0.05	(0.22)
Retired	0.38	(0.49)	0.41	(0.49)	0.36	(0.48)	0.38	(0.49)	0.35	(0.48)	0.34	(0.48)
Student	0.00	(0.06)	0.02	(0.13)	0.01	(0.08)	0.01	(0.11)	0.00	(0.00)	0.01	(0.09)
Other	0.04	(0.21)	0.01	(0.11)	0.04	(0.21)	0.05	(0.22)	0.02	(0.14)	0.02	(0.13)
Wine Consumption												
Regularly (more than twice per week)	0.54	(0.50)	0.55	(0.50)	0.57	(0.50)	0.57	(0.50)	0.58	(0.50)	0.60	(0.49)
Occasionally (more than once per month)	0.44	(0.50)	0.44	(0.50)	0.40	(0.49)	0.40	(0.49)	0.39	(0.49)	0.36	(0.48)
Rarely (less than once per month)	0.02	(0.15)	0.02	(0.13)	0.03	(0.16)	0.03	(0.17)	0.03	(0.16)	0.02	(0.15)
Wine Purchase												
Less than one bottle per week	0.55	(0.50)	0.51	(0.50)	0.48	(0.50)	0.48	(0.50)	0.50	(0.50)	0.48	(0.50)
1-2 bottles per week	0.33	(0.47)	0.34	(0.47)	0.37	(0.48)	0.38	(0.49)	0.36	(0.48)	0.33	(0.47)
3 or more bottles per week	0.10	(0.30)	0.13	(0.34)	0.13	(0.34)	0.14	(0.34)	0.14	(0.35)	0.18	(0.38)
Wine App Utilization												
Never	0.44	(0.50)	0.44	(0.50)	0.45	(0.50)	0.44	(0.50)	0.36	(0.48)	0.40	(0.49)
Sometimes	0.44	(0.50)	0.47	(0.50)	0.45	(0.50)	0.44	(0.50)	0.48	(0.50)	0.48	(0.50)
Often	0.07	(0.26)	0.07	(0.26)	0.07	(0.25)	0.08	(0.27)	0.13	(0.33)	0.10	(0.31)
Always	0.04	(0.21)	0.02	(0.15)	0.03	(0.18)	0.03	(0.17)	0.02	(0.14)	0.01	(0.09)
Effort in Survey Completion												
No or almost no effort	0.03	(0.18)	0.01	(0.11)	0.01	(0.08)	0.01	(0.09)	0.03	(0.18)	0.04	(0.20)
Little Effort	0.04	(0.18)	0.01	(0.08)	0.01	(0.11)	0.02	(0.14)	0.03	(0.16)	0.02	(0.13)
Some Effort	0.22	(0.41)	0.22	(0.42)	0.20	(0.40)	0.17	(0.38)	0.22	(0.41)	0.17	(0.38)
A lot of effort	0.70	(0.46)	0.76	(0.43)	0.77	(0.42)	0.80	(0.40)	0.72	(0.45)	0.77	(0.43)

Notes: Table contains the mean and standard deviation (in parentheses) of the demographic variables collected during the survey, by treatment assignment.

Chapter 2

Your Life, Your Money, Or Your Vote? Traffic Enforcement, Road Safety and Local Budgets

2.1 Introduction

Injuries related to road accidents are the world's leading cause of death for children and young adults, and the eighth overall cause of death (World Health Organization, 2018). Casualties are particularly severe in low and middle income countries, where the use of motorized vehicles is increasing over time without adequate traffic regulation. This issue is also critical in advanced countries: the National Highway Traffic Safety Administration (NHTSA) reports that in 2010 the total cost of motor vehicle crashes in the United States was $242 billion, and as high as $836 billion when adjusting for quality-of-life valuations (Blincoe et al., 2015).

Among existing remedies, safety measures and traffic enforcement are important tools available to policy makers as a way to mitigate this problem. A well established literature details the effectiveness of various safety measures including the use of seat belts (Cohen and Einav, 2003), implementation of point-system driving licenses (Bourgeon and Picard, 2007; De Paola, Scoppa and Falcone, 2013), texting bans (Abouk and Adams, 2013), and policies on drinking and driving (Levitt and Porter, 2001*b*; Adams and Cotti, 2008). However, despite a vast body of anecdotal evidence and media reports, the causal literature on the relation between traffic enforcement and road safety is rather scarce and scattered. A

few papers have analyzed developing countries (Law, Noland and Evans, 2009), investigating the role of political institutions and corruption (Anbarci, Escaleras and Register, 2006; Albalate and Yarygina, 2017). In the United States, traffic enforcement has been studied in only a handful of states and in very specific contexts, due to the limitations that the heterogeneous regulations impose on across-states comparisons. Moreover, evident endogeneity concerns warrant against a mere cross-sectional comparison to identify causal relations, given that stricter traffic enforcement can easily arise as a reaction to worsening road safety conditions. A first attempt to circumvent the problem, by Garrett and Wagner (2009), relies on panel data from North Carolina counties, and shows that counties tend to increase traffic tickets in years following a decline in revenues, but not the opposite. Using data from Massachusetts, Makowsky and Stratmann (2009) demonstrate that officers are more likely to impose traffic fines in municipalities where voters rejected a referendum to temporarily increase the property tax, although their results seem to be valid only when the driver is from out of town. In a subsequent paper, they show that more fines lead to reduced crashes (Makowsky and Stratmann, 2011). More recently, DeAngelo and Hansen (2014) exploit a mass layoff of Oregon State Police, and find that the subsequent decrease in traffic enforcement was associated with a significant increase in injuries and fatalities.

The aim of this paper is to study the interaction between local government and road safety policies within a homogeneous institutional framework. While politicians seek votes from both safe and reckless drivers, they also prefer a safe traffic environment, therefore generating a tradeoff on at least two levels. First, a strict enforcement can be costly in electoral times, which is in contrast with career concerns. On the other hand, demonstrating a stronger commitment to road safety might be desirable to voters, suggesting an increase –rather than a decrease– in enforcement during elections. Second, tools that are effective in saving lives may not necessarily be the same that raise revenues efficiently and in a profitable way, repaying the fixed costs incurred to introduce them. An ideal policy –under perfect compliance– would reduce accident casualties without im-

posing traffic fines. However, revenues raised through them are a non-trivial component of local budgets, and an important one, given the difficulties that local governments have in financing a deficit. Policy makers may therefore rely on additional fines in order to compensate for unexpected financial shortfalls.

I make use of two separate quasi-natural experiments in Italy to address potential endogeneity problems and provide evidence that both a political and a budget motive play distinct roles in shaping traffic enforcement policies. I show two sets of results: first –consistent with the findings of Garrett and Wagner (2009) and Makowsky and Stratmann (2009)– mayors increase traffic fines to make up for reduced revenues. Second, I present evidence that mayors reduce enforcement in electoral years, which is consistent with the notion that strict enforcement is costly in electoral times.

In the first exercise, I exploit a large fiscal consolidation in Italy after the 2010 sovereign debt crisis which resulted in a substantial reduction of transfers from the central government to local entities (a 15 percent reduction in 2013), placing budgets under strong fiscal pressure. Using a regression discontinuity approach I provide evidence that municipalities reacted by increasing traffic fines in order to maintain the same level of expenditures. Estimates show that affected towns raised revenues from fines by more than 4 euros per capita in 2013, roughly a 35 percent increase, corresponding to approximately one third of the reduction in transfers. Moreover, the results are larger for municipalities that had high local taxes to begin with, making it difficult to increase them further.

Contrary to the previous literature (e.g. Makowsky and Stratmann, 2011) I find very weak evidence that additional traffic fines had a positive impact on road safety. While the estimates are not very precise, I find an imprecisely estimated increase in the injury rate in years 2013 and 2014, although I cannot exclude that this is due to a spurious correlation, mean reversion or a lack of power of this experiment. Still, the lack of significant results is consistent with the idea that, in that circumstance, improving safety conditions was not a major motivation of increasing traffic enforcement.

In the second exercise, I exploit variation in municipal election dates for the period 2002-2013 and provide evidence that there exists an electoral cycle of traffic enforcement. Estimates show that, during electoral years, traffic fines imposed are roughly 60 cents lower, a 5 percent reduction relative to non electoral years, and that actual money collection is likewise less efficient, with more than 6 percent reduction. In addition, reduced form estimates show a significant worsening of road safety, with an increase in both accidents and injured people. Election years are associated with 0.04 more accidents per one thousand inhabitants and 0.05 more injured, a 1.9 and 1.6 percent increase respectively, while no significant effect is observed on the number of casualties. While justifying an exclusion restriction in this context is not an immediate task, I also provide some arguably causal estimates on the impact of traffic fines on road safety. Conservative estimates, robust to weak instrument bias, show that doubling traffic fines (an increase of about ten dollars) would reduce the accident rate by 0.3 per one thousand inhabitants, a 16 percent reduction. Unfortunately, data availability does not allow me to directly investigate what type of fines are used, and so I cannot investigate which enforcement tools are likely to be most effective in reducing accidents.

My paper relates to several strands of literature on the interaction between central and local governments in funding public expenditures and the determinants of traffic fines. To the best of my knowledge, this is the first paper establishing a budget motive in the context of a European country. Related to my work is the study by Marattin, Nannicini and Porcelli (2019), who investigate the same fiscal consolidation and show that a reduction in transfers to local governments resulted in an increase in taxes and fees rather than a reduction in expenditures. In addition, my study of a political motive investigates the role of strategic behavior of politicians in an electoral cycle. As pointed out by Dubois (2016) in his extensive review of the literature, nearly any public policy topic has been analyzed under the lens of a political business cycle, following Nordhaus (1975)'s seminal work on the subject. This work is no exception, and it is closely linked to a previous study by

Bracco (2018), who uses a panel of Italian budget data to show that traffic fines are lower in municipalities during an election campaign, to which my results are comparable. This paper is also linked to a recent study byBertoli and Grembi (2018), who investigate the impact of elections on traffic fines and road safety in two Italian regions: *Lombardia* and *Veneto*. Their analysis benefits from more detailed data on the type of fines, but comes at the cost of reducing geographical variation. In particular, the sample they analyze focuses on two of the richest regions in the North, with large social capital (Nannicini et al., 2013). Not surprisingly, my results show a larger reduction of traffic fines in electoral years, indicating that the electoral cycle is stronger in areas of the country more prone to political clientelism and corruption.

The remainder of this paper is as follows: Section 2.2 describes the institutional background and the relevant policy, in Section 2.3 I introduce the data, Section 2.4 presents the main empirical model, Section 2.5 reports the main results and Section 2.6 concludes.

2.2 Institutional Background

The focus of my analysis are municipalities in Italy. Italy is historically a country characterized by great administrative fragmentation, and is divided into 20 different regions, 107 provinces and metropolitan cities, and 7,904 *comuni*. The municipality (*comune*) is the smallest autonomous unit, and provides essential services including nurseries, childcare, education-related services (such as meal provision and transportation), waste management, social services to elderly and disabled persons, local police and public order, maintenance of local roads and traffic enforcement, town planning, culture, recreation, and economic development. Municipalities manage roughly 6.8% of current public expenditures and raise revenues from different sources. Until 2014, about half of their current expenditures were financed by horizontal equalization grants, which are allocated using a historical expenditure system, and the remaining expenditures were financed through a system of local taxes and fees. A residual source of funding exists in the form of specific

earmarked grants that target municipality-specific needs. Among the broad category of fees is a unique subcategory which includes traffic fines. Public safety within the borders of the municipality is mainly provided by the municipal police (*polizia municipale*), who are responsible for the enforcement of traffic laws and public order. Additional support, mostly in larger cities, is entrusted to national police bodies, and revenues from fines are collected by the central government, therefore they do not appear in my data. Revenues collected from fines contribute to the total budget, although a provision exists which requires half of these revenues be devoted to specific goals (e.g. road maintenance or improvement). This could provide an alternative channel that may explain why an increase in enforcement may improve road safety.

According to the Italian Constitution, all local governments are subject to a balanced-budget rule and a "Domestic Stability Pact" (DSP), which imposes a ceiling on the size of fiscal gap. Towns below 5,000 inhabitants were exempted from the DSP until 2013. This policy – time-invariant for most of the period I study– may present as an additional confounding element in the analysis of the 2014 budget, hence it will be discussed to a greater extent in Section 2.4. While some regions enjoy a special type of legislative autonomy, I will center my attention the the normal-statute ones, the majority of which are regulated by the national law and account for 83 percent of the municipalities, or 85 percent of the total population. Administrative fragmentation is paired with town size heterogeneity, with an average population of 7,345 inhabitants according to the latest census (ISTAT, 2011), and about half of the population living in municipalities with less than 20,000 inhabitants. Most of the municipalities, about 85 percent, have a population smaller than 10,000 inhabitants, and only six have more than 500,000 inhabitants. Population is an important characteristics as several policies change discontinuously at given thresholds, such as the size of elective bodies, the aforementioned DSP and a number of policies concerning local governments that I will outline in Section 2.2.2. This feature has been widely exploited in the literature (see Eggers et al., 2018) and will be the main source of identification to

understand the impact of the fiscal consolidation.

2.2.1 Fiscal Consolidation

The main identification strategy to investigate a budget motive comes from a substantial fiscal consolidation that the central government undertook in the 2010-2015 period, after the spread of a sovereign debt crisis in several European countries. A series of contractionary measures were taken to slow down the increase in debt-to-GDP ratio, calling for a deficit reduction from 5.3% of the GDP in 2009 to 2.1% in 2018. This was done through a series of interventions, sometimes more than once per year. About one third of the reduction in expenditures (€8.6 billion out of the total €25.1 billion) was obtained by cutting transfers to municipal governments.[1] This policy, combined with the balanced-budget constraint, placed local governments under financial pressure: on average, cuts represented a 16 percent reduction in current expenditures and a 33 percent decrease of capital investments. Importantly, municipalities below 5,000 inhabitants were exempted from one important round of cuts, a €2.5 billion reduction enacted in 2010 (Law Decree 78/2010) and implemented over the following years. Additionally, towns affected by the earthquakes benefited from some exemptions. The discontinuity at 5,000 inhabitants will provide a strong identification strategy for the local effect, with a few caveats discussed in Section 2.4. Overall –as shown in Marattin, Nannicini and Porcelli (2019)– the downward trend followed by government grants was much steeper than the contraction in expenditures, suggesting that the fiscal adjustment involved an increase in alternative sources of revenues.

[1]It is beyond the scope of this paper to provide a detailed description of each law establishing transfers cuts, given that in our data we only observe the amount of transfers at the end of each year. Generally, these type of intervention indicates a total amount of budget reduction, to be distributed proportionally between municipalities based on their population. Marattin, Nannicini and Porcelli (2019) identify six rounds of transfer cuts, affecting years 2009 to 2015.

2.2.2 Local Government

Italian municipalities are governed by three bodies: a mayor (*sindaco),* an executive committee (*giunta comunale*) and a city council (*consiglio comunale*). The executive committee, composed of 2 to 12 members depending on population, is appointed by the mayor and conducts the everyday business of the municipality, making decisions under a simple majority rule. The mayor, with the help of the executive committee, directs and oversees the activities of the local police, whose members are employed by the municipality. Neither the mayor nor the executive committee has any degree of influence on other police forces, such as the national police force. This is an important feature, given that the strictness of enforcement can be varied quickly as conditions change. The city council (10 to 64 members) is mainly responsible for the approval of the yearly budget, and it convenes only a few times a year in small municipalities.

Since 1993, mayors are directly elected and serve for a 5-year term (since 2001, it was 4-year before), facing a 2-term limit, increased to 3 for towns below 3,000 inhabitants. Depending on the population, the electoral law may prescribe a single-round plurality (below 15,000) or a runoff system. In both cases, the winning candidate obtains a majority in the council that should allow for a frictionless governing body. Despite a standard term of 5 years, local government may come to an early dissolution for a variety of reasons (death, resignation, criminal convictions, among the others), and in that case new elections are held, most often between April and June. For this reason, elections are held in a staggered way, with only a fraction of the municipalities having elections each year.

2.3 Data

I gather together data at the municipality level from a variety of administrative sources. For the fiscal consolidation analysis, I restrict the sample to municipalities between 1,000 and 10,000 inhabitants, to ensure comparability between the treatment and control group.

I also exclude towns in the five special autonomy regions, and those affected by the 2009 and 2012 earthquakes (which were exempt from several cuts). This leaves 6,504 observation each year. Summary statistics of the main variables used in the analysis are contained in Table 2.1. For each variable I keep all the non missing observations, which means the sample will be slightly unbalanced depending on the specification, given the different availability of data from different sources.

Budget Data. Information on municipal finances come from the official budget accounts, submitted every year to the Italian Ministry of the Interior. Coverage spans between 2009 and 2014, the period immediately before and after the policy implementation and when other confounding policies are generally constant over time. These official budget accounts include data on both revenues and expenditures, disaggregated by source and accounting basis; in particular, it contains the accrual amount (amount issued in each year) and the cased amount (the sum actually obtained). The main focus of attention is a specific type of fee, that falls under the heading "Local Police – Fines related to the Road Traffic Act", and track all sorts of traffic related fines, including speeding fines, red light cameras and restricted zone violations. Unfortunately, the data do not provide more specific information on the type of fine, and this would require collecting it individually. On average, municipalities raise about 11 euros per capita in traffic fines, slightly less that 1 percent of the total revenues. Of this amount, almost 80 percent is also cashed every year. I show in Section 2.5.2 that the collection ratio – calculated as fines cashed over fines imposed – is another dimension on which the electoral cycle has an impact.

Car Accidents Data. To provide measures of road safety, I obtained data on the universe of severe car accidents, collected from the main Italian statistical authority, the National Institute of Statistics (ISTAT-Eurostat), and aggregated at the municipality level. It provides the dates and locations of accidents with at least one injured individual, and thus represents only a subset of all the accidents occurring in Italy in the covered period

(2002-2014), but also the most relevant to our analysis.[2] From that, I construct three main outcomes of interest: the accident rate (number of accidents per 1,000 inhabitants), the injury rate and a death rate, both of which are constructed in a similar way to the accident rate. Additional information, such as the type of accidents, are not always available, especially at the beginning of the period, and would result in the loss of most observations, therefore I will not employ them. Over the years considered, there are on average 1.9 severe accidents for every 1,000 inhabitants, with a significant decreasing time trend for all the measures considered, as depicted in Figure 2.1. In particular, the death rate almost halved in the 10 years between 2002 and 2012.

Electoral Data. Political and electoral data were collected from the Italian Ministry of the Interior and contains information about election years, results and individual characteristics of the elected mayor, including ideology and the victory margin of the election. As shown in the last rows of Table 2.1, political ideology plays a weaker role in small municipalities, where most candidates, despite being backed by large parties, participate in the competition under personal lists. The main explanatory variable in the electoral motive is a dummy equal to one if a local election is held in a specific municipality. The margin of victory is an important control variable, as it is a strong predictor of electoral competition, hence an increase in the incentives to please voters.

Additional Data at the Municipality Level. I complement the previous data with additional information at the municipality level, which comes from the National Institute of Statistics. Some of them are time invariant, mostly geographical variables; some vary over time, such as the age structure of the population, average income and average house market value. These variables will be used as controls in order to improve precision and as a robustness check, although they are not strictly required for identification in the

[2]Based on their sample, Bertoli and Grembi (2018) claim that accidents with no injuries account for 44% of the total accidents happening in one year, although their estimates are likely to overlook even smaller crashes that go unreported.

regression discontinuity design.

2.4 Empirical Strategy

The goal of this paper is twofold: to providence evidence that both a budgetary and an electoral motive affect traffic enforcement, which is done by exploiting two different econometric designs, best suitable to circumvent the relevant endogeneity issues.

2.4.1 Budget Motive

As briefly discussed in Section 2.2.1, some of the laws establishing transfer cuts present a sharp discontinuity, completely exempting municipalities below 5,000 inhabitants. This creates the perfect environment for a regression discontinuity design.[3] Using population thresholds as an identification tool has undeniable benefits, and has been widely used in the literature to estimate the impact of a multitude of policies,[4] but it necessitates a careful design to address potential pitfalls (Eggers et al., 2018). The main assumptions for a sharp cross-sectional regression discontinuity to provide unbiased causal estimates are: (1) that the conditional distribution functions are continuous around the threshold; (2) the absence of contemporaneous policy changes; and (3) that municipalities are not able to manipulate the threshold. In particular, assumption (2) is not satisfied in this context, since the fiscal consolidation is not the only policy changing at the threshold, but so does the salary of the mayor and – until 2013 – the absence of a Domestic Stability Pact. Both factors have the potential to affect traffic fines and road safety, given that a higher salary may attract better politicians (Gagliarducci and Nannicini, 2013*a*) and the DSP represents an additional fiscal constraint by removing the possibility to run a budget deficit. I address the potential bias in several ways using two main approaches.

[3] See Thistlethwaite and Campbell (1960*a*) for a pioneering paper, and Imbens and Lemieux (2008*b*) for a comprehensive theoretical and practical review of the methodology.

[4] These include evaluating gender quotas Casas-Arce and Saiz (2015); Campa and Bagues (2017), electoral rules Fujiwara et al. (2011); Hopkins (2011*a*), fiscal transfers Brollo et al. (2013*a*); Eggers et al. (2018).

RD over time. The budget data spans over the period before and after the fiscal contraction is implemented, and the confounding policies are constant over the period 2009-2013. By observing the cross-sectional discontinuity over time we can easily detect whether any confounding factor is present absent the policy of interest. I will show that the impact of confounders is negligible, hence cross-sectional discontinuitiy estimates can be interpreted causally. Adding year 2014 in the estimation requires a discussion of the effect that the introduction of the DSP has on small municipalities, because this is a confounding treatment that is not taken care of by the panel structure of the data. However, it represents a budget tightening in the control group, which would induce –if any– a downward bias in my estimate, by providing non treated municipalities an additional incentive to raise revenues through traffic fines. I therefore argue that RD estimates for year 2014 provide a lower bound of the causal effect of the policy.

Practically, estimation is conducted following Fan and Gijbels (1996*a*) and Imbens and Lemieux (2008*a*), using local polynomial regressions. The sample is restricted to municipalities between 1,000 and 10,000 inhabitants, to ensure comparability, and the observations are further limited using the MSE optimal bandwidth selector proposed by Calonico, Cattaneo and Titiunik (2014*b*). For an optimally chosen neighborhood h around the threshold, on each year $t \in [2009, 2014]$, I estimate the following equation:

$$Y_{it} = \lambda_t + \sum_{k=0}^{p} \left(\beta_{kt} \tilde{P}_{it}^k \right) + T_{it} \sum_{k=0}^{p} \left(\gamma_{kt} \tilde{P}_{it}^k \right) + \boldsymbol{\delta}_t' \boldsymbol{X_{it}} + \epsilon_{it} \tag{2.1}$$

where Y_{it} denotes the outcome (e.g. traffic fines issued, road accidents), T_{it} is a dummy equal to 1 for municipalities above 5,000 inhabitants, $\tilde{P}$ is the municipality population normalized at the threshold, λ_t is a time fixed effect (a constant for each year), ϵ_{it} is the error term, and γ_{0t} is the coefficient of interest, representing the impact of a tighter budget. $\boldsymbol{X_{it}}$, a vector of controls, is not required for identification but can help in reducing small sample bias and improve the precision of the estimates. It is also used to provide robustness checks on the baseline results.

Difference in Discontinuity. To provide a more formal control for unobserved and confounding factors, I also adopt a difference-in-discontinuities approach (Grembi, Nannicini and Troiano, 2016*a*) that provides a neat treatment of time-invariant differences in a two-period setting. Under the assumption that: (1) the confounding discontinuity is constant over time, and (2) it is the same with and without treatment (i.e. the local parallel trends assumption in Eggers et al., 2018 is met), the average treatment effect at the threshold can be estimated with the following equation:

$$\tilde{Y}_{it} = \lambda_t + \sum_{k=0}^{p} \left(\beta_{kt} \tilde{P}_{it}^k \right) + T_{it} \sum_{k=0}^{p} \left(\gamma_{kt} \tilde{P}_{it}^k \right) + \boldsymbol{\delta}_t' \boldsymbol{X}_{it} + \epsilon_{it} \tag{2.2}$$

where $\tilde{Y}_{it} = Y_{it} - Y_{it^*}$, and t^* denotes a baseline year in the pre-policy period. In the two-period scenario this is unique, but in my case several periods are observed before policy implementation. For robustness, I use years 2009, 2010, 2011 and the 2009-2010 average as alternative baseline measures in the subsequent analysis, although results are largely unaffected by the baseline choice, given the negligible impact of confounding policies.

Validity Tests. In both approaches, the regression discontinuity and the difference in discontinuity, I will present a variety of validity tests to provide a stronger case for causal interpretation. Specifically, I report in the Appendix results from a standard event study analysis in a suitable restricted sample, and evaluate the sensitivity of results to different bandwidths. In Section 2.5.3, I present a local balance table for observable characteristics of the municipalities, and test for manipulation of the running variable following McCrary, 2008*a*.

2.4.2 Political Motive

In the second part of this paper, I exploit the variation in local election dates to provide evidence on the existence of a political motive in traffic enforcement. Figure 2.2 shows the number of elections held every year. with a clear concentration on specific years. It

is important to stress – though – that the reason why some municipalities had them on different dates can have several explanations. Sometimes they date back to the very first democratic election held in 1946, or are just the legacy of random events from the past. Therefore, it is very unlikely they are systematically different with respect to unobservable characteristics.

My estimation strategy follows the standard approach in the literature, that is an OLS model with fixed effects and an election year dummy (Levitt, 1997; Shi and Svensson, 2006; Cole, 2009; Baskaran, Min and Uppal, 2015; Bee and Moulton, 2015*a*; Alesina and Paradisi, 2017; Bracco, 2018)

$$Y_{ipt} = \alpha_p + \lambda_t + \beta EY_{ipt} + \boldsymbol{\gamma}' \boldsymbol{x_{ipt}} + \epsilon_{ipt} \tag{2.3}$$

where each observation is a municipality, λ_t a year fixed effect, α_p a province fixed effect, EY_{ipt} is a dummy for an election year and Y_{it} is the outcome of interest. The strategy is similar to that used by Bertoli and Grembi (2018) for a sample of municipalities in two Italian regions. I also provide a two stage least square estimation where measures of road safety are regressed on the amount of traffic fines, instrumented using the election year dummy. In order for the instrumental variable approach to provide meaningful results, relevance of the instrument and the exclusion restriction have to be satisfied. As to the relevance of my instrument, I provide results from the first stage, and compute the weak-instrument robust confidence interval following (Andrews, Stock and Sun, 2019). As to the exclusion restriction, the instrument would be invalid if elections affect road safety in ways different than traffic enforcement, for example through a reduction in infrastructure maintenance. However, during an electoral campaign, incumbent mayors are more, not less, likely to invest in public safety (Levitt, 1997). I find therefore that the only reason why electoral campaigns may negatively affect safety, is through unforeseen and unwanted consequences of policy aimed at pleasing voters, and that the exclusion restriction is likely to be satisfied given this context.

2.5 Results

In this section I provide the main results and a discussion of them in two separate subsections, the first regarding the fiscal consolidation, and the second focusing on election years.

2.5.1 Fiscal Consolidation

Figure 2.3 provides a description of the policy variable, that is the amount of grants transferred each year from the central government to the municipalities. Values are larger, in per capita terms, for small towns reflecting the presence of fixed costs and economies of scale, but are in general flat as population increases. Another noteworthy feature is the substantial heterogeneity between municipalities with similar populations. This happens because the total amount observed in the budget data represent the sum from all sources of revenues, including earmarked funds. Nevertheless, it can be clearly seen that, starting in 2011, a discontinuity arises between observations below and above the relevant threshold. Appendix Figure 2.A.1 reports the same graphs, using variables that differ with respect to 2010 values. It shows that while grants went down for every municipality, the effect is larger for the treated municipalities starting in 2011.

Government Transfers. Figure 2.4 contains the first stage result of the fiscal consolidation, a measure of the discontinuity in government transfers over time. Figure 2.4a reports the cross section RD coefficients at the 5,000 inhabitant threshold over time. Figures 2.4b to 2.4d graph the difference-in-discontinuity coefficient. It is evident that municipalities right above the threshold receive less money starting in 2011, and that the difference is highly significant in 2013. However, the result is not as sharp as we would have imagined by looking at the design of the policy. In particular, it appears that several treated towns had the opportunity to, at least partially, compensate the budget contraction with alter-

native grants. This reduces the power of my instrument, although it is still significant. Table 2.2 lists the coefficients for each specification, showing that treated municipalities had an additional transfer reduction of about 25 euro per capita in 2012, and 30 euro in 2013 terms, a 13 percent decrease in both cases. Comparing the four different estimation procedures, difference-in-discontinuity estimates are never significantly different for cross-sectional results, although the point estimates are larger. This suggests that unobserved cross-sectional differences, while introducing noise to the data, do not induce a substantial bias in the results.

The appendix reports some robustness checks, with the implementation of alternative estimation strategies. Appendix Table 2.A.2 provides detailed estimates for several specifications, and coefficients are pictured in Appendix Figure 2.A.2. Appendix Figures 2.A.4a and 2.A.4b contain difference-in-difference estimates, restricting the sample to a symmetric bandwidth of 1,000 inhabitants. Appendix Figures 2.A.4c and 2.A.4d display results from a local polynomial regression on the same subsample estimated using 2.1. It fits a second degree polynomial in population on both sides of the threshold. Results used for the graphs are presented in Appendix Table 2.A.1. As an additional robustness check, Appendix Figure 2.A.3 provides a measure of the sensitivity of the estimates as the bandwidth varies, for the diff-in-diff and LPR specifications. Overall, results are consistent and not too sensitive to small changes in the sample.[5]

Traffic Fines. Having established that the policy provides a significant first stage, I now turn to the reduced form analysis of its impact on traffic fines. Figure 2.5 contains a graphical representation of the main results, and more details are reported in Table 2.3. First, it is fair to say that precision is sometimes an issue, and despite similar pre-trend coefficients, the cross-sectional RD presents larger point estimates than the difference-in-discontinuity approach. Nevertheless, there is a clear increasing trend after the fiscal

[5]Due to the availability period of some variables, 2014 is omitted from the specifications with additional controls.

consolidation is implemented, and the difference between treatment and control municipalities becomes significant in 2013 and 2014. Diff-in-discontinuity results are about half the size of the cross-sectional coefficients, roughly 4 euro per capita in 2013, and 6.5 euro per capita in 2014. This represents an almost 50 and 80 percent increase, respectively, over the control group mean. Put it in a different way, the average increase in traffic fines in treated municipalities was roughly 10 percent of the budget reduction in 2013, and 60 percent in 2014.

Appendix Figure 2.A.4 provides results for alternative specifications, as does Appendix Table 2.A.2. Estimates from the difference-in-difference estimator show no sign of discontinuity, suggesting that the effect is indeed local, and that controlling for the distance from the threshold is crucial for identification. In the appendix, I also look at the intensity of money collection, evaluating the impact of fiscal contraction on the collection ratio variable. Results are displayed in Appendix Figure 2.A.5, and reported in Appendix Table 2.A.3. It appears that the policy did not induce any significant change in the collection policy of the municipalities, and that most of the additional revenues came from newly issued fines.

The imprecise effect measured in some of the specifications may have several explanations. One possibility is the weakness of the instrument, which was previously discussed in the preceding paragraph, but it may also mask significant heterogeneity between municipalities. In particular, I show some suggestive evidence that traffic fine are used as a residual source of revenues to maintain the pre-policy level of current expenditure. Figure 2.6, and Table 2.4 present discontinuity estimates on the total current expenditures, showing an increase in per capita expenditures after the policy implementation. This could, in principle, be a consequence of the budget contraction. In fact, politicians may have some incentive to introduce additional sources of revenues (such as fines or increases in taxation), but when they are forced to do that because of the exogenous budget contraction, they also collect more money than expected. It would be especially true if the increase in

funds comes from fixed investments, like the introduction of new speed cameras or red light cameras. In addition to small changes in expenditures, municipalities with similar population differ quite a bit in terms of local taxation, in particular with respect to the Income Tax Surcharge Rate, that varies from 0 to 0.8 percent, the maximum allowed by law. Figure 2.7 provides a distribution of the Surcharge Rate in 2010, for municipalities between 1,000 and 10,000 inhabitants. It shows that a quarter of the municipalities had a 0 local surcharge rate, but more than half were already above 0.4. Having a large Income Tax Rate in the pre-policy period means being constrained on that particular instrument, and having to rely on alternative tools in the aftermath of the fiscal consolidation. In Figure 2.8 I display estimates from the discontinuity design for subgroups of municipalities based on the pre-policy tax rate, and report details in Table 2.5. Towns with a value larger than 0.5 in 2010 are in the top 25 percent, 0.6 in the top 20 percent, and 0.7 in the top 10 percent. Despite the reduced precision due to sample size, it is evident that fiscally constrained towns relied more heavily on additional traffic fines, especially in 2014, where towns in the top 20 percent of the surcharge distribution raised fines by 50 euro per capita, a 6-fold increase with respect to comparable control group towns.[6] Overall, this result provides additional evidence that traffic fines are indeed intended as a residual source of revenues when other tools are not feasible.

Road Safety. The subsequent step is to evaluate the reduced form impact of the policy on road safety measures. Graphical results are presented in Figure 2.9, with additional details in Table 2.6. These results present regression discontinuity and difference-in-discontinuity estimates for the accident rate (per 1,000 inhabitants), the injury rate and the death rate. The effects estimated are very noisy, with large standard errors and shifting signs between 2012 and 2013. In particular, a spike in the death rate in 2012 is particu-

[6] A word of caution on this results is needed. The optimal bandwidth estimator for the top 20 and 10 percent restrict sample size to a very narrow neighborhood around the threshold, meaning that the coefficient is very sensitive to outliers. Expanding the band a little further bring the point estimate down to values comparable to 2013.

larly difficult to explain. Overall, it is troublesome to interpret the reduced form results, and we do not observe the significant and robust positive effects reported, for example, in Makowsky and Stratmann (2011). On the contrary, results from the reduced form are not encouraging and they seem to show that increasing traffic fines had a negative effect on road safety. This is true if, for example, tools that are more efficient in raising revenues (e.g. red light or speed cameras) have also drawbacks such as drivers breaking abruptly to avoid the fine.[7] Insofar as compliance to regulation is imperfect, we can use treatment status to isolate an exogenous variation in traffic fines, and estimate its impact on road safety. One possible concern is that the fiscal consolidation violates the exclusion restriction, that is it has an impact on accidents through channels different from traffic fines, e.g. by reducing the fiscal capacity of the municipality, hence investment in the road infrastructure. However, as I have already shown in Figure 2.6, there is no evidence of changes in overall spending after the policy implementation. A second concern is the strength of the instrument, in particular with respect to the staggered phase in of the policy described in Figure 2.4. As Figure 2.5 shows, the impact of the fiscal consolidation on traffic fine is gradual and never very strong, making the instrumental variable approach very prone to bias. I performed weak instrument robust inference using the Anderson and Rubin (1949) test presented in Andrews, Stock and Sun (2019), and could not reject a confidence interval equal to the real line. Results were therefore omitted.

2.5.2 Electoral Cycle

The second experiment of my analysis investigates the presence and effect of an electoral cycle in traffic enforcement. Table 2.7 provides results from equation (2.3) on two different sets of outcomes. In Panel A the dependent variables are measures related to traffic

[7] A summary made by Li, Graham and Majumdar (2013) suggests this is not true, and that on average speed camera tend to reduce car crashes. However, endogeneity of the location is likely to play a substantial role, possibly overlooked by the literature, a have a strong interaction with the policy makers interests. Depending on the main short run goals, cameras can be placed in areas where they save most lives or in areas where they raises most revenues.

enforcement: traffic fines imposed, traffic fines cashed and the collection ratio. The full specification of column (3) shows that in an electoral year there is a 0.6 euro reduction in traffic fines per capita imposed, equivalent to about a 5 percent decrease with respect to the average during non election years. A similar picture is present in column (6) where the effect is measured with respect to the money actually collected. It shows a 6 percent reduction, similar to the reduction in column (9), where the ratio between the two is used. It seems that the collection channel is more significant, in line with the results in Bracco (2018) and Bertoli and Grembi (2018), and that non cashed fines play a major role in driving down revenues. In Panel B, I study the reduced form impact of elections on accident rates and injuries, finding again meaningful and significant results. In electoral years there are about 0.04 more accidents per 1,000 inhabitants, that is a 1.9 percent increase in the accident rate, and a 1.6 percent increase in the injury rate. On the other hand, no effect on the number of deaths, or the death rate, is detected.

A likely interpretation of these results is that a generalized mitigation of traffic enforcement, justified by political reasons, can induce welfare costs in terms of road safety. It is beyond the scope of this paper to investigate what is the optimal level of enforcement, but this paper does attempt to provide a measure of how it affects road safety. As discussed in Section 2.5.1, a suitable instrument for a change in traffic fines can be useful in this context, and I am confident that a dummy for local elections can be such an instrument. Panel A of Table 2.8 displays results for several two stage least squares estimation, where the amount of traffic fines is instrumented using the election year dummy. In Panel B, I also report some first stage diagnostics, in particular the Kleibergen and Paap (2006) and Montiel Olea and Pflueger (2013) robust F-tests to detect weak instruments. Given the low values, I also compute the Anderson and Rubin (1949) weak-instrument robust confidence intervals as suggested in Andrews, Stock and Sun (2019). Robust estimates reduce the amount of information that we can get, but confirm the sign of the effect in most specifications, without ruling out the possibility that magnitudes are even larger. A

conservative lower bound of the magnitude, for both accident and injury rate, show that a euro per capita decrease in the fines imposed increases the accidents and injury rate by at least 0.035 per 1,000 inhabitants, representing a 16 percent increase. This result is quite substantial, although we should be cautious in extrapolation because these types of effects are very likely to be non-linear. In addition, the interpretation is made more difficult by the fact that we do not observe what type of fines are lowered during electoral years, nor can we compare this experiment to the fiscal consolidation presented before. Nevertheless, we can at least speculate that the different effects on road safety can be attributed to different enforcement tools that are affected, and that obtaining this data could provide a valuable addition to the analysis presented in this paper.

2.5.3 Validity Tests

Although the assumptions associated with the validity of a regression discontinuity approach cannot be directly tested, I can provide some justification on why I believe they are satisfied. Table 2.9 presents the RD equivalent of a balance table, where I evaluate the presence of a discontinuity in a number of control variables. However, note that under the assumptions of the difference-in-discontinuity we do not require the absence of a discontinuity, but it is enough that they are constant over the period analyzed. This is what happens in Table 2.9, where the point estimate is roughly equal over the different years examined.

An additional assumption, is that municipalities are not able to sort themselves on a particular side of the threshold, generating a discontinuity in the distribution. First of all, it is important to stress that the relevant population variable is the so-called legal population, the official count performed every 10 year with the official census. Variation to this number can happen only in very specific cases, for example the splitting or merging of different municipalities. Due to the time a census requires before finalizing data, the 2001 legal population was in force until the end of 2012. It is therefore immediate to see

why mayors could not have affected it strategically. As a robustness check, Figure 2.10 provides the results from a McCrary (2008*a*) test, showing the absence of a discontinuity in both the 2001 and 2011 legal population.

2.6 Conclusion

Traffic enforcement is a potentially effective tool to mitigate road casualties, with the additional benefit of being used as a tool to raise revenues for local governments. An ideal policy would only have accident minimization as its goal, but policymakers face contrasting incentives. In this paper I provided evidence that both a budgetary and an electoral motive shape the amount of traffic fines raised in Italian municipalities.

In the first part of my analysis, I evaluated the impact of a large fiscal consolidation that happened in Italy between 2010 and 2014, and resulted in large transfer cuts from the central government to local municipalities. Towns with less than 5,000 inhabitants were largely exempted from the policy, generating a discontinuity in treatment. Using a series of regression discontinuities and differences-in-discontinuities, I quantify the extent of a revenue motive in traffic enforcement, showing that municipalities affected by the cuts increased fine collection to cover roughly 10 percent of the reduced revenues. This represents a 40 percent raise compared to 2010, the last year before policy implementation. A stricter enforcement, motivated by budget considerations, did not lead to a long lasting increase in road safety, measured in terms of both the accident rate and the injury rate, and I cannot exclude that it actually worsened road safety. This result has an important implication on income inequality. In most countries, traffic fines are set as a lump sum amount, without regards to the individual's income, and the rationale is that larger sums should dissuade people from committing the most serious infractions. However, as we have shown, there is a taxation component in traffic fines, and, if we take this into account, an income-based system of sanctions could be fairer from a distributional standpoint. This argument would reinforce the other claim that the incidence of fines is regressive

in nature, especially if richer and poorer individuals have the same tendency to violate traffic laws.

In the second part of the paper, I provided some evidence that fine collection may also be motivated by an electoral motive. Using local elections in Italian municipalities, I estimate that the total amount of traffic fines raised is reduced by 5 percent during an electoral year, and this is associated with a 3 percent increase in car accidents and injuries, with no effect on casualties. An instrumental variable analysis shows that weakening traffic enforcement can have a substantial negative effect on road safety. Unfortunately, data limitation on the type of fines prevented a better understanding of what types of violations are more likely to be ignored and cause the increase in accidents. Overall, these results should suggest a deeper reflection on how political incentives may affect the decisions of elected officials, and caution against delegating to them road safety competencies, without substantial counter balances from politically independent institutions.

References

Abouk, Rahi, and Scott Adams. 2013. "Texting Bans and Fatal Accidents on Roadways: Do They Work? Or Do Drivers Just React to Announcements of Bans?" *American Economic Journal: Applied Economics*, 5(2): 179–99.

Adams, Scott, and Chad Cotti. 2008. "Drunk driving after the passage of smoking bans in bars." *Journal of Public Economics*, 92(5-6): 1288–1305.

Albalate, Daniel, and Anastasiya Yarygina. 2017. "Road safety determinants: Do institutions matter?" *Rivista Internazionale di Economia dei Transporti / International Journal of Transport Economics*, 44.

Alesina, Alberto, and Matteo Paradisi. 2017. "Political budget cycles: Evidence from Italian cities." *Economics & Politics*, 29(2): 157–177.

Anbarci, Nejat, Monica Escaleras, and Charles Register. 2006. "Traffic Fatalities and Public Sector Corruption." *Kyklos*, 59(3): 327–344.

Anderson, Theodore W, and Herman Rubin. 1949. "Estimation of the parameters of a single equation in a complete system of stochastic equations." *The Annals of Mathematical Statistics*, 20(1): 46–63.

Andrews, Isaiah, James H. Stock, and Liyang Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics*, 11(1): 727–753.

Baskaran, Thushyanthan, Brian Min, and Yogesh Uppal. 2015. "Election cycles and electricity provision: Evidence from a quasi-experiment with Indian special elections." *Journal of Public Economics*, 126: 64 – 73.

Bee, C. Adam, and Shawn R. Moulton. 2015*a*. "Political budget cycles in U.S. municipalities." *Economics of Governance*, 16(4): 379–403.

Bertoli, Paola, and Veronica Grembi. 2018. "The Political Cycle of Road Traffic Accidents." CERGE-EI Working Paper No. 633.

Blincoe, Lawrence, Ted R. Miller, Eduard Zaloshnja, and Bruce A. Lawrence. 2015. "The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised)." National Highway Traffic Safety Administration Report DOT HS 812 013, Washington, DC.

Bourgeon, Jean-Marc, and Pierre Picard. 2007. "Point-record driving licence and road safety: An economic approach." *Journal of Public Economics*, 91(1): 235 – 258.

Bracco, Emanuele. 2018. "A fine collection: The political budget cycle of traffic enforcement." *Economics Letters*, 164: 117–120.

Brollo, Fernanda, Tommaso Nannicini, Roberto Perotti, and Guido Tabellini. 2013*a*. "The Political Resource Curse." *The American Economic Review*, 103(5): 1759–1796.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014*b*. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*, 82(6): 2295–2326.

Campa, Pamela, and Manuel Bagues. 2017. "Can Gender Quotas in Candidate Lists Empower Women? Evidence from a Regression Discontinuity Design." IZA IZA Discussion Paper 10888.

Casas-Arce, Pablo, and Albert Saiz. 2015. "Women and Power: Unpopular, Unwilling, or Held Back?" *Journal of Political Economy*, 123(3): 641–669.

Cohen, Alma, and Liran Einav. 2003. "The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities." *The Review of Economics and Statistics*, 85(4): 828–843.

Cole, Shawn. 2009. "Fixing Market Failures or Fixing Elections? Agricultural Credit in India." *American Economic Journal: Applied Economics*, 1(1): 219–50.

DeAngelo, Gregory, and Benjamin Hansen. 2014. "Life and Death in the Fast Lane: Police Enforcement and Traffic Fatalities." *American Economic Journal: Economic Policy*, 6(2): 231–257.

De Paola, Maria, Vincenzo Scoppa, and Mariatiziana Falcone. 2013. "The deterrent effects of the penalty points system for driving offences: a regression discontinuity approach." *Empirical Economics*, 45(2): 965–985.

Dubois, Eric. 2016. "Political business cycles 40Â years after Nordhaus." *Public Choice*, 166(1): 235–259.

Eggers, Andrew C., Ronny Freier, Veronica Grembi, and Tommaso Nannicini. 2018. "Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions." *American Journal of Political Science*, 62(1): 210–229.

Fan, Jianqing, and Irene Gijbels. 1996*a*. *Local polynomial modelling and its applications: Monographs on statistics and applied probability 66*. Vol. 66, CRC Press.

Fujiwara, Thomas, et al. 2011. "A regression discontinuity test of strategic voting and duverger's law." *Quarterly Journal of Political Science*, 6(3-4): 197–233.

Gagliarducci, Stefano, and Tommaso Nannicini. 2013*a*. "Do better paid politicians perform better? Disentangling incentives from selection." *Journal of the European Economic Association*, 11(2): 369–398.

Garrett, Thomas A., and Gary A. Wagner. 2009. "Red Ink in the Rearview Mirror: Local Fiscal Conditions and the Issuance of Traffic Tickets." *The Journal of Law and Economics*, 52(1): 71–90.

Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano. 2016*a*. "Do Fiscal Rules Matter?" *American Economic Journal: Applied Economics*, 8(3): 1–30.

Hopkins, Daniel J. 2011*a*. "Translating into Votes: The Electoral Impacts of Spanish-Language Ballots." *American Journal of Political Science*, 55(4): 814–830.

Imbens, Guido W., and Thomas Lemieux. 2008*a*. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics*, 142(2): 615 – 635. The regression discontinuity design: Theory and applications.

Imbens, Guido W., and Thomas Lemieux. 2008*b*. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics*, 142(2): 615–635.

ISTAT. 2011. "15th Italian General Population and Housing Census."

Kleibergen, Frank, and Richard Paap. 2006. "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics*, 133(1): 97–126.

Law, Teik Hua, Robert B. Noland, and Andrew W. Evans. 2009. "Factors associated with the relationship between motorcycle deaths and economic growth." *Accident Analysis & Prevention*, 41(2): 234 – 240.

Levitt, Steven. 1997. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." *American Economic Review*, 87: 270–90.

Levitt, Steven D., and Jack Porter. 2001*b*. "How Dangerous Are Drinking Drivers?" *Journal of Political Economy*, 109(6): 1198–1237.

Li, Haojie, Daniel J. Graham, and Arnab Majumdar. 2013. "The impacts of speed cameras on road accidents: An application of propensity score matching methods." *Accident Analysis & Prevention*, 60: 148 – 157.

Makowsky, Michael D, and Thomas Stratmann. 2009. "Political Economy at Any Speed: What Determines Traffic Citations?" *American Economic Review*, 99(1): 509–527.

Makowsky, Michael D., and Thomas Stratmann. 2011. "More Tickets, Fewer Accidents: How Cash-Strapped Towns Make for Safer Roads." *The Journal of Law and Economics*, 54(4): 863–888.

Marattin, Luigi, Tommaso Nannicini, and Francesco Porcelli. 2019. "Revenue vs Expenditure Based Fiscal Consolidation: The Pass-Trough from Federal Cuts to Local Taxes." IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University Working Papers 644.

McCrary, Justin. 2008*a*. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics*, 142(2): 698–714.

Montiel Olea, Jose Luis, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics*, 31(3): 358–369.

Nannicini, Tommaso, Andrea Stella, Guido Tabellini, and Ugo Troiano. 2013. "Social Capital and Political Accountability." *American Economic Journal: Economic Policy*, 5(2): 222–50.

Nordhaus, William D. 1975. "The Political Business Cycle." *The Review of Economic Studies*, 42(2): 169–190.

Shi, Min, and Jakob Svensson. 2006. "Political budget cycles: Do they differ across countries and why?" *Journal of Public Economics*, 90(8): 1367 – 1389.

Thistlethwaite, Donald L, and Donald T Campbell. 1960*a*. "Regression-discontinuity analysis: An alternative to the ex post facto experiment." *Journal of Educational psychology*, 51(6): 309.

World Health Organization. 2018. "Global Status Report on Road Safety 2018." World Health Organization.

2.7 Figures and Tables

2.7.1 Figures

Figure 2.1. Accidents and Deaths over Time

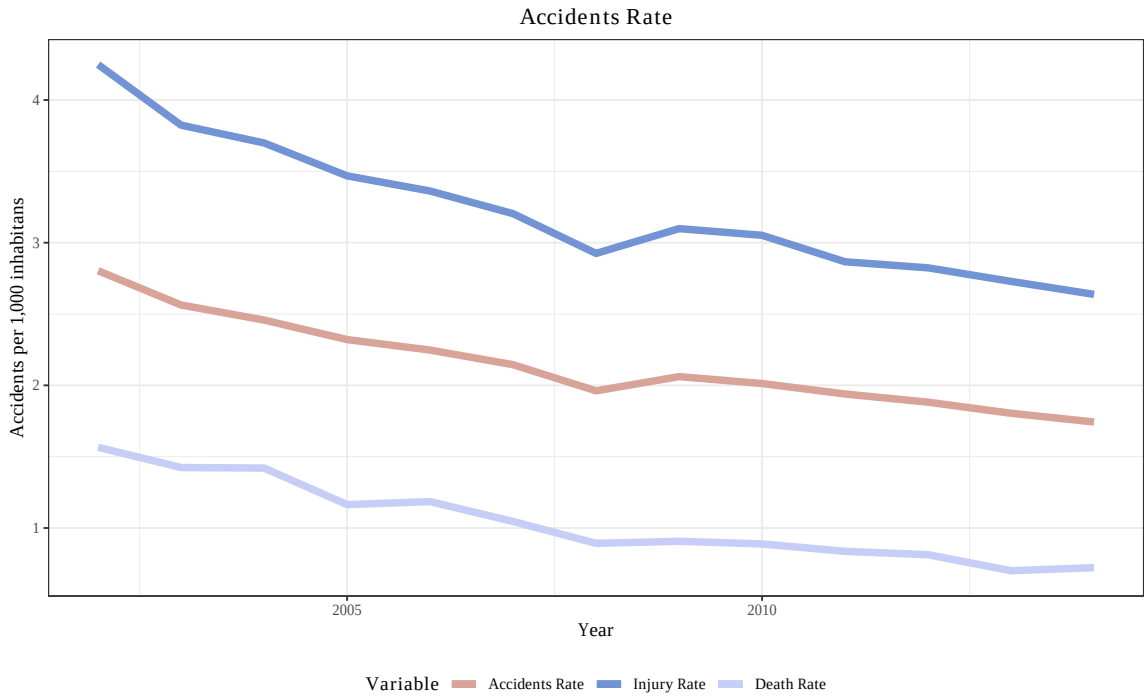


Figure 2.2. Local Elections over Time

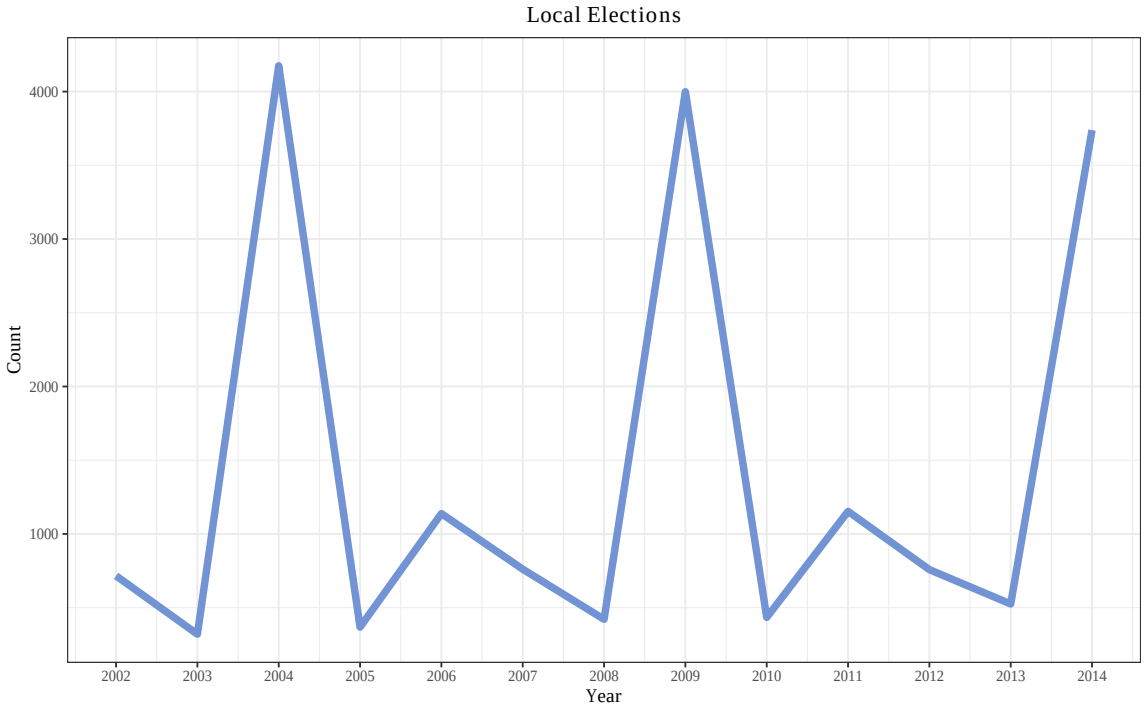
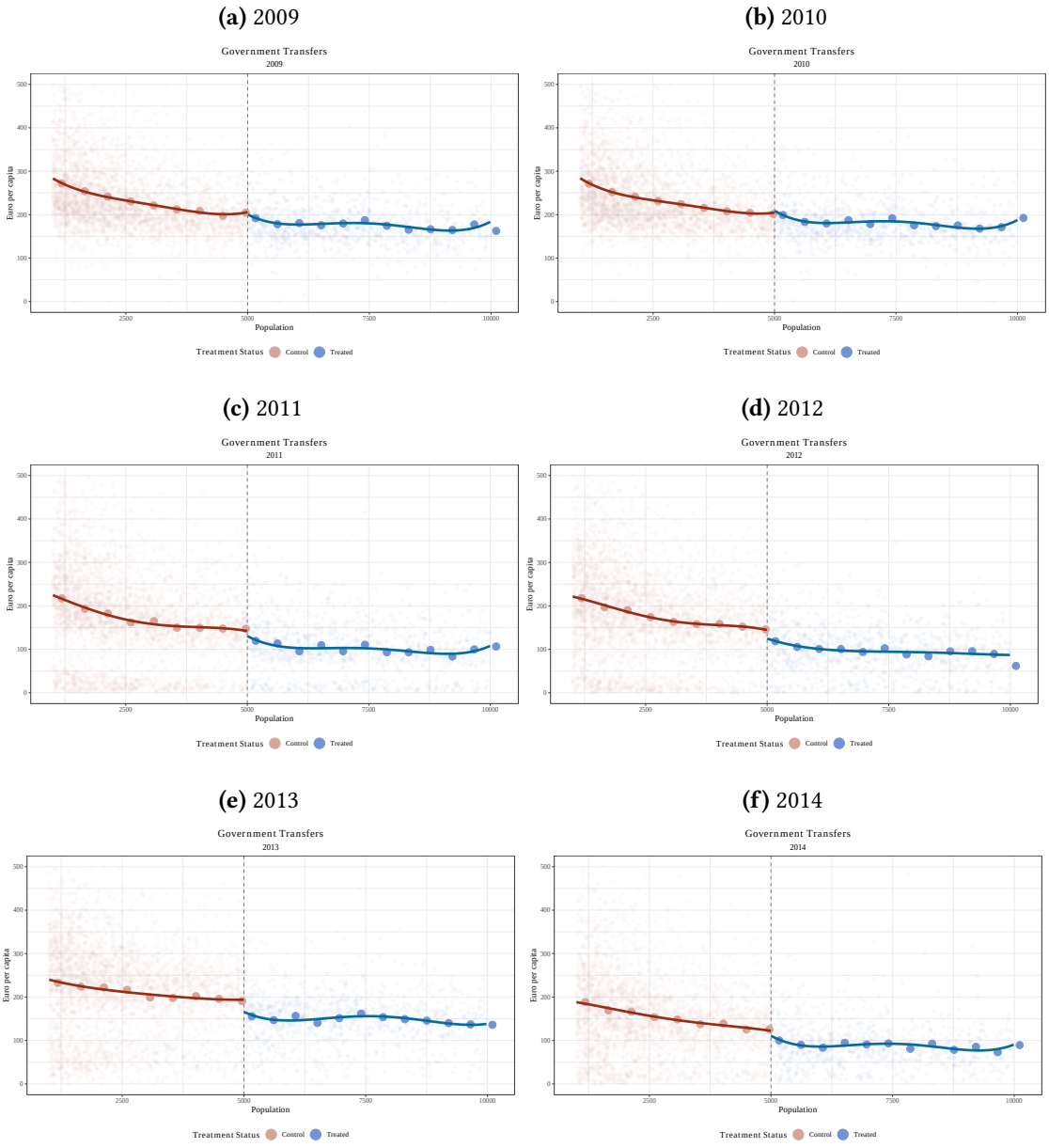
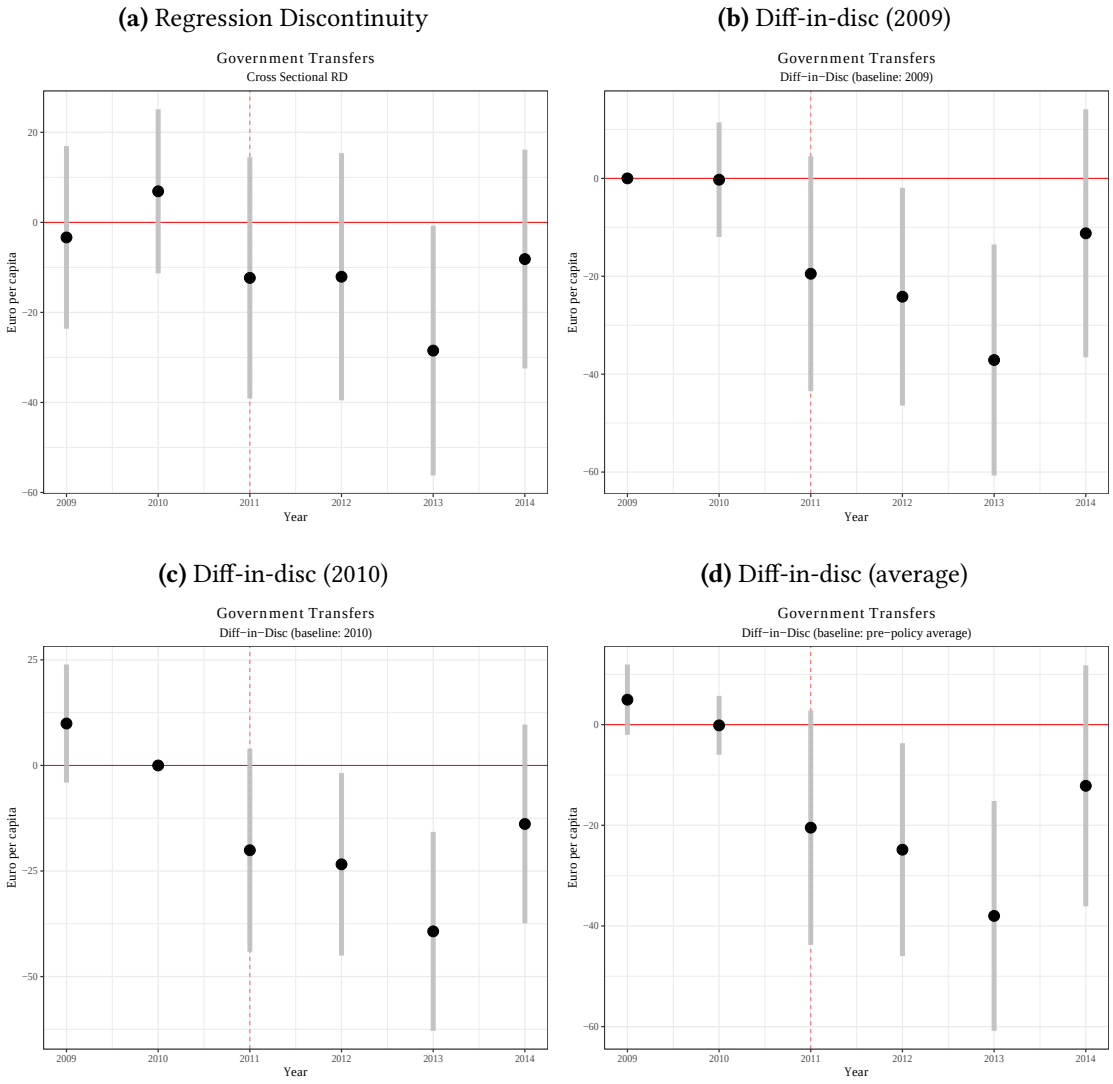


Figure 2.3. Government Transfers (2009-2014)



Note. Panels (a) to (f) plot the amount of grants received by each municipality in per capita terms. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants. Darker dots represent bin scatters.

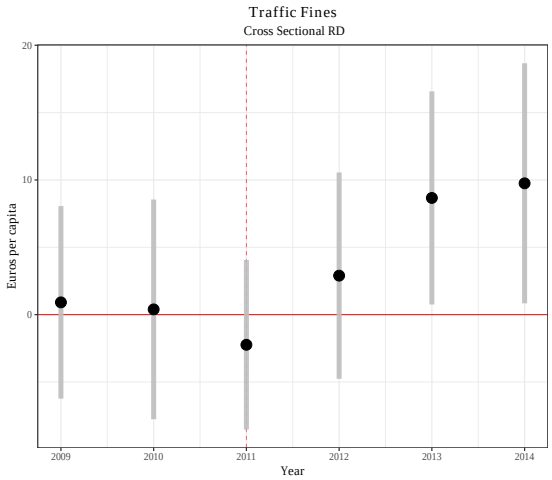
Figure 2.4. Discontinuity Estimates of Government Transfers



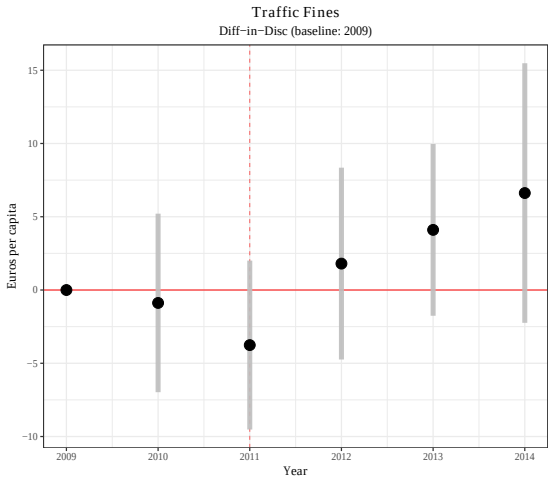
Note. Panels (A) to (D) plot the discontinuity coefficients from equations (2.1) and (2.2) using central government transfers as the dependent variable. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*).

Figure 2.5. Discontinuity Estimates of Traffic Fines

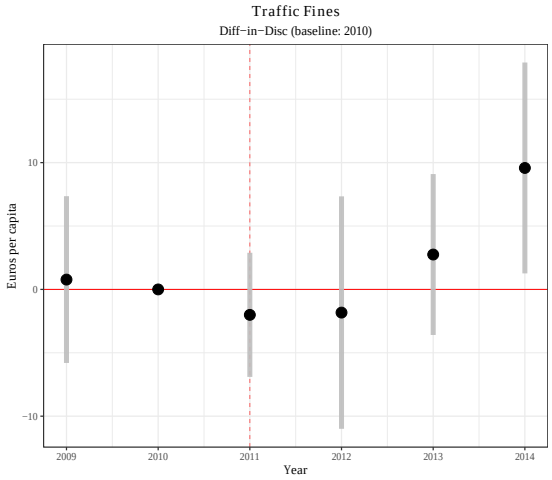
(a) Regression Discontinuity



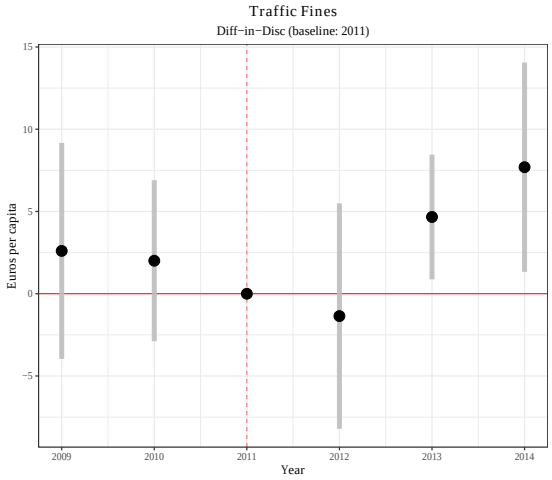
(b) Diff-in-disc (2009)



(c) Diff-in-disc (2010)



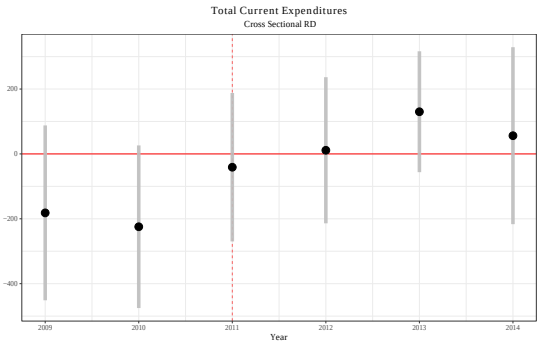
(d) Diff-in-disc (2011)



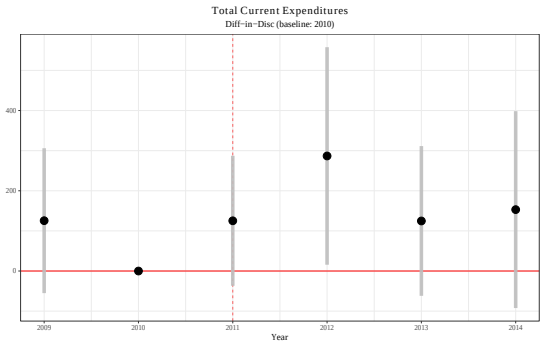
Note. Panels (A) to (D) plot the discontinuity coefficients from equations (2.1) and (2.2) using imposed traffic fines as the dependent variable. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*).

Figure 2.6. Discontinuity Estimates of Total Expenditures

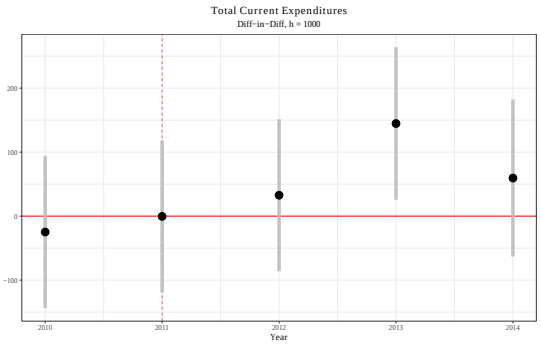
(a) Regression Discontinuity



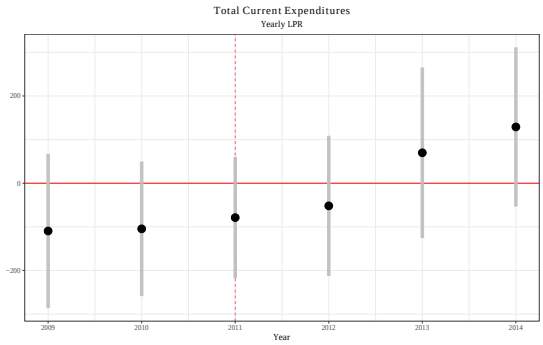
(b) Diff-in-disc (2010)



(c) Difference in Difference



(d) Local Polynomial Regression



Note. Panels (a) to (d) plot the discontinuity coefficients from equations (2.1) and (2.2) using total expenditures as the dependent variable. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*).

FIGURE 2.7. INCOME TAX SURCHARGE RATE DISTRIBUTION

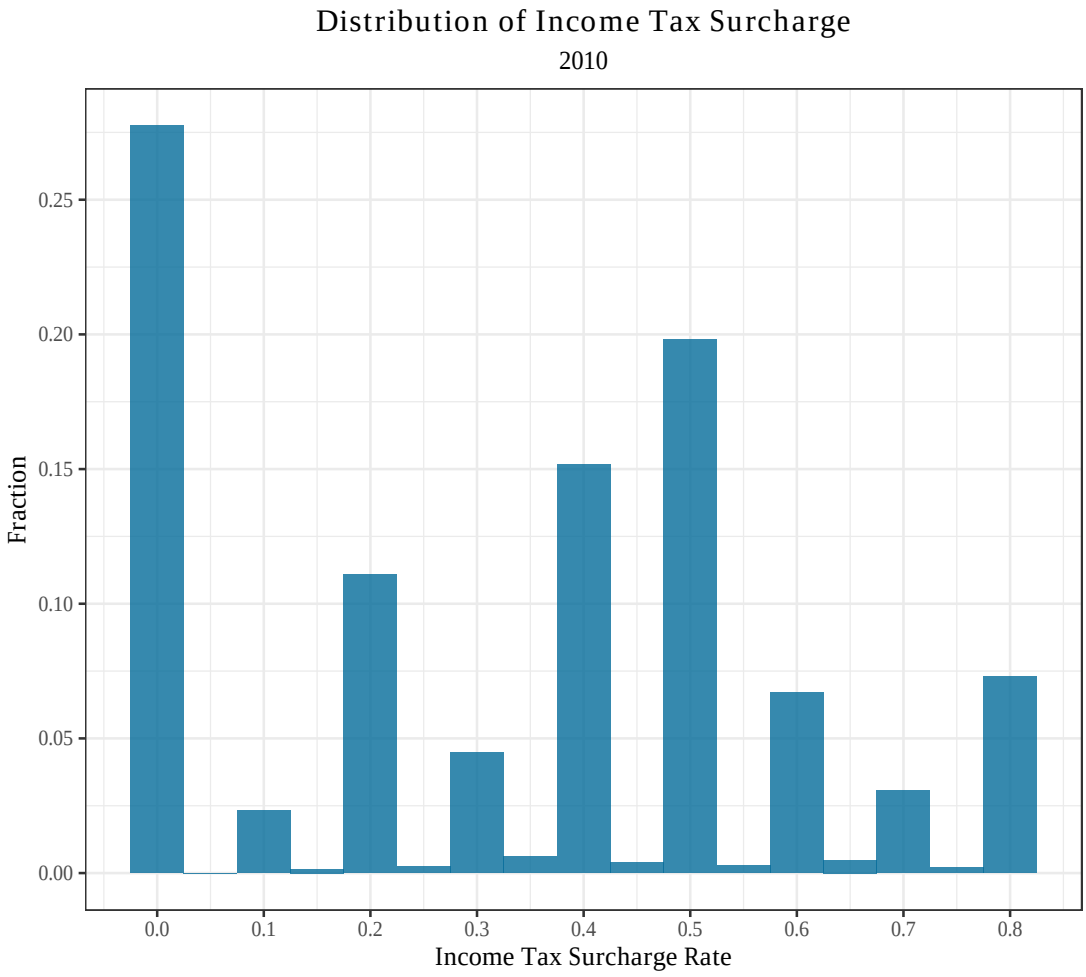
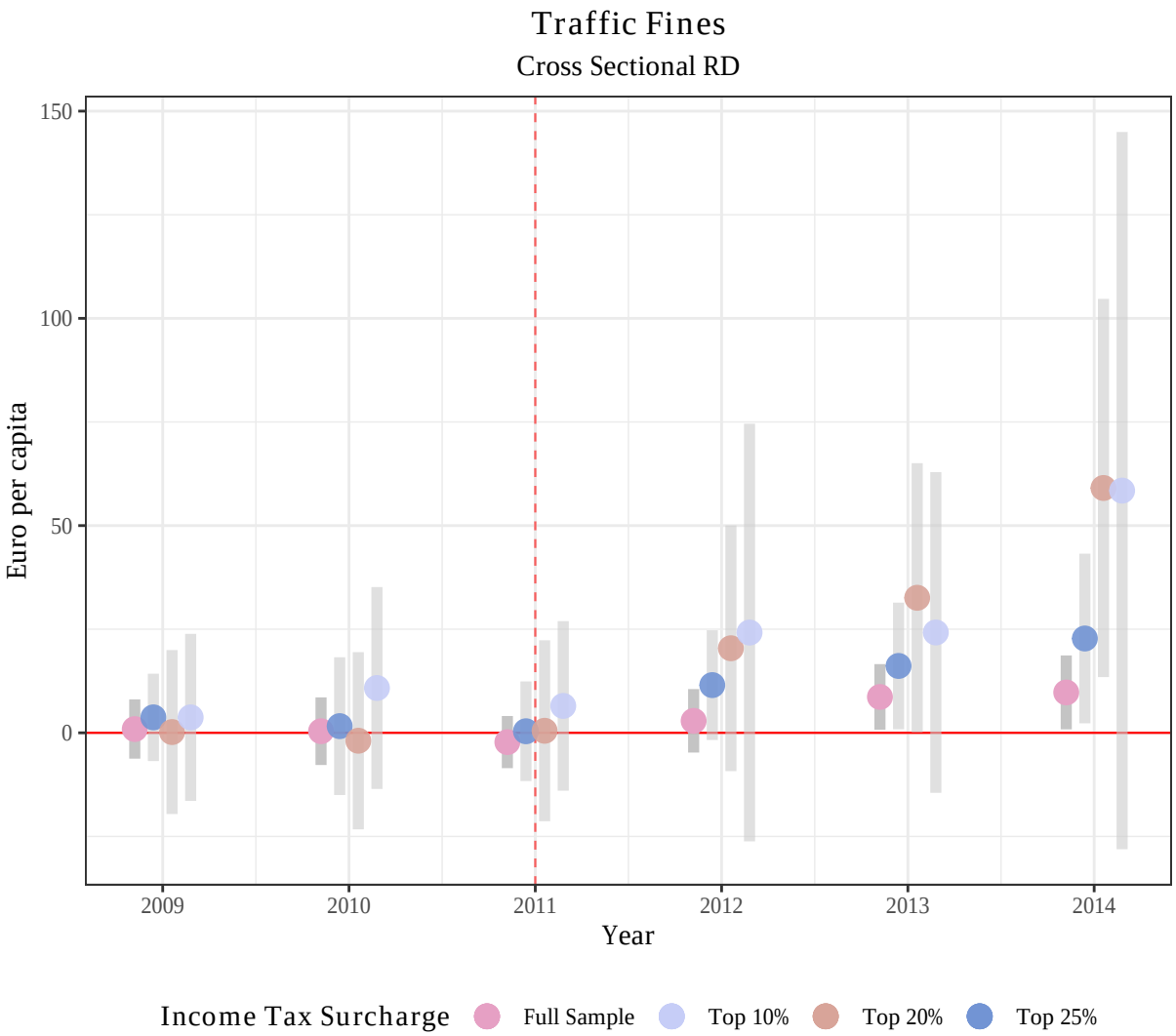
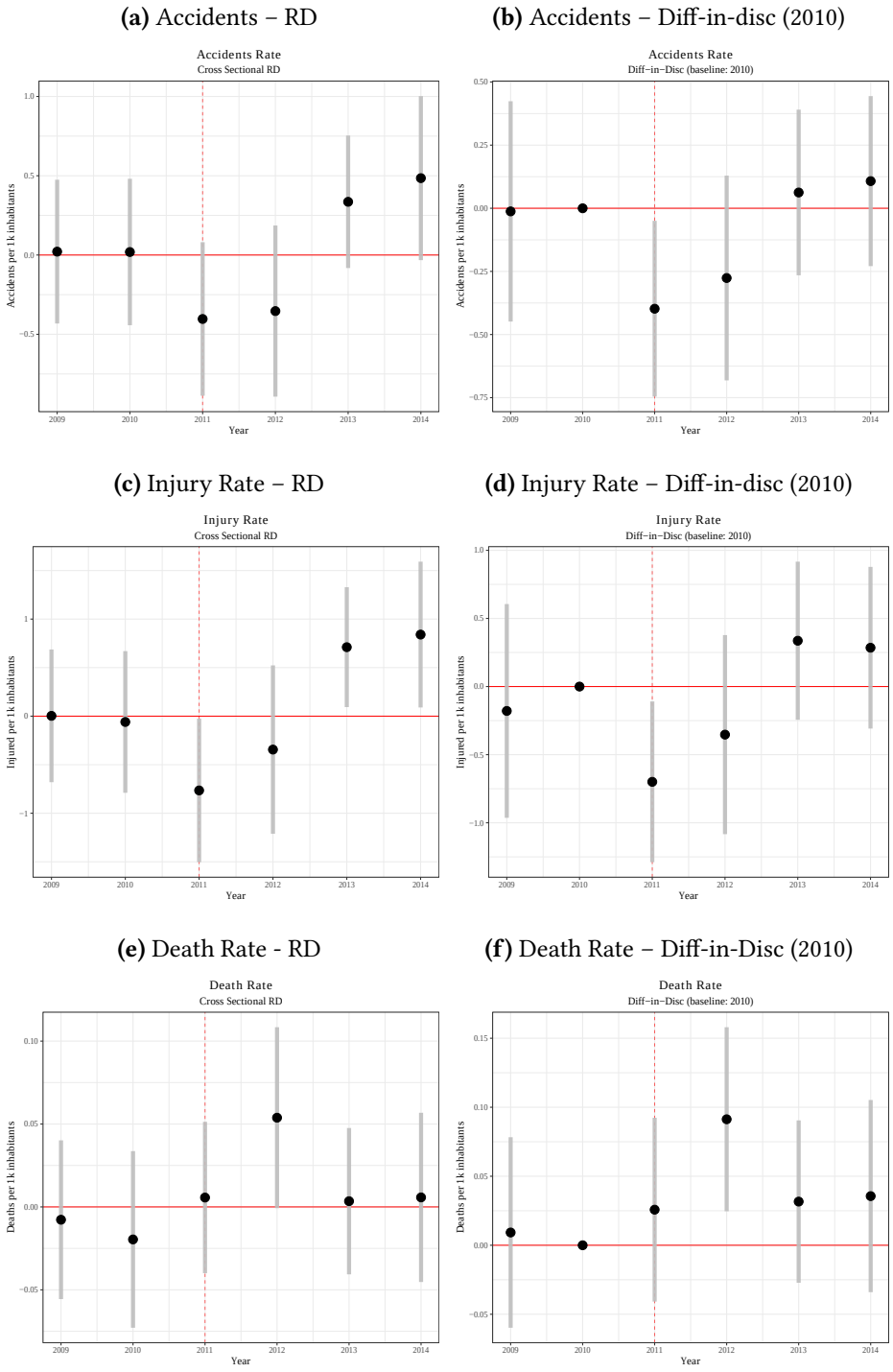


Figure 2.8. Traffic Fines and Taxes



Note. The graph plots cross-sectional RD estimates as in (2.5) for different subgroups depending on the pre-policy Income Tax Surcharge Rate. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*).

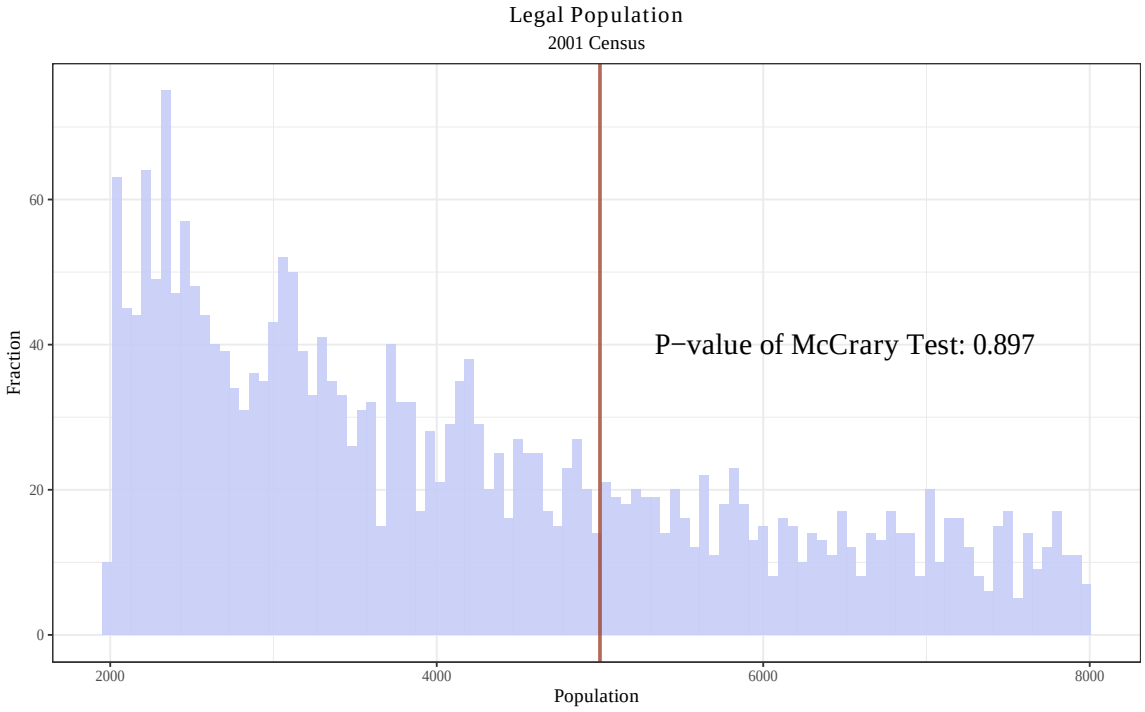
Figure 2.9. Discontinuity Estimates on Road Safety



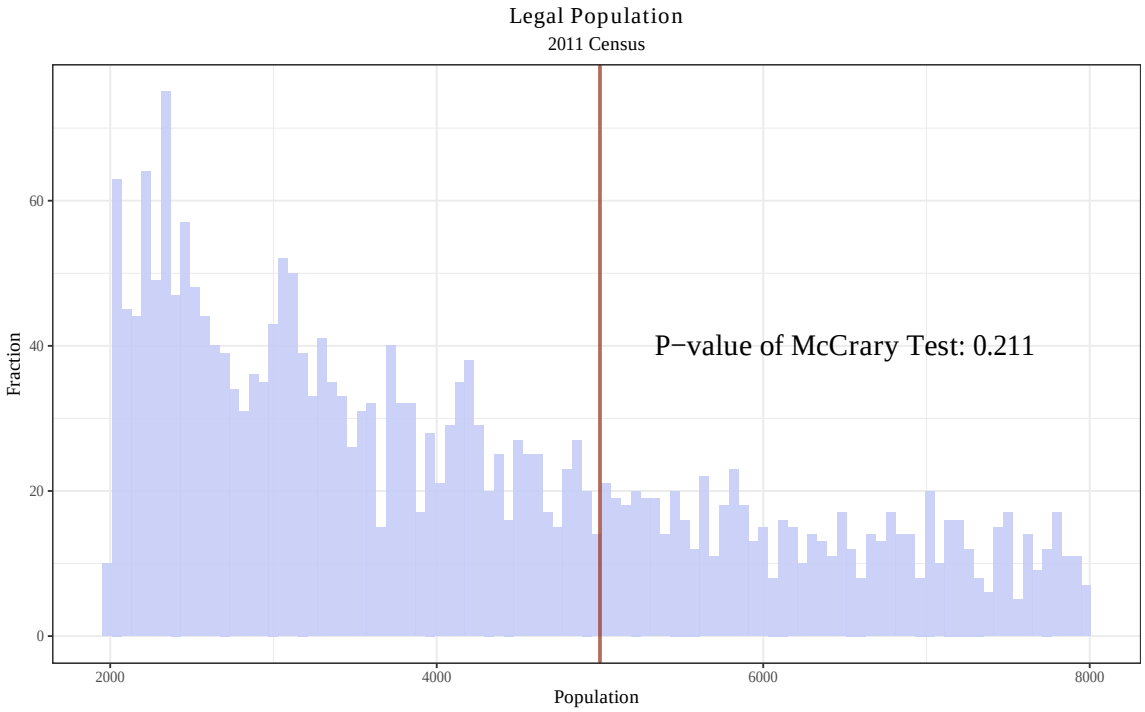
Note. Panels (A) and (B) plot the discontinuity coefficients from equations (2.1) and (2.2) using the rate of car accidents per 1,000 inhabitants as the dependent variable. Panels (C) and (D) use the injury rate as the dependent variable, while panels (E) and (F) use the death rate as the dependent variable. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*).

Figure 2.10. Threshold Manipulation Test

(a) 2001 Population



(b) 2011 Population



Note. The graph represent the distribution of legal population based on the 2001 and 2011 census, and report p-value for a manipulation test following McCrary (2008*a*).

2.7.2 Tables

TABLE 2.1
SUMMARY STATISTICS

Variable	Mean	Std. Dev.	N
Budget Variables			
Current Transfer from Gov't, euros per capita	183.685	115.774	56841
Total Current Transfer, euros per capita	245.084	229.047	62754
Total Revenues, euros per capita	1373.68	1422.281	56840
Fines Data			
Traffic Fines Imposed, euros per capita	10.99	51.495	56840
Traffic Fines Collected, euros per capita	7.049	28.95	56840
Traffic Fines Ratio (collected/imposed)	0.793	0.31	56840
Traffic Fines Share of Revenues	0.008	0.023	56838
Accidents Data			
Accidents (per 1k inhabitants)	1.926	1.988	62067
Injured in car accidents (per 1k inhabitants)	2.93	3.187	62067
Deaths in car accidents (per 1k inhabitants)	0.099	0.243	62067
Political variables			
Average Income, euros	11161.484	3224.965	57618
Female Mayor	0.102	0.303	52647
Margin of Victory at previous election	18.292	15.466	50429
Mayor Party Left	0.148	0.355	52493
Mayor Party Right	0.087	0.282	52493

Table 2.2
Discontinuity Estimates on Government Transfers

Specification	Dependent Variable: Government Transfers (euro per capita)					
	2009	2010	2011	2012	2013	2014
Cross-Section RD	-3.34	7.03	-12.07	-11.31	-28.75	-8.42
	(10.34)	(9.56)	(13.89)	(13.67)	(13.99)	(12.25)
h	1820	1840	1868	1563	1789	1911
N	827	839	853	667	785	772
Diff-in-Disc (2009)	-	-0.11	-19.26	-24.02	-37.58	-11.22
	-	(6.10)	(12.50)	(11.44)	(12.35)	(14.26)
h	-	1636	1906	1809	1934	1705
N	-	720	878	810	871	658
Diff-in-Disc (2010)	9.93	-	-20.06	-23.31	-39.69	-13.85
	(6.91)	-	(12.28)	(11.07)	(12.48)	(12.45)
h	1510	-	1695	1585	1843	1866
N	650	-	746	672	807	737
Diff-in-Disc (2009-2010 avg)	4.96	-0.05	-20.46	-24.58	-38.6	-12.34
	(3.46)	(3.05)	(11.89)	(10.90)	(12.06)	(13.04)
h	1510	1636	1842	1734	1945	1786
N	650	720	837	761	869	696

Note. Table reports discontinuity coefficients from equations (2.1) and (2.2) with robust standard errors in parentheses. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth h is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*), N represents the number of observations used.

Table 2.3
Discontinuity Estimates on Traffic Fines

Specification	Dependent Variable: Traffic Fines (accrual, euro per capita)					
	2009	2010	2011	2012	2013	2014
Cross-Section RD	0.91	0.39	-2.24	2.9	8.67	9.75
	(3.65)	(4.16)	(3.21)	(3.91)	(4.04)	(4.55)
h	1653	1982	1882	1400	1604	1950
N	727	928	864	594	689	785
Diff-in-Disc (2009)	-	-0.88	-3.76	1.8	4.11	6.62
	-	(3.11)	(2.94)	(3.34)	(2.99)	(4.52)
h	-	1870	1714	1653	1603	1645
N	-	851	762	726	685	633
Diff-in-Disc (2010)	0.77	-	-2.01	-1.83	2.75	9.58
	(3.36)	-	(2.50)	(4.68)	(3.24)	(4.24)
h	1900	-	1492	1389	1995	1986
N	876	-	643	585	902	797
Diff-in-Disc (2009-2010 avg)	2.6	2.01	0	-1.36	4.67	7.69
	(3.35)	(2.50)	(0.00)	(3.50)	(1.93)	(3.25)
h	1726	1492	0	1100	1641	1741
N	775	643	0	439	708	674

Note. Table reports discontinuity coefficients from equations (2.1) and (2.2) with robust standard errors in parentheses. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth h is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*), N represents the number of observations used.

Table 2.4
Discontinuity Estimates on Total Expenditures

Specification	Dependent Variable: Total Current Expenditures					
	2009	2010	2011	2012	2013	2014
Cross-Section RD	-181.66	-224.46	-40.97	11.28	130.07	56.22
	(137.50)	(127.89)	(116.72)	(114.92)	(95.07)	(139.17)
h	1552	1234	1095	1911	1924	1397
N	680	518	452	869	865	522
Diff-in-Disc (2010)	125.47	-	125.19	286.75	124.77	153.06
	(92.13)	-	(82.68)	(138.30)	(95.21)	(125.18)
h	1249	-	1828	1150	1763	1712
N	522	-	832	467	761	659
DiD	-	-24.76	-0.37	32.71	144.75	59.55
	-	(60.68)	(60.61)	(60.54)	(60.78)	(62.58)
h	1000					
N	3979					
Yearly LPR	-109.51	-104.55	-78.8	-51.87	69.81	128.99
	(90.05)	(78.59)	(70.91)	(81.83)	(99.68)	(92.98)
h	1000	1000	1000	1000	1000	1000
N	682	677	679	675	667	599

Note. Table reports discontinuity coefficients from equations (2.1) and (2.2), and results from diff-in-diff and LPR of a second degree. Robust standard errors in parentheses. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth h is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*), or set equal to 1,000. N represents the number of observations used.

Table 2.5
Traffic Fines and Taxes

Sample Restriction	Dependent Variable: Traffic Fines					
	2009	2010	2011	2012	2013	2014
Top 25% Tax	3.72	1.61	0.37	11.51	16.14	22.76
	(5.38)	(8.48)	(6.13)	(6.76)	(7.80)	(10.44)
h	2130	1509	1511	1569	1457	1958
N	390	256	257	265	244	288
Top 20% Tax	0.18	-1.91	0.49	20.4	32.58	59.07
	(10.09)	(10.91)	(11.14)	(15.12)	(16.57)	(23.28)
h	1930	1403	1385	1332	1286	1379
N	171	111	109	103	99	96
Top 10% Tax	3.72	10.81	6.49	24.19	24.21	58.44
	(10.28)	(12.43)	(10.44)	(25.71)	(19.74)	(44.15)
h	1861	1354	1979	1471	1908	1487
N	75	51	80	58	71	48

Note. Table reports regression discontinuity coefficients from equations (2.1) and (2.2) with robust standard errors in parentheses. Subgroups are based on the 2010 distribution of the local Income Tax Surcharge Rate. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth h is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*), N represents the number of observations used.

Table 2.6
Discontinuity Estimates on Road Safety

Specification	Years					
	2009	2010	2011	2012	2013	2014
Panel A. Dependent Variable: Accidents Rate (per 1,000 inhabitants)						
Cross-Section RD	0.02	0.02	-0.4	-0.35	0.34	0.48
	(0.23)	(0.24)	(0.25)	(0.28)	(0.21)	(0.26)
h	1966	2185	1715	1442	2012	1505
N	927	1040	756	607	923	613
Diff-in-Disc (2010)	-0.01	-	-0.4	-0.28	0.06	0.11
	(0.22)	-	(0.18)	(0.21)	(0.17)	(0.17)
h	1656	-	1686	1614	2060	1500
N	720	-	738	694	947	609
Panel B. Dependent Variable: Injury Rate (per 1,000 inhabitants)						
Cross-Section RD	0	-0.06	-0.76	-0.34	0.71	0.84
	(0.35)	(0.37)	(0.38)	(0.44)	(0.32)	(0.38)
h	1942	2175	1827	1451	2006	1538
N	910	1035	825	610	919	626
Diff-in-Disc (2010)	-0.18	-	-0.7	-0.35	0.34	0.29
	(0.40)	-	(0.30)	(0.37)	(0.30)	(0.30)
h	1573	-	1719	1509	2039	1705
N	676	-	758	633	934	722
Panel C. Dependent Variable: Death Rate (per 1,000 inhabitants)						
Cross-Section RD	-0.01	-0.02	0.01	0.05	0	0.01
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
h	1561	1732	1664	2174	1922	2217
N	682	775	727	1018	870	1017
Diff-in-Disc (2010)	0.01	-	0.03	0.09	0.03	0.04
	(0.04)	-	(0.03)	(0.03)	(0.03)	(0.04)
h	2017	-	1597	2168	2171	1940
N	939	-	689	1015	1016	858

Note. Table reports discontinuity coefficients from equations (2.1) and (2.2) with robust standard errors in parentheses. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth h is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*), N represents the number of observations used.

Table 2.7
Political Cycle – OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Traffic Fines									
	Fines Imposed			Fines Collected			Fines Ratio		
Election Year	-0.0792	-0.266	-0.569*	-0.197	-0.290	-0.475**	-0.0155***	-0.0139***	-0.0108***
	(0.298)	(0.283)	(0.304)	(0.195)	(0.187)	(0.216)	(0.00213)	(0.00211)	(0.00235)
Observations	75,602	75,602	75,602	75,601	75,601	75,601	75,600	75,600	75,600
R-squared	0.012	0.035	0.036	0.014	0.037	0.037	0.023	0.054	0.055
Controls	X	X	X	X	X	X	X	X	X
Year FE	.	.	X	.	.	X	.	.	X
Province FE	.	X	X	.	X	X	.	X	X
Panel B. Road Safety									
	Accidents Rate			Death Rate			Injury rate		
Election Year	0.0786***	0.0776***	0.0365**	0.00566	0.00580	0.00189	0.118***	0.115***	0.0493*
	(0.0149)	(0.0147)	(0.0152)	(0.00445)	(0.00444)	(0.00399)	(0.0256)	(0.0253)	(0.0268)
Observations	74,780	74,780	74,780	74,780	74,780	74,780	74,780	74,780	74,780
R-squared	0.100	0.145	0.166	0.001	0.011	0.014	0.065	0.100	0.119
Controls	X	X	X	X	X	X	X	X	X
Year FE	.	.	X	.	.	X	.	.	X
Province FE	.	X	X	.	X	X	.	X	X

Note. Standard Errors clustered at the municipality level are reported in parenthesis. Significance level: ***<0.01, **<0.05, *<0.1. Controls include: population, population squared, per capita income, gender of the mayor, victory margin, party of the mayor, transfers from the central government.

Table 2.8
Political Cycle – TSLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. TSLS Results									
	Accidents Rate			Injury Rate			Death Rate		
Fines Imposed	-0.108			-0.156			-0.007		
	(0.0731)			(0.111)			-0.009		
Fines Collected		-0.119*			-0.171*			-0.008	
		(0.0671)			(0.104)			-0.010	
Fines Ratio			-3.909***			-5.612**			-0.276
			(1.397)			(2.367)			-0.317
Observations	70,038	70,037	70,036	70,038	70,037	70,036	70,038	70,037	70,036
Controls	X	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X	X
Province FE	X	X	X	X	X	X	X	X	X
Panel B. First Stage Statistics and Weak Instrument Inference									
KP F-Stat	2.75	4.74	29.22						
MP F-Stat	2.75	4.74	29.22						
AR 95% Lower Bound	(-Inf)	(-Inf)	-7.39	(-Inf)	(-Inf)	-11.33	(-Inf)	(-Inf)	-0.967
AR 95% Upper Bound	-0.035	-0.042	-1.53	-0.037	-0.044	-1.393	(Inf)	-0.133	-0.338

Note. Standard Errors clustered at the municipality level are reported in parenthesis. Significance level: ***<0.01, **<0.05, *<0.1. Controls include: population, population squared, per capita income, gender of the mayor, victory margin, party of the mayor, transfers from the central government.

Table 2.9
Balance Table

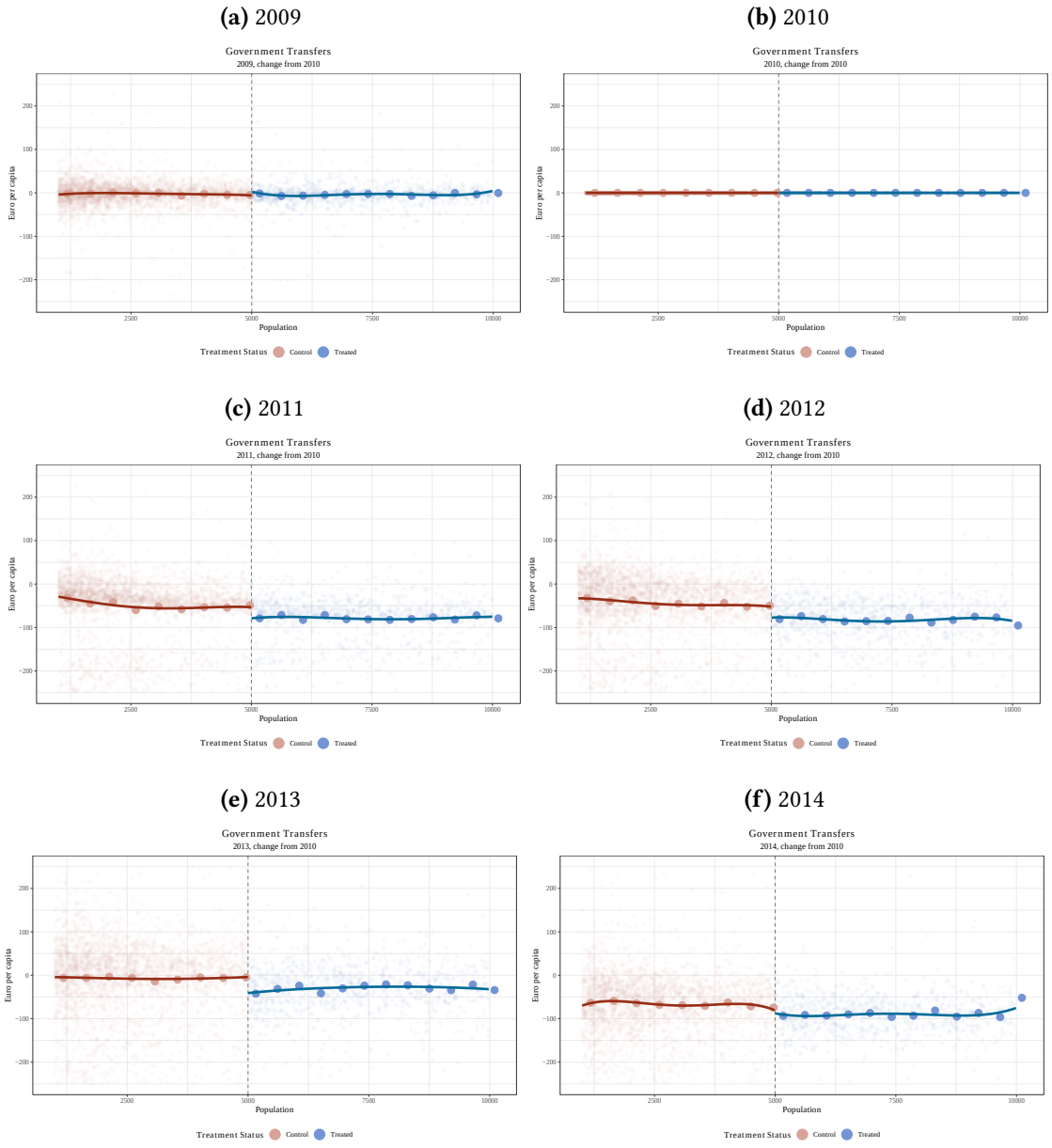
	Year				
Dependent Variable	2009	2010	2011	2012	2013
Income	-70.4	-828.06	-811.48	-1281.32	-907.71
	(654.42)	(555.11)	(584.59)	(485.34)	(478.54)
p-value	0.91	0.14	0.17	0.01	0.06
Female Mayor	0.01	-0.13	-0.1	-0.07	-0.13
	(0.05)	(0.04)	(0.05)	(0.06)	(0.05)
p-value	0.88	0	0.04	0.25	0.01
Victory Margin	-1.6	-0.69	0.06	-2.26	0.1
	(3.13)	(2.50)	(2.38)	(2.56)	(2.42)
p-value	0.61	0.78	0.98	0.38	0.97
Mayor Left Party	-0.04	-0.04	-0.01	0.07	0.06
	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
p-value	0.36	0.29	0.67	0.04	0.04
Mayor Right Party	0.05	0.05	0.01	-0.02	0.1
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
p-value	0.42	0.42	0.83	0.76	0.1

Note. Table reports RD coefficients from equation (2.1) with robust standard errors in parentheses. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth h is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*). The p-value tests difference at the 95% confidence level.

2.A Appendix

2.A.1 Additional Figures

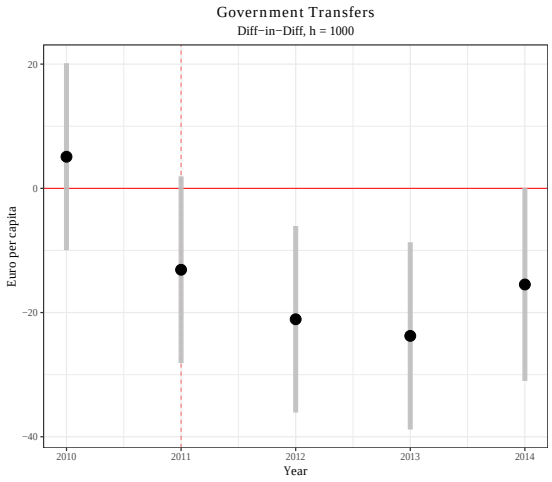
Figure 2.A.1. Difference in Government Transfers (2009-2014)



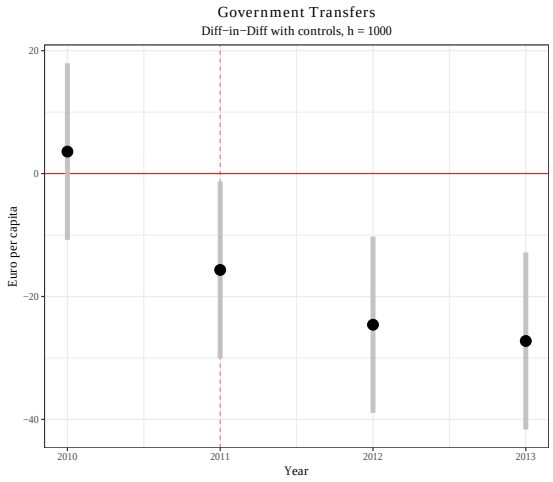
Note. Panels (a) to (f) plot the difference between grants received by each municipality in each year and the 2010 value. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants. Darker dots represent bin scatters.

Figure 2.A.2. Government Transfers – Alternative Specifications

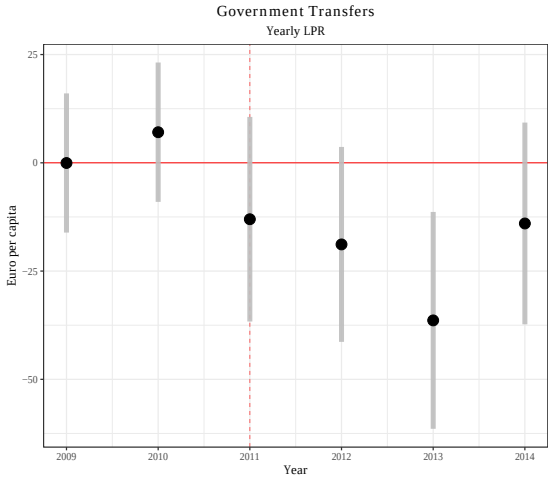
(a) Difference in Difference



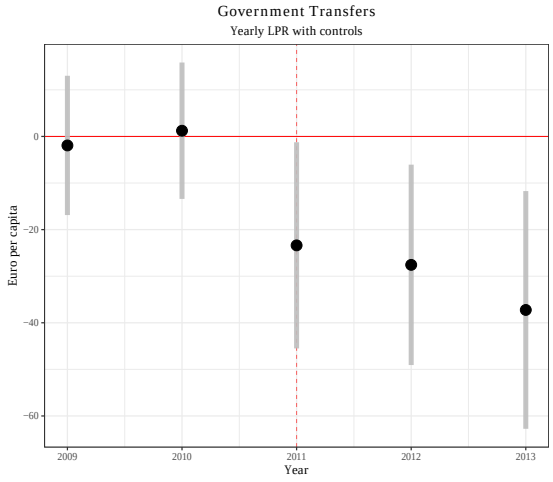
(b) Difference in Difference with Controls



(c) Local Polynomial Regression



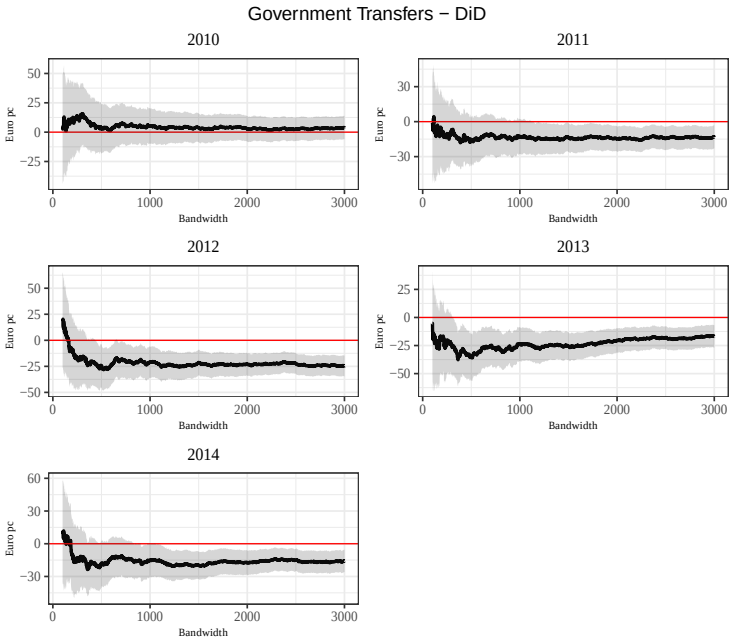
(d) Local Polynomial Regression with Controls



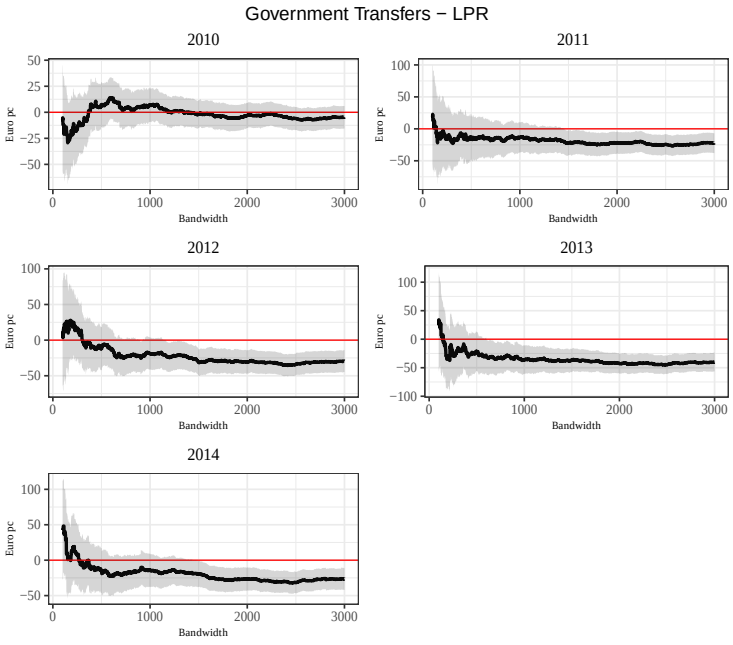
Note. Panels (a) and (b) plot Diff-in-Diff coefficients from Table (2.A.1), Panels (c) and (d) results from a second degree Local Polynomial Regression with. Bandwidth is 1,000 inhabitants on both sides. Controls include population, population squared, per capita income, gender of the mayor, victory margin, party of the mayor.

Figure 2.A.3. Government Transfers – Bandwidth Sensitivity

(a) Difference in Difference



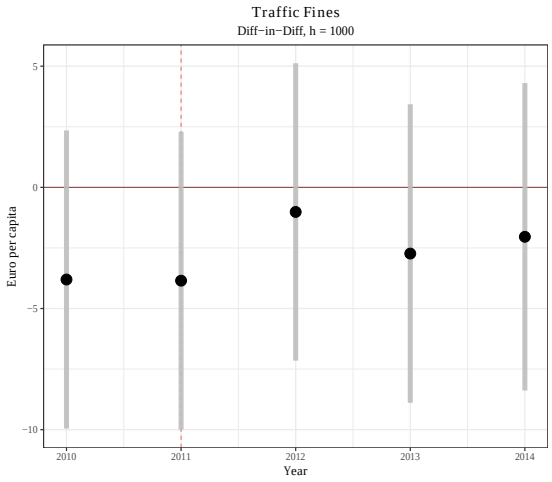
(b) Local Polynomial Regression



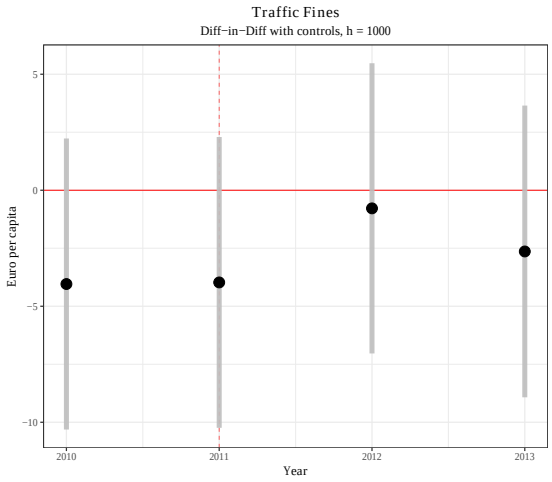
Note. Figure plots Diff-in-Diff and LPR coefficients constructed as in Figure (2.A.4), allowing for multiple values of the bandwidth.

Figure 2.A.4. Traffic Fines – Alternative Specifications

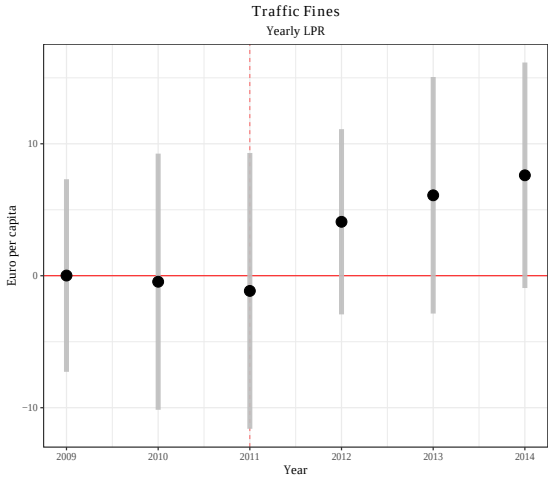
(a) Difference in Difference



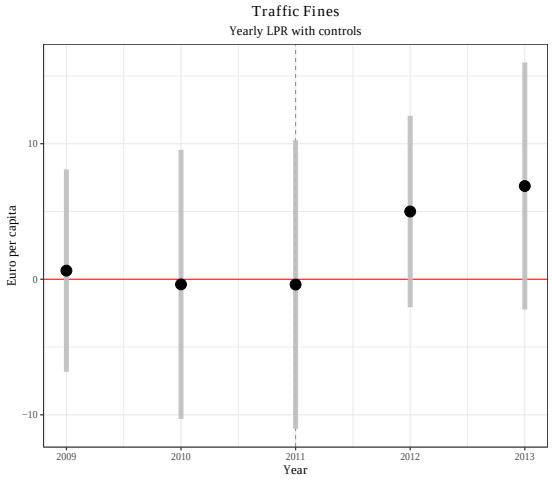
(b) Difference in Difference with Controls



(c) Local Polynomial Regression



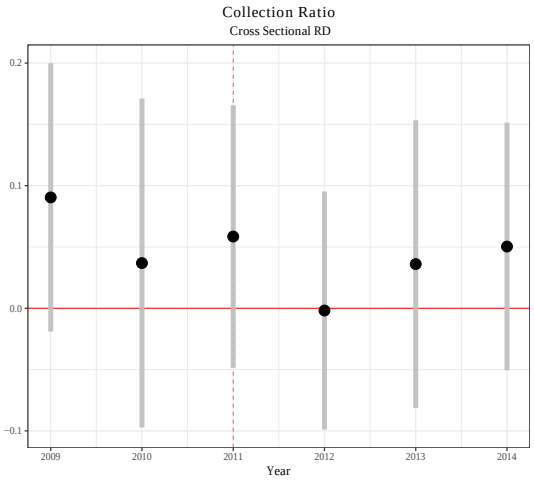
(d) Local Polynomial Regression with Controls



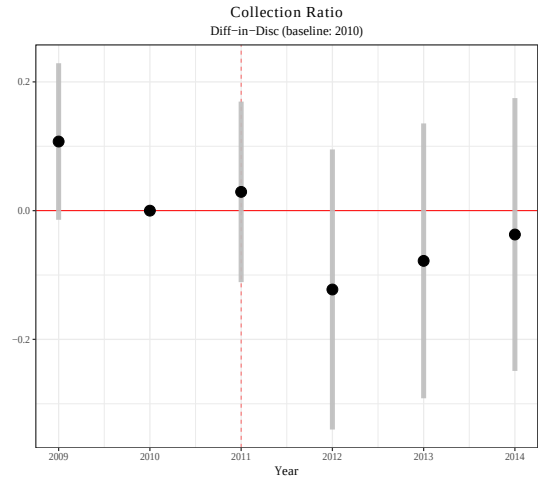
Note. Panels (a) and (b) plot Diff-in-Diff coefficients from Table (2.A.2), Panels (c) and (d) results from a second degree Local Polynomial Regression with. Bandwidth is 1,000 inhabitants on both sides. Controls include population, population squared, per capita income, gender of the mayor, victory margin, party of the mayor.

Figure 2.A.5. Traffic Fines – Collection Ratio

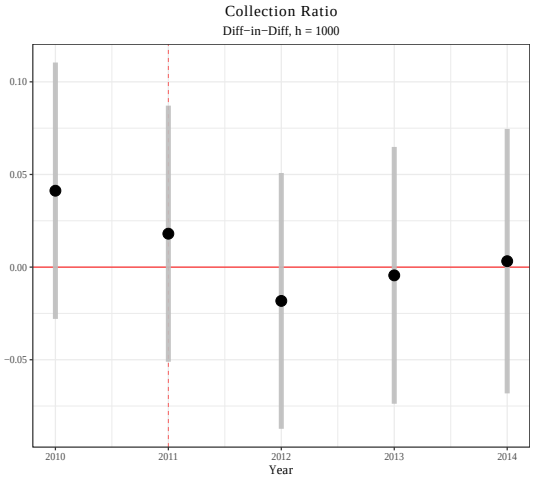
(a) Regression Discontinuity



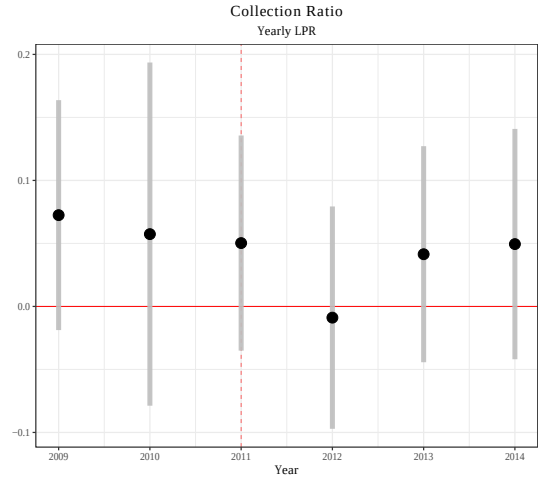
(b) Diff-in-disc (2010)



(c) Difference in Difference



(d) Local Polynomial Regression



Note. Panels (a) and (b) plot Diff-in-Diff coefficients from Table (2.A.3), Panels (c) and (d) results from a second degree Local Polynomial Regression with. Bandwidth is 1,000 inhabitants on both sides. Controls include population, population squared, per capita income, gender of the mayor, victory margin, party of the mayor.

2.A.2 Additional Tables

Table 2.A.1
Government Transfers – Alternative Specifications

Specification	Dependent Variable: Government Transfers (accrual, euro per capita)					
	2009	2010	2011	2012	2013	2014
Difference in Difference	-	5.09	-13.11	-21.08	-23.76	-15.48
	-	(7.68)	(7.67)	(7.66)	(7.69)	(7.92)
h	1000					
N	3979					
DiD w/ controls	-	3.59	-15.67	-24.59	-27.23	-
	-	(7.33)	(7.33)	(7.31)	(7.35)	-
h	1000					
N	3318					
Yearly LPR	-0.04	7.07	-13.03	-18.84	-36.38	-14
	(8.18)	(8.20)	(12.03)	(11.46)	(12.75)	(11.86)
h	1000	1000	1000	1000	1000	1000
N	682	677	679	675	667	599
Yearly LPR w/ controls	-1.93	1.23	-23.38	-27.56	-37.24	-
	(7.61)	(7.46)	(11.27)	(10.95)	(12.99)	-
h	1000	1000	1000	1000	1000	-
N	666	665	668	666	653	-

Note. Robust standard errors in parenthesis. Bandwidth h is equal to 1,000 inhabitants on both sides of the threshold. N represents the observation used in the estimation. Local Polynomial Regressions include a second degree polynomial in population with intereaction. Additional controls include: per capita income, gender of the mayor, victory margin, party of the mayor.

Table 2.A.2
Traffic Fines – Alternative Specifications

Specification	Dependent Variable: Traffic Fines (accrual, euro per capita)					
	2009	2010	2011	2012	2013	2014
Diffrence in Difference	-	-3.8	-3.85	-1.01	-2.73	-2.04
	-	(3.14)	(3.13)	(3.13)	(3.14)	(3.24)
h	-	1000	1000	1000	1000	1000
N	-	3979	3979	3979	3979	3979
DiD w/ controls	-	-4.04	-3.97	-0.78	-2.64	-
	-	(3.20)	(3.20)	(3.19)	(3.21)	-
h	1000					
N	3318					
Yearly LPR	0.02	-0.45	-1.15	4.09	6.1	7.61
	(3.72)	(4.94)	(5.32)	(3.57)	(4.56)	(4.35)
h	1000	1000	1000	1000	1000	1000
N	682	677	679	675	667	599
Yearly LPR w/ controls	0.64	-0.37	-0.39	5	6.88	-
	(3.80)	(5.05)	(5.42)	(3.59)	(4.64)	-
h	1000	1000	1000	1000	1000	
N	666	665	668	666	653	

Note. Robust standard errors in parenthesis. Bandwidth h is equal to 1,000 inhabitants on both sides of the threshold. N represents the observation used in the estimation. Local Polynomial Regressions include a second degree polynomial in population with intereaction. Additional controls include: per capita income, gender of the mayor, victory margin, party of the mayor.

Table 2.A.3
Traffic Fines – Collection Ratio

Specification	Dependent Variable: Fines Collection Ratio					
	2009	2010	2011	2012	2013	2014
Cross-Section RD	0.09	0.04	0.06	0	0.04	0.05
	(0.06)	(0.07)	(0.05)	(0.05)	(0.06)	(0.05)
h	1476	1761	1695	2205	1586	2029
N	634	797	748	1055	672	821
Diff-in-Disc (2010)	0.11	-	0.03	-0.12	-0.08	-0.04
	(0.06)	-	(0.07)	(0.11)	(0.11)	(0.11)
h	1277	-	1550	2354	2203	2465
N	539	-	667	1143	1025	1082
DiD	-	0.04	0.02	-0.02	0	0
	-	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
h	1000					
N	3979					
Yearly LPR	0.07	0.06	0.05	-0.01	0.04	0.05
	(0.05)	(0.07)	(0.04)	(0.04)	(0.04)	(0.05)
h	1000	1000	1000	1000	1000	1000
N	682	677	679	675	667	599

Note. Table reports regression discontinuity coefficients from equations (2.1) and (2.2) with robust standard errors in parentheses, diff-in-diff estimates and a Local Polynomial Regression of a second degree. Sample is restricted to municipalities between 1,000 and 10,000 inhabitants, bandwidth h is estimated separately on each year following Calonico, Cattaneo and Titiunik (2014*b*), or set equal to 1,000. N represents the number of observations used.