Centre for European Economics Research

Research Proposal Summer School Revealed Preferences

Revealed Preferences under Framing: Users valuation of privacy

Johannes Walter

Abstract

Diese Dokumentation enth"alt eine sortierte Liste der wichtigsten LaTeX–Befehle. Die einzelnen Listeneintr"age sind untereinander durch viele Querverweise verkettet, die ein Auffinden inhaltlich zusammengeh"origer Informationen erheblich erleichtern.

Summer School Revealed Preferences
Nicolai Kuminoff

July 12, 2020

1 Introduction

In stated preference studies people tend to express a high valuation for privacy, yet observed behavior is typically at odds with stated preferences. This phenomenon has come to known in the literature as privacy paradox. Some have argued that this observation is merely illusory, since one can state a high valuation of privacy in general, but in a specific situation a cost-benefit analysis might lead people to discount privacy concerns (Acquisti, Brandimarte, and Loewenstein 2015, p. 2). This refutation falls short insofar, as it is a well documented result in behavioral economic research that people' decision making capabilities are only in part rational. Especially in situations when people are uncertain about the consequences of their actions, when they are unsure about their preferences or when they are under time pressure, people often search for clues in their surrounding to provide guidance (Acquisti, Brandimarte, and Loewenstein 2015, p. 3). One possible source of orientation can stem from the specific way in which a decision is framed. Deriving revealed preferences in the presence of framing can be problematic, if people's preference is not consistent over the set of framings. Goldin and Reck (2020) propose a new method to recover consistent population preferences even when framing effects are present.

Their methodology rests on the insight that even though framing-consistent preferences cannot be observed on the *individual level* in typical datasets, on the *population level* the fraction of consistent decision-makers can be identified. From the subgroup of consistent decision-makers their approach extrapolates to the entire population. Their method is described in detail in section 2.

This research draft proposes two innovations. First, it is proposed to collect an online dataset recording people's decisions about browser cookie settings via a type of data collection akin to the one in Levy (2020). Goldin and Reck (2020)'s method can subsequently be applied to the analysis of this dataset. The second innovation is methodological in nature. The dataset collected online will contain repeated observations of the same individual both under an ad-tracking cookie opt-in framing and an ad-tracking cookie opt-out framing. This additional information can be used to overcomes the central limitation in most datasets. This means that the necessity for Goldin and Reck (2020)'s model vanishes. Instead, the subset of framing-consistent people can be observed on the individual level. Once these individuals are identified, one can again extrapolate to the entire population, as is done in Goldin and Reck (2020)'s method. Finally, the results from the first approach and the second approach can be compared, yielding an empirical test of Goldin and Reck (2020)'s method.

What would ideal results of this research look like? If the proposed research succeeds it find whether Goldin and Reck (2020)'s new method generates estimates that are qualitatively similar and quantitatively close to estimates based on repeated observations of the same individual. By doing so, it will produce estimates of people's privacy preferences which are robust with respect to framing. Here two scenarios are conceivable: Either it will be found that From a policy point of view, the latter result could possibly be interpreted to imply that the rules governing cookie default settings would have to be further tightened, such that websites default setting is to only use technically necessary cookies. Ad-tracking cookies would then mandatorily be opt-in.

This research lies in the intersection of three strands of literature, each of which it will attempt to contribute to. First, it will contribute to the literature on measuring preferences for and attitudes towards privacy. Both in the social and behavioral sciences There is a large body of literature that examines the nature of privacy preferences. This research will contribute by uncovering preferences for online privacy settings in particular. Second, by carrying Goldin and Reck (2020)'s model into a new setting and confronting it with an empirical test, this research will contribute to the literature on revealed preferences under framing. And finally, there is a growing literature on the effects of the GDPR. The change in laws regarding cookie settings is what enables this research in the first place. The results of this study will have practical implications for the question whether even stricter privacy regulations seem justified.

2 Economic Model

The description of this model follows chapter one of Acquisti, Brandimarte, and Loewenstein (2015). A fully detailed description of the model can be found there and is beyond the scope of this research proposal. Here, only the main concepts necessary for understanding this proposal are presented.

2.1 Basic model notation

A decision maker i chooses from a binary decision set $\mathbf{S} = \{0, 1\}$ und two possible frames $D_i \in \{0, 1\}$. In the context of this research proposal, the decision coded with a 0 could refer to an individual's decision to accept only technically necessary cookies and the decision coded with a 1 could refer to an individual's decision to allow for non technically necessary cookies. Frame 0 could be the situation where the default setting is such that technically necessary cookies are the only pre-chosen ones and the user would have to actively engage in clicking on all non-technically necessary cookies she wishes to allow. Frame 1 could then refer to the situation where in the default setting both types of cookies are pre-chosen and the user accepts both types with one click. We can adapt a notation that is akin to what many researchers like to use in a potential outcome setting. $Y_i(0)$ and $Y_i(0)$ denotes then the decision individual i makes under frame $D_i = 0$ and $D_i = 1$, respectively. Decision makers are assumed to have strict ordinal preferences over the set of available options. $Y_i^* \in \{0, 1\}$ denotes the most preferred option. Each decision maker is characterized by a vector of random variables $(Y_i(0), Y_i(1), D_i, Y_i^*)$, which are drawn from an underlying population distribution. For each i, the researcher observes the pair (Y_i, D_i), where $Y_i = Y_i(0)D_i + Y_i(1)(1 - D_i)$. Goldin and Reck (2020) assume for their model that the researcher does not observe Y_i^* and only observes one of $Y_i(0)$ and $Y_i(1)$, depending on the frame D_i.

The data that is suggested to be collected in this proposal deviates in a crucial way from this assumption. In contrast to the datasets Goldin and Reck (2020) have in mind, the dataset in this proposal will contain *repeated* observations of the same individual. The number of observations for each i can be denoted with a subscript $k \in \{1, \ldots, K\}$ such that $Y_{i,k}$ is observed. Importantly, here it is assumed that K is sufficiently large to allow for at least one observation of each framing $D_i \in \{0, 1\}$.

The mean choices among decision makers assigned to a frame is denoted by $\bar{Y}(1) \equiv E[Y_i|D_i = 1]$ and $\bar{Y}(0) \equiv E[Y_i|D_i = 0]$.

Goldin and Reck (2020) continue their description of the model as follows: Each decision maker can choose either consistently, i.e. the same choice under each frame, or choose in a way that is responsive to the frame. Consistency is denoted by $C_i = 1$ iff $Y_i(0) = Y_i(1)$ else $C_i = 0$. Again, they assume that each i is observed only under one frame, such that C_i is not observed. In the context of this proposal, as described above, it is assumed that each i is observed under both frames, such that C_i is identified. This difference is crucial for one of the main contributions of this proposal. The additional information in this dataset allows to calculate two sets of results. For the first one, the additional information in this dataset is not used. The analysis will proceed under the assumption of Goldin and Reck (2020). The second set of results will be obtained using the full amount of information in the dataset. As such, the "ground-truth" that remains hidden under Goldin and Reck (2020)'s assumptions, is observed. In other words, with the dataset proposed here, it will be able to identify all individuals who make consistent choices, independent of the frame. Having these two sets of results, one can compare the estimates from both to arrive at an assessment of the quality of Goldin and Reck (2020)'s model accuracy.

2.2 Model assumptions

One of Goldin and Reck (2020)'s main contributions is to make the necessary assumptions for their analysis explicit. Furthermore, the assumptions are fundamental to their approach and will be required in section 3 to derive framing-consistent estimates for the entire population. For these reasons, the assumptions will be presented shortly here as well. Each assumption is first presented in the way Goldin and Reck (2020) lay them out and subsequently followed by a comment on how each assumption relates to the approach suggested in this research proposal.

- Assumption A1) *Frame separability*: For all i, Y_i^* does not depend on D.

For each individual, the most preferred option does not depend on the framing. This is an assumption about the content of a decision makers' preferences and is useful to define which features of the environment are considered to be part of the framing (Goldin and Reck 2020, p. 2764).

- Assumption A2) *Frame exogeneity*: $(Y_i(0), Y_i(1), Y_i^*) \perp\!\!\!\perp D_i$.

This assumption refers to the data generating process by which decision makers are assigned to frames. It is similar in nature to the assumption in the potential outcome setting that says treatment and assignment to treatment need to be independent. A2) makes sure that observed differences are due to the effect of the frames, rather than due to differences in the groups of individuals assigned to each frame.

- Assumption RPA) *Revealed-preference assumption*: For all $i, Y_i^* = Y_i$.

In Goldin and Reck (2020)'s setting, a framing effect is observed when assumptions A1) and A2) are satisfied and one observes $\bar{Y}(1) \neq \bar{Y}_i$. In the context of this research proposal, a framing effect occurs if $Y_{i,k}(0) \neq Y_{i,\neg k}(1)$ for at least one k.

- Assumption A3) *Consistency Principle*: For all $i, C_i = 1 \Rightarrow Y_i = Y_i^*$.

This assumption tells us that preferences are only guaranteed to be revealed by choices for those decision makers who choose consistently across frames. This assumption also applies in the context of the data that this research proposal suggests.

- Assumption A4) *Frame monotonicity*: For all $i, Y_i(1) \geq Y_i(0)$.

Assumption A4) allows Goldin and Reck (2020) to recover aggregate information about consistency despite the fact that C_i is not observable on the individual level in their setting. The assumption requires that when a frame does affect choices, it does so in the same direction for each affected decision maker. Since C_i can be observed directly in the setting of this proposal, A4) is not necessary for the analysis of this dataset.

3 Data and Methods

Subsection 3.1 describes how the data is proposed to be collected and the kind of data that will be collected. It will also contain a short overview of browser cookies and their functionality. Section 3.2 will present how Goldin and Reck (2020) identify their parameters of interest and the simplifications that are possible when one has access to a richer dataset.

3.1 Data

The process of data collection is inspired by Levy (2020). In this paper, the author wrote a browser extension for Google's Chrome browser. Via a Facebook ad campaign, 8084 participants for a survey were attracted. At the end of the survey, they were offered a small reward in exchange of installing the browser extension on their device (Levy 2020, p. 9). 2262 survey participants installed the extension, and 1835 kept the extension installed for at least two weeks.

This research proposal suggests to apply the same design. Participants can be attracted via a Facebook or Google ad campaign. In a short survey, their age, gender and nationality can be elicited. Afterwards, they are asked to install a small browser extension in exchange for a small reward. The extension will use a short javascript code that writes a log file. This log file will keeps track whether a participant has visited a website that asks for cookie settings. It will also log whether the default setting is to accept all cookies, or whether by default only the technically necessary cookies are selected. Importantly, the extension needs not collect any information about the content of the visited websites. From a privacy point of view, this design is less invasive than the one in Levy (2020). This observation motivates the believe that there is a sufficient amount of people who are willing to participate in such a study.

A hypertext transfer protocol (HTTP) cookie is a small piece of data that is sent from a website and stored on a users device. For the purposes of this research proposal, cookies can be classified into two groups. First, there are technically necessary cookies that ensure a website is working as intended. Such cookies could remember preceding user interactions with the site, e.g. adding an item to a shopping cart, or logging in. Second, there are cookies that are not technically necessary to ensure the website's functionality. A common example are for example third-party tracking cookies which record previously visited websites. These non-essential third-party tracking cookies can generate long-term records of a user's browsing history and can therefore manifest a potential privacy concern. Websites have an incentive to allow these cookies, as they receive third-party payments in return (common tracking cookies are e.g. Google AdSense or Facebook remarketing). The GDPR requires that all websites which target European Union member states ask their users for consent before they are allowed to store non-essential cookies on a user's device. The host of a website still has a choice on how to frame this question. Either non-essential cookies are unselected by default and if the user clicks once on "accept" no tracking is possible. Or non-essential cookies are selected by default and if the user wants to avoid them it is necessary to go into the website's privacy settings and unselect tracking cookies manually.

3.2 Method

This section will describe how Goldin and Reck (2020) identify their model parameters. It will also contain a description of how the parameter identification is simplified by the method proposed here. To keep this section concise and within the scope of this proposal, only the main ideas are introduced here. For a detailed discussion, see Goldin and Reck (2020, p.2767).

In the first step, Goldin and Reck (2020) focus on the consistent decision makers, i.e. the subgroup of individuals whose decision does not depend on the frame. With a dataset as suggested above, it is trivial to find this group. Once the consistent decision makers are found, using assumption A3), the consistency principle, the options they preferred can be found.

Goldin and Reck (2020)'s method is also applicable to datasets that do not contain repeated observations for each individual. It is worth noting, that Goldin and Reck (2020) mention that there method can also improve datasets that do contain repeated observations. In particular, this is the case when the *order* in which the repeated observations are generated can itself be considered a framing. For instance, this could be an issue if researchers were to analyse survey data, where the questions were not sufficiently randomized. For the dataset that is described in section 3.1 the order in which individuals see each question about cookies is, as described, assumed to follow a random process.

Goldin and Reck (2020) continue by describing the conditions for which consistent decision makers preferences can be identified when each decision maker is observed under a single frame.

- Proposition 1: $\bar{Y}_C \equiv \bar{Y}(0)/(\bar{Y}(0) + 1 - \bar{Y}(1))$
- Proposition 1.1: Under assumptions A1 - A4, $E[Y_i^*|C_i = 1] = \bar{Y}_C$.
- Proposition 1.2: Under assumptions A1 - A3, $\bar{Y}_C \geq 1/2 \Rightarrow \bar{Y}_C \leq E[Y_i^*|C_i = 1] \leq 1$, and $\bar{Y}_C \leq 1/2 \Rightarrow 0 \leq E[Y_i^*|C_i = 1] \leq \bar{Y}_C$.

Proposition 1.1 follows from the insight that, given frame monotonicity, only consistent decision-makers choose against the frame (Goldin and Reck 2020, p. 2768). Proposition 1.2 provides a partial identification result, as it defines upper and lower bounds for the fraction of frame consistent decision makers. The results under proposition 1.2 are robust to failures of frame monotonicity. In other words, it guarantees bounds in cases when there are decision makers who decide against the frame, but are still in truth inconsistent. These decision makers are defined by Goldin and Reck (2020, p. 2768) as frame defiers.

The logic of proposition 1 is illustrated in table 3.2, which is taken from Goldin and Reck (2020, p. 2769). To understand this table, it is important to know the conditions that produced the data. Here it is assumed that the researchers observe answers under both types of framing. Yet each individual's answer is only observed once under one type of framing.

TABLE 1
ILLUSTRATION OF PROPOSITION 1

	Choose Not to Enroll, Opt-In Regime, $Y_i(0) = 0$	Choose to Enroll, Opt-In Regime, $Y_i(0) = 1$
Choose not to enroll, opt-out regime, $Y_i(1) = 0$	.35	.00
Choose not to enroll, opt-out regime, $Y_i(1) = 1$	.25	.40
Fraction consistent	$E[C_i] = 0.4 + 0.35 = 0.75$	
Fraction of consistent decision makers preferring option 1	$E[Y_i^*|C_i = 1] = \bar{Y}_C = 0.40/(0.4 + 0.35) \approx 0.53$	
Bounds on consistent preferences, without assumption A4	$0.53 \leq E[Y_i^*|C_i = 1] \leq 1$	

NOTE.—Under frame monotonicity, top-right quadrant = 0; top-left quadrant = $1 - \bar{Y}(1)$; bottom-right quadrant = $\bar{Y}(0)$; bottom-left quadrant = $\bar{Y}(1) - \bar{Y}(0)$. $\bar{Y}_C$ is biased toward 1/2, as frame defiers are equally assigned to the top-left and bottom-right quadrants by frame exogeneity (by definition, they cannot be in the bottom-left quadrant).

To

the best of my understanding, the table contains a mistake. I believe that the description for the second row of the table should read "Choose to enroll, opt-out regime, $Y_i(1) = 1$" (instead in the version in the article, the description mistakenly contains a "not": "Choose not to enroll, opt-out regime, $Y_i(1) = 1$") Now, it becomes easy to see how the fraction of consistent decision makers can be found even if each decision makers is only observed under one frame. In the example in table 3.2, 0 % of observations fall into the top-right quadrant by assumption. The top-left quadrant is the result of $1 - \bar{Y}(1)$, where $\bar{Y}(1)$ is the fraction of decision makers who chose to enroll under the opt-out framing. This number is observed by the researchers. The bottom left quadrant is the difference between the fraction of decision makers who enroll under opt-out framing and the fraction of decision makers who enroll under opt-in framing. Both fractions are observed. The bottom right quadrant is the fraction of decision makers who enroll under opt-in framing, which is again, observed. Now the fraction of consistent decision makers who prefer option 1 can be calculated according to proposition 1. When assumption A4 is relaxed, at least a lower bound can be calculated. This way, despite the researcher not observing each decision maker under each framing, the preferences of the consistent decision makers can be recovered.

When a researcher has access to a dataset as proposed in this research proposal, the fraction of consistent decision makers preferring option 1 can be calculated directly and could be compared to the result from Goldin and Reck (2020)'s method (in table 3.2 approximately 53%). Thus, providing a way to check the method for its validity in a certain setting.

In the next step, Goldin and Reck (2020, p.2769) describe several methods to carry the preferences of the consistent decision makers over to the rest of the population.

References

Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein (2015). “Privacy and human behavior in the age of information”. In: *Science* 347.6221, pp. 509–514.

Goldin, Jacob and Daniel Reck (2020). “Revealed-Preference Analysis with Framing Effects”. In: *Journal of Political Economy* 128.7, pp. 000–000.

Levy, Ro’ee (2020). *Social Media, News Consumption, and Polarization: Evidence from a Field Experiment.* https://www.roeelevy.com/research. [Online; accessed 09-July-2020].