Three Studies in Empirical Applications of Microeconomic Theory

by
Sinan Ozel
BSc. EE, Istanbul Technical University, 2002
MBA, Koc University, Istanbul, 2004

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
In the Department of Economics

We acknowledge with respect the Lekwungen-speaking peoples on whose traditional territory the university stands and the Songhees, Esquimalt and WSÁNEĆ peoples whose historical relationships with the land continue to this day.

University of Victoria

Three Studies in Empirical Applications of Microeconomic Theory

by
Sinan Ozel
BSc. EE, Istanbul Technical University, 2002
MBA, Koc University, Istanbul, 2004

Supervisory Committee Members

Dr. Elisabeth Gugl, Co-Supervisor
(Department of Economics)

Dr. Pascal Courty, Co-supervisor
(Department of Economics)

Dr. Martin Farnham, Departmental Member
(Department of Economics)

Dr. Catherine Worthington, Outside Member
(School of Public Health and Social Policy)

Abstract

This dissertation is comprised of three stand-alone articles, two of them co-authored, and one solo. The solo article, "Increases in Victim Mortality Rates in the Aftermath of Mandatory Arrest Laws: A Study in Unintended Effects" is in Part I. The first co-authored article "Teachers' Strikes and Standardized Test Scores: Impact on Performance & Participation", is in II. The last chapter, Part III, is already published in the journal "Information Economics and Policy" (ISSN: 0167-6245), accessible under the title "The Value of Online Scarcity Signals".

Contents

List of Tables

List of Figures

Acknowledgements

To my partner-in-life, and my love, Melis Ciner, for her patience, first and foremost, but also her willingness to read, share ideas, and keep on cheering. This dissertation would not have been possible without her.

To my mom Ayse, for guiding me in this direction and for putting the dissertations of both her children first, even in the midst of health issues.

To my brother Selim, for sharing all family responsibilities and always caring for my mental well-being.

To my maternal uncle Erdal, for making all this possible in the first place by being there for our family.

To my advisors Elisabeth and Pascal - I was fortunate to have you both as mentors and enjoyed every moment of learning from you. You have all been available to give me invaluable advice every time I needed it, and have shown an extraordinary amount of patience.

Part I

Increases in Victim Mortality Rates in the Aftermath of Mandatory Arrest Laws: A Study in Unintended Effects

Abstract

From the late 1970s into the 1990s, 20 states in the USA passed 'mandatory arrest laws', designed to curb domestic violence, by requiring the responding officer to arrest the offender. I show that these laws led to an increase in the mortality rates of African-American women. The increase takes place two to four years after the law has been passed and loses its statistical significance as time passes. I infer from these results that mandatory arrest laws had unintended consequences for the victims whose partners were arrested, but that these unintended consequences impacted the victims only in the few years following the law. I conclude that increased awareness of the law and offender deterrence eventually mitigates these unintended consequences.

1 Introduction

In the current paper, I present evidence that mandatory arrest laws, intended to protect victims of domestic violence, led to a temporary increase in the mortality rates of African-American women. While this result may seem unexpected at first, prior literature suggests that such an outcome is possible: in an experimental setting, African-American victims were twice as likely to be dead after 23 years if their offending partners are arrested (Sherman and Harris, 2014). If arresting a domestic violence offender can unintentionally increase the probability of death for a victim, then mandating arrest as a policy for domestic violence can also unintentionally increase mortality rates of African-American women.

The main purpose and contribution of this paper is to estimate the impact of mandatory arrest laws on mortality rates of African-American women. For this purpose, I construct a data set of yearly mortality rates by state, gender and race. Results from Sherman and Harris (2014) suggest that victim mortality increases with partner arrest, but this result is limited to an experimental setting where the deterrence effects of the new law are non-existent or limited: offenders were not aware that their actions could lead to an arrest, and could not have been deterred from committing violence in the first place. Outside of the experiment, once the mandatory arrest laws are passed and known by the population, potential offenders can be deterred from offending in the first place.

For the identification of effects, I exploit the variation in the timing of the passing of the mandatory arrest laws using a fixed effects model (Angrist and Pischke, 2008). Similar fixed effects models have been used by Stevenson and Wolfers (2006), Aizer (2010) and Chesson et al. (2000), and are sometimes described as a difference-in-differences approach (Lovenheim and Willen, 2019; Anderson and Walker, 2015; Huebener and Marcus, 2017). In this approach, the main identification assumption requires the timing of the laws to be independent of confounding factors. These confounding factors may come from data that is not observed or collected. To control for the unobserved confounding factors that are specific to a state, state-fixed controls are included in the model, and to control for the confounding factors that are common across all states but vary over time, year-fixed controls are included. What remains are factors that may be correlated to a particular timing of the law passing in a given state. In the current setting, it was nation-wide events, and not state-specific triggers, that led to the passing of the laws (Iyengar, 2009; Buzawa and Buzawa, 1996). This increases my confidence that the identification assumptions are met. In spite of this, I also include state-year economic, demographic and political controls, as well as state-specific linear time trends. The results remain statistically significant.

This paper has two main results: mandatory arrest laws led to an unintentional increase in the mortality rates of African-American women, and this increase was temporary. I interpret these results as the consequence of deterrence of partners by mandatory arrest. The mandatory arrest laws created unintended consequences for the victims in the first couple of years after they were passed, but in the following years, they helped the victims by deterring the offenders, an effect that was not captured in the experimental setting by Sherman and Harris (2014). I discuss possible mechanisms and alternative explanations in Section 6.

The rest of the paper is organized as follows. In Section 2, I give the background information regarding mandatory arrest laws, the social experiments regarding spousal arrest, and a brief review of the literature concerning mandatory arrest laws. In Section 3, I explain the construction of the data set used for the analysis. In Section 4, I explain my methodology. In Section 5, I demonstrate that there is an impact of the mandatory arrest laws on the mortality rate of African-American women, and that this impact was not sustained in the long run. Finally, in Section 7, I finish by presenting my conclusions and the corresponding policy implications.

2 Background & Prior Literature

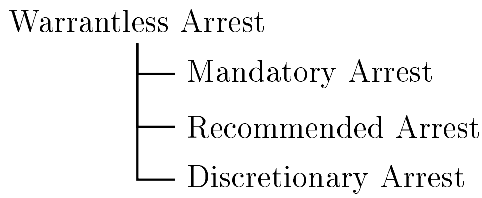
Mandatory arrest laws require officers to make an arrest when responding to a domestic violence incident. Prior to the end of the 70s, no states had any arrest laws concerning domestic violence: it was regarded as a "private affair" in the USA. An arrest would have required the police officer to witness the crime, or obtain an arrest warrant, which rarely happened (Fedders, 1997; Zeoli et al., 2011; Maxwell et al., 2001; Pavlidakis, 2009; Dugan, 2003). The increasing influence of women's right movements, including a nationally-known lawsuit against a police department[1], started to bring focus to domestic violence as a policy concern (Buzawa and Buzawa, 1996). In 1981, researchers working with the Minneapolis police department conducted a random arrest experiment, Minneapolis Domestic Violence Experiment (MDVE), and concluded that arrest decreases the repetition of domestic violence in the following six months (Sherman and Berk, 1984). The Attorney General released a report recommending arrest as an appropriate response to family violence (Hart, W. L., 1984). These events, particularly the MDVE, triggered a nation-wide response to deal with domestic violence as a public concern (Iyengar, 2009; Fedders, 1997). By the mid-90s, twenty states had mandatory arrest laws. The remaining states either had laws that recommended arrest as a response to domestic violence incidents, or left the decision to arrest at the discretion of the responding police officer.[2]

[1]The case is Thurman v. City of Torrington (1984). The police officer arriving on the scene failed to make an arrest despite seeing the suspect with a bloody knife and kicking his wife in the head. The wife survived and was eventually awarded $ 2.3 million after suing the city for failing to protect her. For more details, see footnote 36 in Fedders (1997), or the first paragraph of Section II in Sherman et al. (1992a).

[2]Even though warrantless arrest laws vary in their wording and structure, criminal law literature categorizes them into three categories (Frye et al., 2007; Iyengar, 2009; Hirschel

Around the same time that mandatory arrest laws were being passed, the results from MDVE were challenged by a series of experiments. These experiments are collectively called the Spouse Arrest Replication Program, or SARP. For the most part, the SARP replication experiments do not confirm the original findings from MDVE (Sherman et al., 1992a; Maxwell et al., 2001; Berk et al., 1992). Despite this rejection, Berk et al. (1992) show that there is a decrease in repeated offense for offenders with a strong "stake in society", for instance, employed offenders. Finally, Maxwell et al. (2002) conduct a meta study combining all experiments from the SARP program. They conclude that arrest does in fact reduce repeated offense. They also confirm the "stake in society" argument by showing that the decrease is larger for employed offenders.

Figure 1: The Types of Domestic Violence Arrest Laws



Criminal law literature categorizes domestic violence arrest laws into three categories: mandatory, recommended and discretionary (Frye et al., 2007; Iyengar, 2009; Hirschel et al., 2007a).

The largest of these replication experiments was the Milwaukee Domestic Violence Experiment (MilDVE). Recently, Sherman and Harris (2014) conducted a follow-up study on the victims 23 years after the experiment. Surprisingly, they found that spouses of offenders who were arrested were twice as likely to be dead. It is not clear what causes the deaths, and it is also not clear how long after the arrest the deaths take place.[3] Their results were driven by the mortality rates of African-American women.[4] While

et al., 2007a). *Mandatory arrest laws* require the officer to make an arrest. *Recommended (or preferred / pro-arrest) arrest laws* articulate that arrest is the recommended action, but do not make it mandatory. *Discretionary arrest laws* leave the decision to the police officer. See 1.

[3]Even though Sherman and Harris (2014) do not report on the particulars of the cause of death, they are able to discern between violent deaths, heart disease, and cancer. Their results come entirely from non-violent deaths. In the current study, I double check my results by similarly distinguishing non-violent deaths. The results are very similar to the main results. (Not reported)

[4]Specifically, among the African-American population, they found that among 529 victims treated with a partner arrest, 52, or almost 10% were dead, as opposed to only 13 fatalities, or 5%, among the 262 victims whose partners were not treated with partner arrest.

there have always been concerns about racialized applications of the law[5], Sherman and Harris (2014) provide the first empirical evidence of a possible causal impact of arrest on victim outcomes. Mandatory arrest laws have been criticized for various reasons after their inception[6], but an increased likelihood of victim's death is arguably its most serious unintended consequence in the literature.

At first, it seems intuitive that mandatory arrest laws should not *increase* victim mortality rates: mandatory arrest laws are aimed at protecting the victims, either by removing the offender from the scene, or by deterring the offender from hurting the victim in the first place. Sherman and Harris (2014) discuss possible causal pathways that could explain this unexpected increase in mortality, but they do not bring evidence on any particular pathways. Given that their results are driven by African-American women, it is possible that potential racialized applications of the law by police departments may have played a role. Sherman and Harris (2014) speculate that an "interaction between arrest, race and employment" may explain the causal pathway that leads to the victim's death. They note that the increase in mortality was attributable to African-American female victims, and that employed victims and victims with partners who had no prior arrest were impacted the most. Possibly, the probability of death of a victim might be increasing due to the loss of a job right after the incident. Alternatively, the loss of a job of a partner may have a similar impact.[7] Domestic violence is a particular type of crime where the interaction of the victim and the offender can continue after an arrest or a prosecution. As such, victims can be impacted indirectly by the loss of income opportunities of the offender. For example, alimony payments to a victim may decrease if the offender faces a decrease in their income. Such a pathway would be in parallel with microeconomic theory: because members of a household may benefit from pooled resources, loss of resources of the offender may impact the victim.

The results from Sherman and Harris (2014) motivate the need to focus on African-American women. This is perhaps expected: women are the primary victims of more physical and criminal forms of domestic violence, and African-American women are particularly at risk (Farmer and Tiefenthaler, 2003; Aizer, 2010). However, African-American men, white women, and

[5] See Fedders (1997); Ruttenberg (1994); Desmond and Valdez (2012); Miller (1989)

[6] See Humphries (2002) for a good list and discussion.

[7] Pager (2003); Bushway (1998) show that a criminal record decreases employment opportunities.

white men may also be impacted by mandatory arrest laws. How these groups are impacted can present evidence about the causal pathway. In particular, if the causal pathway to mortality is mainly through the interaction of arrest and loss of employment opportunities, I would expect to see an increase in the mortality rates of all four groups: African-Americans and whites, and both genders. These groups can be impacted through the loss of resources related to employment whether they are offenders or victims. Conversely, if racialized applications of the law play an important role, I would expect to see an impact on African-American women and men, but not white men and women. How much the consequences of arrest can be transferred to the victim can also show up on the data. If the mortality rates of women increase, this serves as evidence that the law is impacting women, even though they are less likely to be offenders of domestic violence. If the mortality rates of women and men increase by similar amounts, this would serve as evidence that both offenders are victims are likely to be impacted by the law.

The randomized experiment in MilDVE was designed in the following manner: first, a domestic violence incident needs to take place, and the victim needs to call the department. Once an offender and a victim unknowingly choose to be part of the experiment, the police department made sure that they met the eligibility criteria: the offense had to be a misdemeanor, so more serious offenses that involved injury to the parties would not be included in the randomization. Furthermore, the responding police officer had to have probable cause that the offense occurred recently. After these two criteria were met, the treatment (arrest) would be assigned randomly.

There are certain difficulties with this design: first, the offender is unaware of the ongoing experiment, so they do not have a chance of being deterred from the act of violence by the threat of arrest. For this reason, the randomized experiments were able to test reductions only in repetition of the offense. Reductions in potential domestic violence in the overall population are not easy to test through randomized design. Furthermore, the design applied to only offenses that are not serious enough to always warrant an arrest. Arguably, arrests of spouses could be more beneficial and less harmful to the victim when the underlying offense is more serious. Finally, the researchers themselves suspected that police officers may have subverted the experiment where they felt that the given experimental treatment was inappropriate. All of these measures point to a situation where the results could be biased due to attrition from the experiment, or due to subversion (Sherman and Berk, 1984).

Typically, mortality rates of African-American women follow similar trends in mandatory and discretionary arrest law states. In the current paper, I present evidence that mortality rates for African-American women deviate from their trends in two to four years after the mandatory arrest laws were passed. With the quasi-experimental design in the current study, I estimate the impact of mandating a treatment on an entire population, rather than a particular subsample. This allows me to overcome the challenges faced by the researchers in experiment design. Most importantly, the impact of deterrence of the first domestic violence incident is not possible to capture in this randomized experiment design. In addition, the situations with more serious offenses are not even in the experiment. Finally, the experiment design suffers from the possibility of subversion, an issue that the authors themselves noticed. Mortality rates of entire state populations do not suffer from these issues, and the variation in the timing of the arrest laws create a quasi-experimental setting, which I exploit to overcome the difficulties from the original experiment.

3 Data

In the current paper, the outcome variable is the mortality rate of African-American populations for each state-year pair. Mortality rates are defined as the number of deaths in a given year per 100,000 people of the relevant population. They are used extensively in the medical and health economics literature as an outcome of interest. Since Sherman and Harris (2014) observe an increase in mortality of African-American victims, I also focus my analysis on African-American mortality.[8] Similar to Sherman and Harris (2014), I use changes in crude mortality among all of African-Americans in response to arrest law changes. I calculate these rates from the mortality data available at the National Bureau of Economic Research (NBER).[9] The descriptive statistics for African-American mortality rates are found in Table 1, broken down by gender, and the type of arrest law.

[8]I also ran the same analysis for white populations as well and found no significant change.

[9]For the number of deaths, I use data available at http://www.nber.org/data/vital-statistics-mortality-data-multiple-cause-of-death.html. For population estimates for each state year pair, I use data available at http://www.nber.org/data/census-intercensal-county-population.html

Table 1: Mortality Rates African-American Women

	Mortality Rates	Obs	Mean	Std Dev	25th Percentile	Median	75th Percentile
Women	All	1836	638	240	497	667	791
	No Law	1031	663	239	522	685	809
	Discretionary	416	626	241	501	675	779
	Recommended	80	543	276	318	512	820
	Mandatory	309	593	216	466	554	767
	Before 1985	816	660	235	526	676	811
	1985 And After	1020	620	242	475	662	784

Table 1 shows the descriptive statistics for the mortality rates. The statistics come from state-year pairs, which correspond to each row in the data. For instance, "Mandatory" refers to the data points for states that passed a mandatory arrest law, but includes only the years after the law has been passed. Overall, mortality rates for African-American have been decreasing all across the United States. (The year 1985 is chosen arbitrarily from the middle of the data set to show this decrease.)

In any given state and year, a state can either have a warrantless arrest law or no law governing arrest. When the state has an arrest law, it can be either mandatory, recommended, or discretionary. For each of these classifications, I use a separate binary variable. As an example, the binary variable for mandatory arrest laws takes the value one in all years beginning with the passing of a mandatory arrest law, and zero in all preceding years. No states repeal warrantless arrest laws. Since a variable constructed in this manner is going to be serially correlated, I estimate standard errors clustered at the state level.

Warrantless arrest laws vary significantly in wording across states. It is therefore necessary to find a reasonable simplification that categorizes all different laws as one of the three classes. Zeoli et al. (2011); Hirschel et al. (2007a); Ruttenberg (1994) and Iyengar (2009) are the four papers that include such a clear classification. Among these, only Iyengar (2009) and Zeoli et al. (2011) present a categorization that includes the year of passing. From the two, I prefer Zeoli et al. (2011) because it was written by legal experts particularly in response to this need for classification. It is also the most recent classification available.[10] In addition, Zeoli et al. (2011) classification, unlike Iyengar (2009) classification, includes the year of passing for discretionary laws as well as the mandatory and recommended. By using the Zeoli et al. (2011), I am able to include binary variables for a discretionary arrest law and confirm that these coefficients are not significant. This classification is presented in Table 2.

According to Zeoli et al. (2011), there are 20 states and the District of Columbia with the expression "shall arrest" in their wording. These are classified as mandatory arrest law states. Similarly, there are six states with the expression "is the preferred/recommended action" in their arrest law wording. These are classified as recommended arrest law states. The remaining 25 states have laws with only the expression "may arrest" in their wording and these are classified as discretionary arrest law states. This procedure covers all warrantless arrest laws currently in effect in 50 states in the USA as well as the District of Columbia. This categorization is guided by the summary of expressions in Table 2 in Zeoli et al. (2011). All of these warrantless arrest laws were passed over a course of two decades - see Table 2.

4 Empirical Methodology

To identify the effect of mandatory and recommended arrest laws on the mortality rates of African-American women, I exploit the fact that the laws passed in different years in different states. For identification, I employ a difference-in-differences strategy and implement this strategy in a fixed effects estimation. This allows me to exploit the variation in passing of the laws across states and over time by including state-fixed effects and year-fixed effects. The baseline estimation equation is:

$$y_{s,t} = \beta_m M_{s,t} + \beta_r R_{s,t} + \beta_d D_{s,t} + \psi X_{s,t} + \alpha_s + \lambda_t + \epsilon_{s,t} \tag{1}$$

Here, the subscripts s and t denote state and year. Correspondingly, α_s and λ_t refer to the set of binary variables that control unobservable factors particular to each state, and unobservable factors that change over time across all states. Coefficients β_m and β_r are the effect of mandatory and recommended arrest laws on mortality rates. $M_{s,t}$ and $R_{s,t}$ are binary variables

[10]In fact, Iyengar (2009) is one of the papers that Zeoli et al. (2011) cite as justification for their paper.

Table 2: Year of the Passing of the Arrest Law

State	Categorization	Year of Passing			
AK	Mandatory	1996	MT	Recommended	1991
AL	Discretionary	1989	NC	Discretionary	1991
AZ	Discretionary	1991	ND	Recommended	1989
AR	Recommended	1991	NE	Discretionary	1989
CA	Recommended	1996	NH	Discretionary	1979
CO	Mandatory	1994	NJ	Mandatory	1991
CT	Mandatory	1986	NV	Mandatory	1985
DC	Mandatory	1986	NM	Mandatory	1987
DE	Discretionary	1984	NY	Mandatory	1996
FL	Discretionary	1992	OH	Mandatory	1995
GA	Discretionary	1981	OK	Discretionary	1987
HI	Discretionary	1980	OR	Mandatory	1978
IA	Discretionary	1986	PA	Discretionary	1986
ID	Discretionary	1979	RI	Mandatory	1988
IL	Mandatory	1993	SC	Mandatory	1995
IN	Discretionary	2000	SD	Mandatory	1996
KS	Mandatory	1992	TN	Recommended	1995
KY	Discretionary	1980	TX	Discretionary	1989
LA	Mandatory	1985	UT	Mandatory	1995
MA	Recommended	1988	VA	Mandatory	1997
MD	Discretionary	1986	VT	Discretionary	1985
ME	Mandatory	1980	WA	Mandatory	1995
MI	Discretionary	1978	WI	Mandatory	1989
MN	Discretionary	1978	WV	Discretionary	1994
MO	Mandatory	1989	WY	Discretionary	1982
MS	Mandatory	1995			

Year the arrest law was passed in treatment (mandatory & recommended), and in control (discretionary) states. The list includes all 50 states and the District of Columbia.

and denote whether a mandatory or a recommended arrest law is in effect in a given state and year. I also include a binary variable $D_{s,t}$ indicating whether a discretionary law was passed in a state, and the corresponding coefficient β_d helps determine if these states have significant changes after the the discretionary laws are passed. $X_{s,t}$ denotes additional political and economic controls particular to each state and year, and ψ is the corresponding coefficient.

Using the model in Equation 1, I am able to control for state-specific and year-specific unobservables using state- and year-fixed effects. It is important to control for year-specific unobservables for two reasons. First, mortality rates for African-American women show a downward trend. Year-fixed binary variables control for this trend. Second, a number of nationwide policies may have affected mortality rates and domestic violence rates during this period. Notably, the African-American mortality rates may be affected by tough-on-crime measures such as the Anti-Drug Abuse Act of 1986, and domestic violence rates may have been affected by the Violence Against Women Act of 1994. Adding year-fixed effects allows us to control for omitted variables bias related to unobserved factors changing over time across all states. Similarly, state-fixed effects allow us to control for state-specific unobservable variables. For instance, certain states may have particular practices specific to them that may impact African-American mortality rates. The fixed-effect model is considered a generalization of the difference-in-differences model (Lovenheim and Willen, 2019; Anderson and Walker, 2015; Huebener and Marcus, 2017; Allison, 1994). By controlling for state fixed effects, we effectively control for treatment and control units, at the same time allowing treatment and control units to vary over state. By controlling for time fixed effects, we control for pre- and post- treatment, while allowing for time trends.

Controlling for factors specific to each state and each year leaves the factors that may be correlated with the timing of laws in each state. I find no literature claiming that mandatory arrest laws were passed as a reaction to state-specific triggers or trends. Contrarily, the literature is unanimous in that mandatory arrest laws were triggered by nationwide events, such as the MDVE, a number of lawsuits and the pressure exerted by women's rights groups Buzawa and Buzawa (1996); Iyengar (2009).

A particular concern is that other policy changes could have been passed together with the arrest laws or that state-specific political trends could be correlated with the timing of the mandatory arrest laws. There is no evidence of this in the literature. Zeoli et al. (2011); Hirschel et al. (2007a);

Dugan et al. (2003) and Zorza (1992), among others, describe the conditions under which mandatory arrest laws were passed, and make no mention of any other laws that were passed at the same time. However, it is possible that mandatory arrest laws were possibly part of some other political trend that was unique to these states. While we control for factors specific to each state that are not changing in time, if there are political factors that changed over time and are correlated with the timing of mandatory arrest laws, these could bias the estimates. To be able to control for such factors, I include binary variables indicating party control in each one of the houses in each state for all given years. The binary variable denoting the composition of the state senates does not yield a significant estimate, but the one denoting the composition of the houses of representatives is significant. Addition of state-year specific political controls do not alter the significance of my variables of interest.

Economic factors may also be important for mortality rates. If there are factors particular to a specific state, state-specific binary variables in the model will control for these. For instance, if there are economic factors that correlate with the particular timing of the passing of mandatory arrest laws in each state, they can bias the coefficient estimates. I include economic controls for each state-year for the average personal income, GDP and employment rates from the Bureau of Economic Analysis. If there are such economic factors associated with the particulars of each state and year, I should observe a change in the coefficients. Instead, the introduction of these controls do not change the significance or the magnitude of the coefficients of interest.

When using fixed-effects models for causal inference, Angrist and Pischke (2008) advise including state-specific linear time trends. State-specific time trends are a final check on the identification strategy. Even after the literature review, controlling for unobservable factors relating to each state, and the inclusion of political and economic controls particular to a specific state and year, there may be remaining unobserved confounding factors. State-specific linear time trends can capture any remaining unobserved confounding factors. Addition of state-specific time trends does not change the significance of the coefficients of interest, but lowers the magnitude of the coefficient. I discuss the implications of adding state-specific trends in Section section §6.

In addition to the main results, I use the model in Equation (2) to test the timing of the effect. In Equation (2) is a time-varying extension of Equation equation (1). Here, I change the binary variables of interest, $M_{s,t}$ and $R_{s,t}$, as follows: I include six variables, $M_{s,t-k}$ where $k \in \{0, 5\}$, capture the effect in the year of passing and the five subsequent years. Different from Equation (1), the variable $M_{s,t-k}$ takes the value one only in the k-th year after a mandatory arrest law has been passed, and the value zero in all other years. I then include a final variable, $M_{s,LT}$, which is similar to the original $M_{s,t}$ but is lagged by five years, i.e. it takes the value one in all years subsequent to the fifth year after the law has passed. This way, $\beta_{m,LT}$ captures the long-term effect of mandatory arrest laws. The variables $R_{s,t-k}$ and $R_{s,LT}$ are defined in the same way, but are for recommended arrest laws rather than mandatory arrest laws.

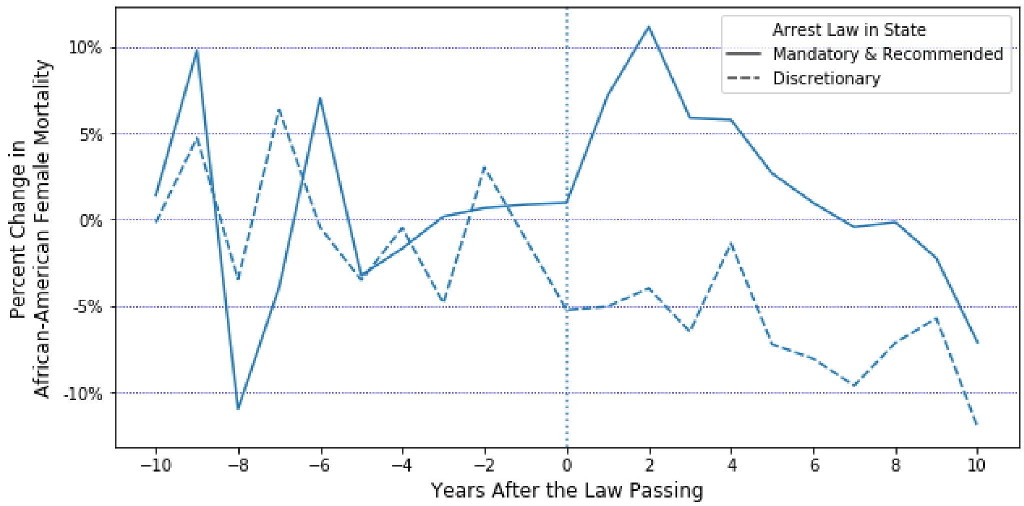
$$
\begin{aligned}
y_{s,t} = \sum_{k}^{k\in\{0,5\}} \beta_{m,k} M_{s,t-k} + \sum_{k}^{k\in\{0,5\}} \beta_{r,k} R_{s,t-k} + \sum_{k}^{k\in\{0,5\}} \beta_{d,k} D_{s,t-k} + \\
\beta_{m,LT} M_{s,LT} + \beta_{r,LT} R_{s,LT} + \beta_{d,LT} D_{s,LT} + \psi X_{s,t} + \alpha_s + \lambda_t + \epsilon_{s,t} \quad (2)
\end{aligned}
$$

Equation (2) allows me to capture the timing of the effect of the mandatory laws. Estimating if the impact is temporary or permanent is important to understand how society reacts as a whole to mandatory arrest laws. If the death is due to an immediate event, such as a violent retaliation or suicide, and the impact is permanent, the coefficients $\beta_{m,0}$ through $\beta_{m,5}$ and the long-term coefficient $\beta_{m,LT}$ should all have similar magnitudes. However, if the impact is immediate but temporary, the coefficient $\beta_{m,0}$ should be high in magnitude, but the following coefficients $\beta_{m,1}$ through $\beta_{m,5}$ and the long-term coefficient $\beta_{m,LT}$ should be small and statistically insignificant. In another extreme, the deaths can be due to a long-term and sustained decrease in life expectancy of the victims. If this is the case, all of the coefficients $\beta_{m,0}$ through $\beta_{m,5}$ should have small coefficients, and $\beta_{m,LT}$ should be large in magnitude. The results show that the impact is concentrated in the third and fourth years after the arrest law has been passed, and that it is not sustained in the long term. The implications of these results are in Sections 6 and 7.

5 Results

Figure 2 is a visual representation of my main results and shows a deviation from the ongoing trend after a mandatory arrest law has passed. In the long run, this deviation decreases. To obtain Figure 2, I normalize the mortality

Figure 2: Average Mortality Rates of African-American Women Following the Law Change



rate changes of each of the states by their values before the law changes. This serves as a visual representation of controlling for state level unobservable confounding factors. I also shift the timelines to meet at the year that a law was passed; this serves as a visual equivalent of controlling for time trends. Finally, I take the average of each state-year pair within the treatment (mandatory and recommended) and the control (discretionary) states. The line in the middle divides the data into before and after, and the dashed and the solid lines show the treatment and control. The mortality rate of African-American women in discretionary states follow a downward trajectory, while the mortality rate of African-American women in mandatory and recommended states deviate from their similar trajectory after the law has passed. After a few years, the difference begins to diminish and becomes indistinguishable from statistical noise.

Table 3 shows the results from Equation (1). Specifically, columns (2) corresponds to Equation (1). Column (1) is the same as column (2) except that column (1) does not include any controls other than the binary controls for state-specific and year-specific effects. In column (3), I add a linear time trend. In column (4), I allow for the states to have different slopes for this time trend by interacting the linear trend with state-level binary variables. The results from columns (1) through (4) can be interpreted as the variation in mortality rates, i.e. the annual number of deaths per 100,000 people.

Table 3: Main Results: Impact of Mandatory and Recommended Arrest Laws on Mortality Rates of African-American Women

	Mortality Rates				Percent Change (Poisson)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discretionary (β_d)	-10.12	-3.563	-3.563	-3.102	-0.925%	0.238%	0.238%	-0.824%
	(19.79)	(16.84)	(16.84)	(12.88)	(3.26%)	(2.85%)	(2.85%)	(1.87%)
Recommended (β_r)	24.70	36.35**	36.35**	27.38	3.56%	5.84%**	5.84%**	5.79%
	(19.11)	(15.95)	(15.95)	(17.23)	(3.34%)	(2.75%)	(2.75%)	(3.92%)
Mandatory (β_m)	41.59**	35.07**	35.07**	25.69**	6.58%**	5.39%*	5.39%*	4.26%**
	(17.94)	(15.22)	(15.22)	(12.57)	(3.08%)	(2.83%)	(2.83%)	(2.09%)
Demographic Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Economic Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Political Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Time Trends	No	No	Yes	Yes	No	No	Yes	Yes
State Time Trends	No	No	No	Yes	No	No	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1836	1800	1800	1800	1836	1800	1800	1800

Standard errors in parentheses

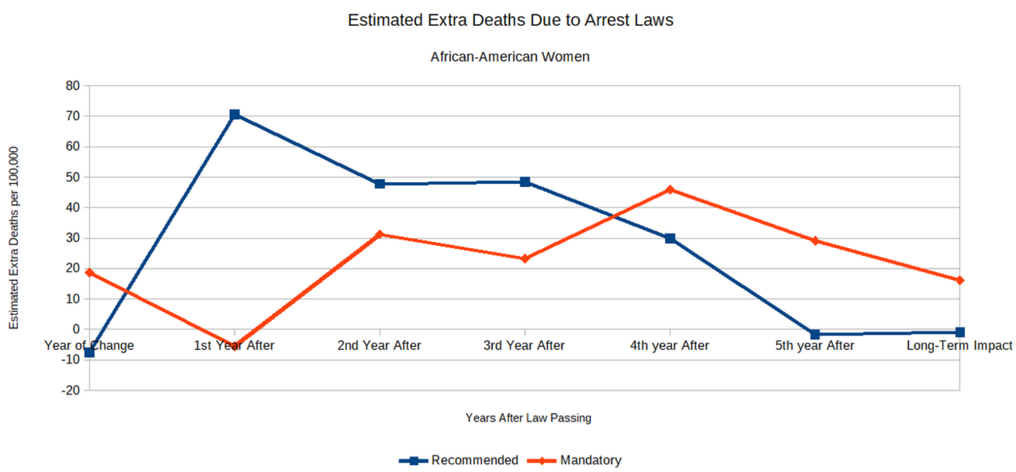
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the impact estimates of mandatory and recommended arrest laws. Robust standard errors, clustered by state, are reported in parentheses. Results are robust to block bootstrapped error estimation. The errors are not guaranteed to be normally distributed, in fact, for the log-linear specification, they are not normally distributed even asymptotically (Giles, 2011). Block bootstrapped errors are very similar, and the star category remains the same.

When the outcome variable is a mortality rates or an incident rate, some researchers prefer a linear model similar to Equation (1)(Stevenson and Wolfers, 2006; Winegarden and Bracy, 1995; Macinko et al., 2006). However, an alternative is to use a log-linear regression - examples in literature include Baltagi and Levin (1992) and Lochner et al. (2003). In the current paper, I also report on the results from a log-linear regression. In Table 3, columns (5) through (8) show the results from a log-linear model, and they are organized in a similar manner to columns (1) through (4). These results can be interpreted as the percent change in mortality rates. The coefficients estimates are also available in Figure 3.

There are a few additional analyses. One related line of investigation is the cause of death. For this purpose, I run the same regression for mortality by three causes of death: violence, ischemic disease and cancer. I do not observe a statistically significant increase in the mortality rates. This suggests that even though there was an increase in the mortality rates, it was not driven by one particular cause. As such, multiple mechanisms may explain the increased mortality. Another related line of investigation is about mortality of white men and women. White men and women may also be impacted by this policy change. I run these regressions, and do not obtain statistically significant results. These results are available in the Appendix. Finally, to understand better which age groups are impacted, I run the regression on various age groups. I find that the women above the age 60 are the ones experiencing a significant increase in mortality rate. The other groups seem unaffected.

Figure 3: The Timing of the Impact of Mandatory and Recommended Arrest Laws on Mortality Rates of African-American Women



This figure shows the dynamic response of the mortality rates of African-American women to the passing of the arrest laws. The corresponding coefficient estimates and standard error estimates are available in Table 4.

Table 4: The Timing of the Impact of Mandatory and Recommended Arrest Laws on Mortality Rates of African-American Women

	Mortality Rates				Percent Change (Poisson)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\beta_{r,0}$	0.652	5.209	5.209	-7.453	0.256%	0.144%	0.144%	-0.086%
	(29.21)	(24.16)	(24.16)	(16.04)	(5.26%)	(4.74%)	(4.74%)	(2.67%)
$\beta_{r,1}$	79.42**	84.84**	84.84**	70.55	14.6%**	14.7%*	0.147*	0.145
	(31.97)	(37.84)	(37.84)	(48.13)	(6.45%)	(7.12%)	(7.12%)	(9.34%)
$\beta_{r,2}$	55.45**	65.77**	65.77**	47.84	9.68%**	11.3%**	11.3%**	9.83%
	(23.10)	(27.92)	(27.92)	(40.25)	(4.25%)	(4.68%)	(4.68%)	(7.48%)
$\beta_{r,3}$	55.03	66.19	66.19	48.41	9.068%	10.96%	10.96%	9.746%
	(39.19)	(42.34)	(42.34)	(52.93)	(6.68%)	(7.16%)	(7.16%)	(9.52%)
$\beta_{r,4}$	42.37	46.88	46.88	29.81	7.423%	7.993%	7.993%	7.165%
	(31.16)	(28.48)	(28.48)	(41.14)	(4.98%)	(4.78%)	(4.78%)	(7.33%)
$\beta_{r,5}$	13.53	17.71	17.71	-1.670	2.768%	3.293%	3.293%	2.347%
	(22.72)	(16.32)	(16.32)	(24.99)	(3.91%)	(2.86%)	(2.86%)	(4.78%)
$\beta_{r,LT}$	7.017	16.90	16.90	-0.809	0.057%	2.03%	2.03%	2.286%
	(24.35)	(23.80)	(23.80)	(33.17)	(4.64%)	(4.48%)	(4.48%)	(5.65%)
$\beta_{m,0}$	25.86	18.10	18.10	18.67	4.3%	2.685%	2.685%	0.03231
	(16.45)	(13.75)	(13.75)	(13.69)	(2.67%)	(2.32%)	(2.32%)	(2.21%)
$\beta_{m,1}$	2.916	-5.578	-5.578	-5.561	0.138%	-1.774%	-1.774%	-1.242%
	(18.95)	(16.66)	(16.66)	(17.85)	(0.0329)	(0.0305)	(0.0305)	(0.0286)
$\beta_{m,2}$	35.26*	28.76*	28.76*	31.29*	5.80%*	4.571%	4.571%	5.633%*
	(18.28)	(16.95)	(16.95)	(18.36)	(2.95%)	(2.94%)	(2.94%)	(3.11%)
$\beta_{m,3}$	31.26*	22.57*	22.57*	23.34*	4.97%*	3.303%	3.303%	4.091%*
	(16.99)	(13.05)	(13.05)	(13.42)	(2.71%)	(2.25%)	(2.25%)	(2.24%)
$\beta_{m,4}$	56.41*	47.84*	47.84*	46.00**	9.38%**	7.724%	7.724%	7.983%**
	(28.27)	(27.40)	(27.40)	(20.40)	(4.49%)	(4.63%)	(4.63%)	(3.62%)
$\beta_{m,5}$	40.67	31.63	31.63	29.07	6.609%	4.812%	4.812%	4.959%
	(25.81)	(23.25)	(23.25)	(17.75)	(4.27%)	(4.05%)	(4.05%)	(3.15%)
$\beta_{m,LT}$	37.72	31.61	31.61	16.17	6.12%	4.991%	4.991%	2.922%
	(27.44)	(23.93)	(23.93)	(23.01)	(4.53%)	(4.29%)	(4.29%)	(3.47%)
Demographic Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Economic Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Political Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Time Trend	No	No	Yes	Yes	No	No	No	Yes
State Time Trends	No	No	No	Yes	No	No	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1530	1500	1500	1500	1530	1500	1500	1500

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the estimates of the impact of mandatory and recommended arrest laws over time. Robust standard errors, clustered by state, are reported in parentheses. The errors are not guaranteed to be normally distributed, in fact, for the log-linear specification, they are not normally distributed even asymptotically (Giles, 2011).

6 Discussion

The results from Table 3 show that the mortality rates of African-American women increased in response to the passing of either mandatory or recommended arrest laws. I estimate a statistically significant increase of 25 to 42 deaths per 100,000 in the mortality rate of African-American women. This can be interpreted as additional deaths attributable to the unintended consequences of mandatory arrest laws per 100,000 people. The estimate for the effect of mandatory arrest laws remains statistically significant even after including state-specific time trends. This confirms that partner arrest has unintended consequences for African-American women; and that these consequences show up in their mortality rates.

My results decrease in magnitude and statistical significance when state-specific time trends are added. State-specific time trends should be interpreted with caution in this setting: if the mechanism leading to mortality is taking place in the short term, say, a few years after the arrest, then adding state-specific time trends acts as a reasonable check of validity on my results. However, if the impact is happening in the long term and is sustained, the state-specific time trends may absorb some of the impact attributable to the arrest. This is because the state-specific time trends are constructed from the entirety of the timeline for each state, and as such, they incorporate the trends in mortality rates after a mandatory arrest law has been passed. In this case, the magnitude of the estimates will be biased downwards. Despite this caveat, I add state-specific linear time trends. This inclusion follows literature, and the advice from Angrist and Pischke (2008) when using fixed effects models for causal inference. The results are still statistically significant.

The effects of mandatory arrest laws are in agreement with the results from Sherman and Harris (2014). In addition, results from the model shown in Equation (2) show that the estimated impact comes from the second to fourth year after the law is passed. The coefficients are statistically significant, and change only slightly in magnitude and significance with the addition of time trends. The coefficient showing the impact in the fifth year and in the years following it are no longer statistically significant. Recommended arrest laws show a similar trend, they are jointly significant, but the significance comes from the first through fourth year coefficients. These results support the findings from Sherman and Harris (2014), and show that they have played out in a similar way in the empirical setting as well.

Taken together, the models present a picture of a short-lived increase in the mortality rates of African-American women in the aftermath of mandatory arrest laws. In the second year after a mandatory arrest law has been passed, there is a clear increase in the mortality rates of African-American women. This increase is sustained in the third year, and increases in magnitude in the fourth. However, it decreases in magnitude and is much lower and statistically insignificant in the fifth year and in the years thereafter.

The victims that were impacted are only the first wave of victims, and the mechanism took place in a few years. The causal pathway between partner arrest and mortality takes place within a few years of partner arrest. More importantly, this causal pathway no longer has a statistically significant impact in the few years after a mandatory arrest law has passed. In other words, the unintended consequences of partner arrest affected victims only in the few years after the laws have passed. Like the victims from the MilDVE experiment, these "first wave" of victims were likely not aware of the consequences of reporting the offense. In this sense, they were similar to the victims in the experimental setting, uninformed, therefore did not have a chance to change their behavior in response to the arrest laws. Consequently, I observe the treatment effect of partner arrest in the first few years in the dynamic model. As time passes, potential offenders and victims are likelier to become informed about the legal change and the consequences of offending and reporting. In the long run, the net policy effect of mandating partner arrest becomes statistically insignificant.

It might be considered surprising that there are differences between the results from the experimental setting in Sherman and Harris (2014) and the results from the empirical setting. In reality, this sort of divergence happens. For example, alcohol consumption has been shown to lead to liver disease in clinical trials (Herd, 1992). Based on this result, a policy maker can decide to outlaw alcohol. However, liver disease did not decrease with state prohibitions (Dills, 2004). Mortality from alcohol-related causes decreased only with the constitutional prohibition, and then actually rose back to pre-prohibition levels, before the constitutional prohibition was lifted (Miron and Zwiebel, 1991). This is a clear example of how a treatment effect can be different in an experimental setting and in an empirical setting. Mandating or outlawing a certain treatment can change behavior of populations in the long run, leading to different outcomes outside of a randomized trial. In the case of prohibition, outlawing alcohol led to an unregulated market of alcohol. In the case of mandatory arrest law, mandating arrest created a deterrent that acted over a longer period of time, in a way that a randomized trial that ran once on a subset of individuals cannot create. This explains the divergence of the results from the experimental setting.

7 Conclusion

The results from the static and the dynamic models paint the following picture: there are unintended consequences of a partner arrest, but these consequences affect only the first wave of victims at a statistically significant level. After a few years have passed under mandatory arrest laws, partner arrest seems to no longer increase mortality rates for victims. I interpret this as a result of potential offenders and victims adapting their behavior as they become informed about the legal regime they live under. In this sense, the victims and offenders in the first wave were similar to the victims and offenders from the experiment: they were not aware that arrest was a potential consequence, and were not able to mitigate whatever unintended consequences led to their demise. This constitutes the main conclusion of this paper.

This conclusion confirms the experimental findings of Sherman and Harris (2014) in an empirical setting in the short term, but also shows that the impact disappears in the long term. In the empirical setting, the first set of offenders were unaware that they could be arrested, and the victims suffered the eventual consequences of arrest. However, the current study shows that the overall impact is different when the offenders are informed. Subsequent potential offenders were likely deterred from offending because they became more and more aware of the legal regime.

An alternative interpretation is possible: victims, rather than the offenders, may have been deterred from reporting the crime. Research suggests that this is less likely: Chesney-Lind (2002) and Langton et al. (2012) show that African-American women are more likely than white women to report domestic violence to the police.[11] Furthermore, arrest levels across mandatory and discretionary states are not significantly different from each other (Hirschel et al., 2007a). However, it is well-known that domestic violence is an underreported crime (Dugan, 2003; Farmer and Tiefenthaler, 2003). The

[11] Miller (1989) notes that this may be due to correlation to economic means, and that women with more income have more access to resources and may not need to rely on the police as much.

impact of mandatory arrest laws on reporting remains an important open question in literature.

The causal pathway that links the offender's arrest to the victim's death is not clear, but it is possible to discuss some possibilities. Lack of a clear cause of death, and the evidence of an increase in the mortality of African-American women (but not other groups) suggest that a complex interaction between racialized applications of policies, arrest and loss of resources may be increasing the likelihood of the victim's death. One such example can involve eviction: Desmond and Valdez (2012) report examples of how victims were evicted because the house they were in has been subject of an arrest. Another type of interaction can involve the loss of spouse's work opportunities: Pager (2003) shows that a criminal record decreases employment opportunities.[12] These losses of opportunity or means can be detrimental to many. For individuals already at risk, it can have disastrous consequences. Domestic violence is already correlated with lower income and a lack of economic means, so individuals at risk of losing healthcare, falling below the poverty line and eviction are over-represented within the victims of domestic violence. Eviction can lead to homelessness for individuals who are already struggling to make ends meet, and loss of spousal income can push an individual below the poverty threshold. Both of these events can have health consequences, and these health consequences can increase the probability of death.

Based on the conclusions, I make the following policy recommendation: if policy makers decide to pass mandatory arrest laws, they should announce and inform the population in advance. This will give ample time for potential offenders to inform themselves and shift their behavior. This is not an endorsement of mandatory arrest laws in general: there are other policy measures that are geared towards supporting the victims rather than punishing the offenders. The results from Sherman and Harris (2014), along with the current results, suggest that policy makers should pause before passing mandatory arrest laws. However, if they are going to be passed, informing the population beforehand is desirable to mitigate potential unintentional consequences of these laws.

[12] There are other studies that find a link between arrest and job market opportunities: Hunter and Borland (1999) show that Indigenous populations suffer an 18% decrease in employment opportunities by being arrested. Bushway (1998) finds a modest impact on the job opportunities of young white American men after being arrested. Bushway (1998) is another example.

Appendix I: Results for Other Groups of Interest

In addition to African-American women, African-American men and white men and women may also be impacted the law. The following table summarizes the results from a similar regression, where only the combined impact of mandatory and recommended arrest laws are estimated. The first 4 columns include the same controls as the specifications in Table 3. In column (5), in addition to the state-specific linear time trend, I add the state-specific quadratic time trend. Columns (6) and (7) are Poisson regressions, and the coefficients can be interpreted as approximate percentage change in mortality, after being multiplied by 100. Column (6) included state-specific linear time trends, and column (7) includes quadratic time trends as well.

Table 5: Estimated Impact of Mandatory and Recommended Arrest Laws on the Four Populations of Interest

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
African-American Women	34.16*	31.12**	31.12**	21.00**	14.89*	0.0389***	0.0268**
	(17.40)	(14.65)	(14.65)	(8.882)	(7.856)	(0.0149)	(0.0134)
African-American Men	52.37*	39.81	39.81	-14.33	-10.09	-0.00900	-0.0129
	(28.00)	(24.65)	(24.65)	(14.24)	(10.92)	(0.0192)	(0.0137)
White Women	-24.20	-7.966	-7.966	1.994	0.323	0.00461	-0.000953
	(20.11)	(11.35)	(11.35)	(5.119)	(4.306)	(0.00596)	(0.00483)
White Men	-16.08	-8.334	-8.334	-2.761	-1.383	-0.000752	-0.00202
	(17.69)	(14.36)	(14.36)	(6.948)	(5.478)	(0.00741)	(0.00532)
[1em] Population Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Economic Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Political Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Time Trend	No	No	Yes	Yes	Yes	Yes	Yes
State Time Trends	No	No	No	Yes	Yes	Yes	Yes
State Quadratic Trend	No	No	No	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1836	1800	1800	1800	1800	1800	1800

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The first 4 columns are include the same controls as the specifications in Table 3. In column (5), in addition to the state-specific linear time trend, I add the state-specific quadratic time trend. Columns (6) and (7) are Poisson regressions, and the coefficients can be interpreted as approximate percentage change in mortality, after being multiplied by 100. Column (6) included state-specific linear time trends, and column (7) includes quadratic time trends as well.

Part II

Teachers' Strikes and Standardized Test Scores: Impact on Performance & Participation

Abstract

Using data on standardized test scores for 4th and 7th graders, I employ a difference-in-differences approach to measure the impact on student outcomes of two province-wide teachers' strikes in 2005 and 2014 in British Columbia, Canada. Unlike previous work, I find that the impacts are not statistically significant on average across all schools. However, I find impacts on certain groups of students, particularly a statistically significant drop on Grade 4 female students' grades and participation rates. This drop is driven by the longer of the two strikes. Overall, the study shows that the impact of strikes on student performance is not statistically significant, and that previous results based on school-level data may suffer from over-rejection problems.

8 Introduction

Teachers' strikes may impact student learning negatively. This potential impact is part of an ongoing debate about teachers' unions and their right to strike. On one hand, student learning is important for society, and we would not desire that this learning be interrupted. On the other hand, union formation and strikes are an important right in contemporary society, and we would not want to deprive teachers of this right. Understanding how, and how much, teachers' strikes impact student scores is an important aspect of this discussion. In the current chapter, I estimate the effect of teachers' strikes on student success rates in standardized tests, namely, the Foundation Skills Assessment (FSA) exams administered in British Columbia. I do not find evidence of an impact on student success rates or the participation rates on average, but I do find evidence of an impact on participation and success rates of Grade 4 female students, and evidence of an impact on a smaller set of schools.

The main purpose of this chapter is to estimate the impact of teacher strikes on student achievement in standardized exams. The identification strategy of the current chapter relies on the fact that strikes impact public schools but not independent schools in British Columbia. The success rates and participation rates of students in public and independent schools in British Columbia follow similar trends.[13] The two largest strikes in the data set took place in the years 2005/2006 and 2014/2015. This treatment provides a quasi-experimental setting that can be exploited in a difference-in-differences approach: if success rates of public schools drop significantly in these years in comparison to those of independent schools, this provides evidence that strikes had a negative impact on students' academic performance.

The current chapter makes two contributions to the literature. First, I estimate the participation rates as well as the success rates for the students. This is important because students' drop in achievement can be masked by concerned parents who might not want to have their children take a standardized exam. Without testing whether there is a drop in participation, it is difficult to understand the full extent of the effect. The second contribution is about the use of standard error estimation, particularly when using school-level data, as it is commonplace to use "clustered" errors with difference-in-differences literature to avoid obtaining false positive results. However, recent studies show that this approach does not always mitigate the possibility of obtaining false positive results, it can actually make the problem worse. I show that this problem exists in the current setting, and address this problem by using placebo experiments to choose a reliable method to address the issue. When addressed in this manner, the results do not show a clear decrease of student performance or attendance, but a more nuanced and complex impact based on gender, grade and the chosen school sample. Furthermore, I show evidence that strikes may actually have a positive impact at least in certain cases.

The rest of the chapter is organized as follows. Section 9 gives background on teachers' strikes and standard examinations in British Columbia. Section 10 presents the current state of the literature. The details for the data are in Section 11. The estimation methodology, model and procedure, are

[13]See Figure 4 and figures 6 and 7.

explained in 12. The results are in Section 13, and Sections 14 and 15 present the discussion and conclusions.

9 Background: Teachers' Strikes and Standard Exams in British Columbia

British Columbia (BC) has a long history of teachers' unions and organizations. The labor union that represents all public school teachers, British Columbia Teachers' Federation (BCTF), was founded in 1917. However, the right for collective bargaining was obtained in 1987, with Bill 20, the Teaching Profession Act (BCTF, 2016c). BCTF conducted 32 local strikes and three lockouts until the right to strike was reduced considerably in 2001 with Bill 18, the Essential Services Legislation. The tension between government and the BCTF increased over the years, and finally, in the school year 2005/2006, BCTF voted to go on strike - the so-called "Illegal Strike". Their stated objectives were improved learning conditions, increased salary and restored bargaining rights. On September 28, BCTF started rotating strikes, and finally on October 7, BCTF started an indefinite strike. After 16 days of strike - 10 school days - BCTF reached an agreement with the government that included restoring the right to strike, and returned to work on October 24 (BCTF, 2016b). This strike and the associated school year 2005/2006 forms the first of the two treatment years in the current study.

The 2005/2006 "Illegal Strike" is in fact the second-longest strike in BC history. The longest strike took place in 2014, when, over similar concerns such as improved learning conditions, salary, and the possibility of constraints on bargaining rights, BCTF started a rotating strike in May 2014 and full strike in June. On September 2, the scheduled opening date for all schools, there was no resolution and the full strike continued. It finally ended on September 18 (Talmazan, 2014; BCTF, 2016a). This strike and the associated 2014/2015 school year forms the second of the two treatment years.

The impact of these strikes may show up on standardized exams in BC, as the province has an even longer history of standardized exams. The first exam was in 1876 for entrance into Victoria High School, the only high school at the time. Larger scale standardized testing began in the 1920s. Between 1924 and 1972, 34 different types of I.Q. tests were administered to students in British Columbia. In 1975, the Provincial Learning Assessment

Program (PLAP) replaced I.Q. tests and continued until 1999, when it was replaced by the Foundation Skills Assessment (FSA) (Raptis and Fleming, 2006). The FSA is an annual province-wide assessment of students' academic performance in BC, held at different times every year. FSA consists of three different assessments: Numeracy, Reading and Writing (British Columbia, 2020). The success rates and participation rates on the standardized FSA examinations are the two outcome variables in the current chapter.

10 Literature Review

Early studies of teachers' strikes find a negative impact on student achievement (Thornicroft, 1994; Caldwell and Moskalski, 1981; Caldwell and Jeffreys, 1983; Zirkel, 1992). These earlier studies are susceptible to omitted variable bias: unobserved school-level characteristics can be systematically different for schools that were impacted by the strikes, or events particular to a school or school district may trigger strikes. More recent studies make various attempts to control for omitted variables. Zwerling (2008), for instance, uses lagged variables to control for existing district-level characteristics, and finds no significant impacts of strikes on school attendance or on graduation rates. Webbink and Belot (2010) use a difference-in-differences strategy to estimate the impact of strikes. They exploit the difference in the Flemish and the French communities in Belgium and find that teachers' strikes increase class repetition and decrease the total years of schooling. Johnson (2011) and Baker (2013) use very similar econometric models to estimate the impact of teachers' strikes on standardized test scores in Ontario. While Johnson (2011) focuses on a larger set of labor disruptions and uses school-level fixed effects to control for unobservables, Baker (2013) focuses only on strikes and uses first differencing to control for school-level effects. They both find statistically significant negative impacts on academic achievement in standardized tests in Maths. Baker (2013) also reports on the impact of strikes by gender and find differences between male and female students. Jaume and Willen (2019) exploit the variation in districts and years in Argentina in a difference-in-differences framework. They conclude that strikes decrease wages, labor force participation, and college enrollment rate. They also conclude that strikes increase the probability of having children and increase household production for women. Along with Webbink and Belot (2006), Jaume and Willen (2019) research the impact of teachers' strikes on long-term labor market outcomes, whereas Johnson (2011) and Baker (2013) research the short-term effects on immediate academic performance.

Currently, the literature remains unclear on the impact of strikes. On one side of the dispute are the papers that find no impact and conflicting results, such as Zwerling (2008), Thornicroft (1994), Caldwell and Moskalski (1981), Caldwell and Jeffreys (1983) and Zirkel (1992). Most of these papers do not offer a clear identification strategy and present cross-sectional results. The only exception is Zwerling (2008) who uses lagged variables, which results in a model that is very close to the difference-in-differences approach, even though it is not explicitly called such. Johnson (2011) and Baker (2013) use various difference-in-differences approaches to estimate the impact of strikes on academic performance. While difference-in-differences is valuable for addressing the omitted variables issues, it can yield false positive results due to the standard error estimation (Bertrand et al., 2003). While the usual method of dealing with these issues is to use "clustered" standard errors, recent studies show that this does not necessarily solve these issues in all cases (Cameron and Miller, 2015; Abadie et al., 2010; MacKinnon, 2019). In the current study, I show that it is easy for school-level data to suffer from these issues and that "clustered" standard errors do not address the issue. Following Bertrand et al. (2003), I use "placebo-law" experiments to find the estimation methodology that suffers little from the risk of false positive estimates.

A related line of research focuses on the teachers' unionization and union bargaining power.[14] As teachers' unionization is indirectly related to teachers' strikes, I include some of the important papers in this line of literature for completeness. The literature in this area is extensive, as a recent review by Cowen and Strunk (2015) lists 39 articles. While the entirety of this line of research is beyond our scope, we present a short summary. The literature focuses on the impact of unionization and union bargaining power on three types of outcomes: teachers' wages and earnings (Rose and Sonstelie, 2010; Zigarelli, 1996; Winters, 2011; Lovenheim, 2009), district spending on education (Brunner and Squires, 2013; Chambers, 1977; Duplantis et al., 1995; Gallagher, 1979; Eberts and Stone, 1986; Hoxby, 1996), and student achievement. Among these, the student achievement is closest to our line of

[14]Unionization is typically defined as 50% of teachers being in a union for a given state or district. The bargaining power of unions is proxied by average union income per teacher or spending per student.

research. Earlier studies find that unionization or collective bargaining agreements are associated with higher academic outcomes for students (Freeman and Ichniowski, 1988; Register and Grimes, 1991; Carini et al., 2000), but these studies are susceptible to omitted variable bias (Zirkel, 1992). More recent studies use various identification strategies and have mixed results: Hoxby (1996) combines a difference-in-differences approach with instrumental variables and finds that unionization increases dropout rates, particularly for low-performing students. Lott and Kenny (2013) exploit differences in timing of state law and also find a negative impact of union resources per student on student performance. In contrast, Lovenheim (2009) finds no impact of unionization on student performance with a difference-in-differences approach. While these papers focus on academic performance, Lovenheim and Willen (2019) research the long-term effects of collective bargaining laws. They use a difference-in-differences approach and exploit the variation in the passage of the laws, and find that laws that support unionization decrease earnings, labor participation, and hours worked.

Another related line of research is about the marginal impact of days in a school year, on academic and economic outcomes. Card and Krueger (1992) find that the number of school days in a year do not impact future earnings, after controlling for state fixed effects. Lee and Barro (2001) report the associations of days in school year with standardized test scores. Pischke (2007), Marcotte and Hemelt (2008), Sims (2008), Hansen (2011), Aucejo and Romano (2016) and Fischer et al. (2017) find that additional days in a school year increase academic performance. With the exception of Lee and Barro (2001) and Marcotte and Hemelt (2008), all of the literature uses an identification strategy. In particular, Hansen (2011) exploits weather-related changes with an instrumental variables methodology, and Marcotte and Hemelt (2008) exploits the variation in unscheduled closings. The rest of the papers use difference-in-differences or a similar strategy based on exogenous variation in school days across time or across panels. A related line of research involves the decrease from five days to four days of schooling per week. In contrast to previous papers, Anderson and Walker (2015) find that reducing the days in a week from five to four increases academic performance, based on a school-level panel data set from Colorado.

11 Data

I use a data set of FSA results at the school level, procured directly from the BC Ministry of Education. This data set includes all years from 1999/2000 to 2018/2019, and all the schools in British Columbia. It contains the assessment results for Numeracy, Reading and Writing, for both Grades 4 and 7. It is further broken down into different sub-populations of students by gender, aboriginal status and English-as-a-second-language status. Each school can be either independent or public.[15] As mentioned earlier, the strike impacted only the public schools. As such, public schools serve as the treatment group, and independent schools serve as the control group.

The two outcome variables are the success rates and the participation rates. Participation rates are an important part of the story, as parents may opt-out of the exam if they feel that their child is not ready due to the impact of the strikes. In fact, the BCTF is against the FSA: since 2008, they have been sending parents a letter every year to explain why they are against it, and encourages the parents to opt out of the FSA. Parents might have elevated concerns around the time of a strike concerning their child's well-being, especially if their child is struggling. Therefore, I include the participation rates as a second variable of interest. To construct success rates, I add the number of students exceeding and meeting expectations, and divide by the number of students writing the exam. To get the participation rate, I divide the number of students writing the exam by the number of expected writers. The descriptive stats are in Table 6.

There are a couple of issues with the data set: first, the BC Ministry of Education has reported that the figures from the year 2007/2008 suffered from data collection issues and are not reliable, so I drop the year 2007/2008. Second, some schools went through a reorganization and merged, so I collapse these schools into one. The results in the current analysis are robust to removing these schools from the analysis.[16] A final issue is that participation can exceed the number of allocated students in previous years, but not all. I deal with this irregularity by marking these data points, and am able to keep them out of the analysis. The results are robust to removing these data points. Finally, in the year 2008, the BC Ministry of Education made

[15] Two schools serve as an exception to this categorization. They are marked as the Department of Indian Affairs and Northern Development. These two schools are dropped from the analysis.

[16] See the appendix for further details on the data cleaning.

Table 6: Descriptive Statistics

	Grade	Skill	Year Count	School Count	District Count	Success Rate	Participation Rate
Independent	4	Numeracy	20	237	51	0.89 (0.15)	0.97 (0.07)
		Reading	20	237	51	0.89 (0.13)	0.97 (0.07)
		Writing	20	237	51	0.93 (0.11)	0.97 (0.08)
	7	Numeracy	20	239	51	0.88 (0.15)	0.97 (0.07)
		Reading	20	239	51	0.88 (0.13)	0.97 (0.07)
		Writing	20	239	51	0.93 (0.11)	0.97 (0.08)
Public	4	Numeracy	20	1430	60	0.79 (0.17)	0.92 (0.14)
		Reading	20	1431	60	0.78 (0.15)	0.92 (0.14)
		Writing	20	1430	60	0.87 (0.14)	0.91 (0.15)
	7	Numeracy	20	1523	60	0.76 (0.18)	0.92 (0.15)
		Reading	20	1523	60	0.75 (0.15)	0.92 (0.15)
		Writing	20	1523	60	0.84 (0.15)	0.91 (0.16)

Table 6 shows the descriptive statistics of the two outcome variables in the current paper. The success rates and participation rates are both presented as between 0 and 1, 1 meaning full success/participation. The figures are population averages, whereas the standard deviations are given in parentheses.

fundamental changes to how they implement, score and report the outcomes for the FSAs.[17] The impact of this change in reporting is visible in the data - see Figure 4. To control for this change, a binary variable is included to mark all the years after 2008 in the methodology. For the most part, the results are robust to removing this term.

As I explain in Section 12, when left at the school level, the specifications are likely to lead to a false rejection of the null hypothesis. This false rejection is relatively easy to see when I run placebo-law experiments: since my treatment variable is defined by two strike years, I am able to use the remaining year pairs as a check on the validity of the estimates. If the two strike years are in fact significantly different than the other pairs of two years in the set, the other pairs of years should not yield statistically significant estimates. Table 8 shows that this is the case only when the data is collapsed at the district level. When left at the school level, any arbitrary pair of two years yields statistically significant results 40% of the time. This is, of course, misleading: in the absence of an actual impact, I should obtain statistically significant results at the 0.05 level only 5% of the time. In other words, to be confident in my results, the two strike years need to be fundamentally dif-

[17]Before 2008, the scoring was done centrally. After 2008, the local districts became responsible for scoring. Furthermore, the administration mode has changed from paper-based to computer-based (Simon et al., 2012)

ferent then the other pairs of two years - not similar to 40% of them. Based on this analysis, I collapse the data to the district level. This ensures that my estimates are actually valid and not just a mechanical construct of the statistical error estimation methodology.

In order to protect the privacy of individual students, the BC Ministry of Education does not report on smaller schools and subpopulations of students. This means that some schools are not represented in all grades and skills for all years. For example, even though there are 1744 schools in the data set, only 1667 of these schools have at least one year for Grade 4 Numeracy, and among these, only 1055 have all 20 years. It is possible to use either the full-but-imbalanced data set, or the smaller-but-balanced subset of schools. When the full data set is used, all 1667 schools are included, but many schools have missing years. I also run the regressions on a "balanced" data set. With a balanced data set, only the 1055 schools are included, but all of these schools have data for all years. The advantage of using a balanced data set is that it avoids the statistical noise that can stem from the missing years. The disadvantage, on the other hand, is that the results are representative only of schools with larger numbers of students, as these schools more likely to have no missing years from the data set. While most of my discussions are based on the full sample, I also include the results from the balanced data set in the last two columns of all results tables, and discuss these briefly. The results are different for this set of schools.

Figure 4 shows the timeline of the average success rate. The average success rate for independent schools and the public schools follow similar trends. This provides a visual check on the identifying parallel trends assumption. The rest of the figures for success rates and participation rates are in the Appendix in Section 15. For most of the success rate timelines, especially for Grade 4 Numeracy, the independent and public schools follow parallel trends. Participation rates, however, start deviating right before the 2014/2015 strike, and can be an concern for the identification assumption. Even in this case, the participation rates follow similar trends up until the later treatment, the second strike. In other words, even in cases where the parallel trends assumption may be violated, the trends prior to treatment are similar to each other. However, it should still be noted that other treatments, in particular, the changing political attitudes towards standardized testing, may explain part of the drop in participation. I discuss this possibility in further detail in Section 15.

The averages and standard deviations for a breakdown of schools by their

Figure 4: Average Success Rates for Grade 4 Numeracy Tests

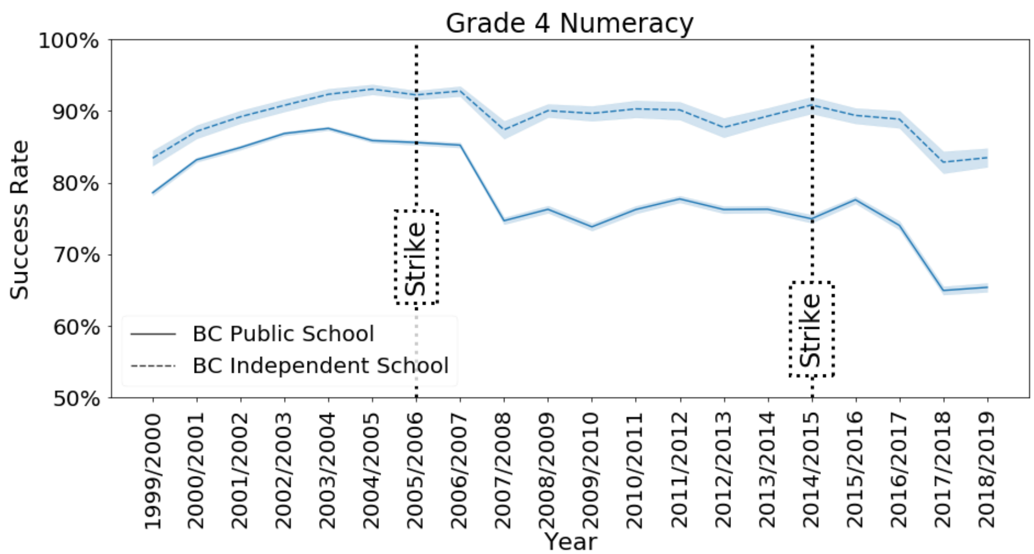


Figure 4 shows the average success rate of schools over time. Success rate is calculated as the rate of students meeting or exceeding expectations, divided by the total number of students writing the exam. The two lines represent the BC Public Schools and the BC Independent Schools. The average is calculated across schools.

public vs independent status and the strike vs non-strike years are given in Table 7. This table provides a cleaner look into the data. The differences between strike years and non-strike years in success rates are relatively similar across public and independent schools.

12 Methodology

We employ a difference-in-differences strategy to identify the effects of teachers' strikes. My baseline equation is as follows:

$$Y_{dpt} = \beta_0 + \beta_1 p_{dp} + \beta_2 s_t + \beta_3 p_{dp} s_t + \varepsilon_{dpt} \tag{3}$$

The unit of observation is a set of schools that share the same district and the same status as either a public or an independent school. The subscript *dp* designates this unit of observation: all schools within a particular district that share the same status as a public or independent school are collapsed into one unit and are represented with this subscript. p_{dp} is a binary variable that takes its value based on whether a given set of schools is public or independent. The binary variable s_t shows whether a particular year had a strike or not: it takes the value 1 for the two strike years, 2005/2006 and 2014/2015, and 0 for all other years. $p_{dp}s_t$ is the interaction of these two binary variables. Its coefficient, β_3, is the main regressor of interest, and captures the impact of the strike.

The strikes impact all public schools at the same time, because of this, the residuals from Equation (3) will be correlated (Bertrand et al., 2003). To deal with this problem, most papers in literature use "clustered" standard error estimation. However, while popular, "clustered" standard errors do not always solve the problem of over-rejection (Angrist and Pischke, 2008; Abadie et al., 2010; Bester et al., 2011). In fact, "clustered" standard errors can produce very misleading results when data does not satisfy certain conditions (MacKinnon, 2019; Cameron and Miller, 2015). In the current setting, the largest unit of observation in the data hierarchy are districts. Clustering at the district level will remove the effect of the serial correlation within the districts, but not across the districts, and because of this, the i.i.d[18] assumption in standard error calculation is violated. Furthermore, the success

[18]Independent and identically distributed: here, the residuals are no longer independent of each other.

Table 7: Public and Independent Schools in Strike vs Non-Strike Years

		Participation Rate				Success Rate			
		BC Independent School		BC Public School		BC Independent School		BC Public School	
		Control Years	Strike Year	Control Years	Strike Year	Control Years	Strike Year	Control Years	Strike Year
Grade	Skill								
4	Numeracy	0.97	0.97	0.92	0.91	0.89	0.92	0.79	0.81
	Reading	0.97	0.97	0.92	0.91	0.89	0.91	0.78	0.79
	Writing	0.97	0.97	0.91	0.91	0.93	0.94	0.87	0.87
7	Numeracy	0.97	0.97	0.92	0.91	0.88	0.90	0.76	0.78
	Reading	0.98	0.97	0.92	0.91	0.88	0.87	0.76	0.74
	Writing	0.97	0.97	0.91	0.91	0.93	0.95	0.84	0.87

Table 7 shows a tabulation of success rates across public schools and independent schools in strike and non-strike years. Public (treated) and independent schools (control) schools between strike (treated) and (non-treated) are similar to each other.

rates within schools and across years are correlated. It possible for school level data sets to suffer from larger correlations across clusters. Previous papers in the teachers' strike literature either do not discuss their standard error estimation methodology, or use "clustered" standard errors and make no detailed explanations. In the current study, I investigate various options to choose the optimal methodology.

To choose a valid methodology, I follow the placebo-law methodology first used by Bertrand et al. (2003). Placebo-law experiments involve changing the treatment variable across repetitions of the same regression. If the standard error estimation is not biased, I expect these regressions to yield a rejection of the null hypothesis approximately 5% of the time, with a p-value less than 0.05. In the current setting, the treatment variable is constructed by the two strike years, 2005/2006 and 2014/2015. Since the 2007/2008 is dropped from the data set due to data collection issues, this leaves 17 control years. From the 17 years, I construct 136 pairs of years that do not include the treatment years. These pairs of two years serve as the basis for alternative treatments. Furthermore, since each year has assessment examinations for two grades (4 and 7) and three examinations, (Numerical, Reading and Writing) I construct 816 alternative null hypotheses. If the strike years truly are different than the others, these alternative null hypotheses should generate statistically significant ($p < 0.05$) results 5% of the time, or around 40-41 of the 816 alternative iterations.

Table 8 shows the results from the placebo experiments. Here, the three columns on the right are estimations from school-level data sets, which is similar to most of the literature. The two leftmost columns are results after collapsing the data set at the district level, these columns correspond to the specification in Equation (3). School level estimations suffer from a clear over-rejection problem. "Clustered" standard error estimation, while popular, does not remove this problem in the current setting. This is consistent with warnings in the standard error estimation literature (Cameron and Miller, 2015; MacKinnon, 2019). Based on the results from this table, I deviate from most of the literature and collapse the data at the district level before running the regressions.

Table 8: Results from Placebo Experiments: Rates Rejection Under the Null Hypothesis of No Impact

School-level			District-level	
No Clustering	Clustering by District	2-way Clustering	No Clustering	Clustering by District
318 (39%)	360 (44%)	329 (40%)	41 (5%)	98 (12%)

Counts show the number of times the placebo estimation yielded a "statistically significant" estimation with a p-value of less than 0.05. If the standard error estimation is not biased, these regressions should reject the null hypothesis approximately 5% of the time, with a p-value less than 0.05. There are 6 classes and 132 two-year pairs, resulting in 816 placebo estimations. For a specification that does not suffer from an over-rejection problem, this corresponds to 40-41 in absolute value of the estimations with a p-value of less than 0.05. In the case of school-level regressions, the 2-way clustering refers to the method detailed in Cameron et al. (2008), and is by district and year.

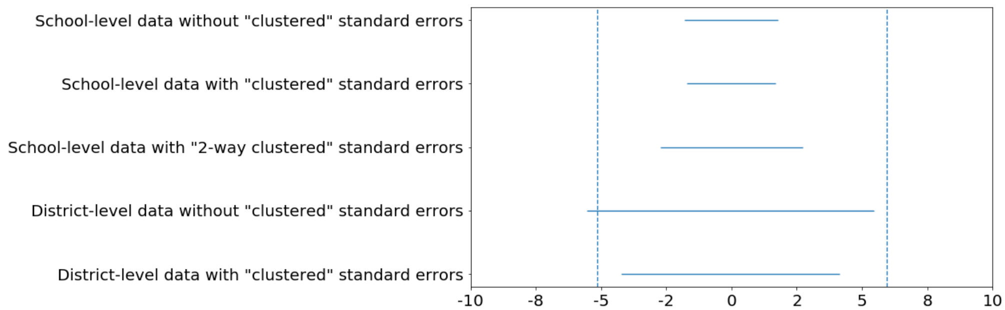
Figure 5 shows the results from Table 8 visually. The two vertical dashed lines show the actual 95% confidence interval from the distribution of the coefficient estimates from the placebo estimations. The horizontal error bars show the 95% confidence internal calculated from the average of the coefficient estimates and the average of the standard error estimates. The specifications that do not fail from an over-rejection problem line up reasonably close to the dotted lines. In contrast, the specifications that suffer from an over-rejection problem are smaller.

When reporting the results, I extend the Equation (3) as follows: In Equation (4), I add the binary variable r_t that takes the value 1 beginning with the school year 2008/2009. This variable captures the change in the FSA exam that happened on the year 2008/2009. As can be seen on the Figure 4, the change had a visible and sustained impact on the scores. In Equation (5), I add the year fixed effects. Finally, in Equation (6), I add both the year and district fixed effects. The results are typically not sensitive to these different specifications. However, they are sensitive to standard error estimation, and the choice between a balanced sample of schools versus the full population of schools. I report the results for all specifications with the full sample of schools, with standard errors estimated with and without "clustering". For the balanced set, I include only the final specification. The mathematical expressions for these models are presented below:

$$Y_{dpt} = \beta_0 + \beta_1 p_{dp} + \beta_2 s_t + \beta_3 p_{dp} s_t + \beta_4 r_t + \varepsilon_{dpt} \tag{4}$$

$$Y_{dpt} = \beta_0 + \beta_1 p_{dp} + \beta_2 s_t + \beta_3 p_{dp} s_t + \beta_4 r_t + \mu_t + \varepsilon_{dpt} \tag{5}$$

Figure 5: Confidence Intervals from the Placebo Analysis: Placebo Intervals vs Theoretical Intervals



The two dashed lines show the boundaries of 95% of the coefficient estimates from the placebo runs. Each one of the error specifications show the 95% confidence interval based on the average coefficient and the average standard error estimations for the corresponding specification. The school-level specifications, even when “clustered” or “two-way clustered” error estimates are used, suffer from an over-rejection problem. Collapsing the data at the district level resolves this issue.

$$Y_{dpt} = \beta_0 + \beta_1 p_{dp} + \beta_2 s_t + \beta_3 p_{dp} s_t + \beta_4 r_t + \mu_t + \alpha_d + \varepsilon_{dpt} \tag{6}$$

In Equation (4), I add the term $\beta_4 r_t$. As explained in Section 11, BC Ministry of Education made changes to the administration and scoring of the FSA in 2008. The variable r_t controls this change, and the β_4 is the corresponding coefficient. While β_4 is statistically significant, the results are robust to removing this term. In Equation (5), I add controls for each year in the data set - this controls for any year-specific unobservable factors non-parametrically. Finally, in Equation (6), I add controls for each district in the data set, allowing us to control for any unobservable characteristics specific to the district.

In particular, adding the year fixed effects μ_t is important: the BC Ministry of Education web site makes a point that year-to-year differences in scores may not be reflective of an actual change. By using year fixed effects, I am able to control for any unobservable changes that may have impacted both independent and public school districts. For instance, if there is a systematic, non-random impact on scores related to classroom composition, the binary variables for year fixed effects will control for this change. This model design allows me to be able to make a year-to-year comparison.

13 Results

Table 9 shows the main results of the current chapter. Since leaving the data at the school level causes a severe over-rejection problem, all of these results come from the estimations aggregated at the district level, as shown in Equations (3) through (6). The column numbers correspond to the equation numbers. In the full sample of schools, there is no evidence of an impact of the strikes. The first four columns report the standard errors with clustering, and the following four columns without clustering. The coefficient estimates can be interpreted as the change in the success rate - the percentage of students who meet or exceed expectations. The average estimated impact of the strikes on these rates is positive for Grade 7 Numeracy and Writing, negative in all other cases, and is not statistically significant. I interpret this as little or no impact of strikes across all districts.

I also focus my results on a smaller set of schools. To protect student privacy, success rates for some years are suppressed by the BC Ministry of Education. This causes missing years for certain schools, particularly schools with a smaller number of students. When aggregated, the missing years may cause extra noise in the data. To deal with this problem, I choose a set of schools that have the full data. The two rightmost columns are based on aggregations of schools that have full data for all years. This can make it easier to detect impacts that are smaller in magnitude. However, it also restricts the analysis to only schools with a large enough number of students. I include the results from this balanced set of schools for completeness in the last two columns. In this case, the coefficient estimates are always negative and statistically significant for Grade 7 Numeracy and Writing.

I interpret the results from Table 9 as a nuanced picture: on average, there is little or no impact of strikes on schools. This interpretation is based on both the statistical significance of the coefficients and the fact that the coefficients are small in magnitude, and in two cases, Grade 7 Numeracy and Writing, positive. However, the same estimations for a smaller set of schools (those with larger number of students) show not just some statistically significant results, but also coefficients that are larger in magnitude, and having the same direction for all specifications and different grades and skills. I discuss further in the conclusions section.

Table 10 shows the estimated impact of the strikes on participation rates. Similar to Table 9, the strikes do not have a significant impact on the participation rates overall. However, for the balanced set of schools, there is a statistically significant impact on the Grade 4 Numeracy participation rates.

Taken together, Table 9 and Table 10 do not show evidence of a average drop in success rates or participation rates across all schools. However, this may be due to increased variance due to the missing data related to the privacy policies of the BC Ministry of Education. The results based on the balanced data set, on the two rightmost columns, address this issue. In this case, I estimate statistically significant drops in most grade 7 scores and statistically significant drops in most Grade 4 participation rates. I discuss the implications of these findings in Section 14.

Table 11 shows the estimated impact of the strikes by gender. The strikes had a negative and statistically significant impact on the success rates of female students for all three skill sets. This impact remains negative and statistically significant, even with different specifications or standard error estimations. Just like in Table 9, I run the same analysis for the balanced set of schools. In this case, the impact estimate remains negative. In particular, the estimate for Reading remains statistically significant.

Table 12 shows that participation dropped for Grade 4 female students for all subjects. Taken together with the results in 11, the success rates and participation rates for Grade 4 female students drop simultaneously. This is arguably the most robust finding in the current study.

The two strikes may have had different impacts. To research this possibility, I amend the variable s_t by changing it into two different variables for each strike year. Tables 13 and 14 show the results broken down by each strikes, respectively, for success rates and for participation rates. Overall, for the full data set, there is no significant drop in success rates except for Grade 4 Reading scores for the year 2005/2006. For participation rates, there is a drop in Grade 7 Reading scores, but this drop is not consistent across all standard error specifications.

Using the balanced set presents a different but clearer picture. The two rightmost columns of Tables 13 and 14 show that there is an estimated increase in both participation and success rates during the first strike, and a decrease in both success rates and participation rates in the second strike. In particular, the decrease in participation in the 2014/2015 strike is statistically significant both grades and all subjects, and the decrease is success rates is statistically significant for all subjects in Grade 7.

Overall, the results do not present clear evidence of an impact of strikes in academic performance. The impact of the success rate is less than 3% in absolute value, and positive in one case. It is also not statistically significant.

Table 9: The Impact of Strikes on Success Rates

		All Available Data								Balanced Set of Schools	
		Clustered				Not Clustered				Clustered	Not Clustered
	Specification	(3)	(4)	(5)	(6)	(3)	(4)	(5)	(6)	(6)	(6)
Grade	Skill										
4	Numeracy	-1.18 [1.96]	-1.55 [1.92]	-1.79 [1.92]	-0.26 [1.83]	-1.18 [3.32]	-1.55 [3.10]	-1.79 [3.07]	-0.26 [2.32]	-2.22 [1.67]	-2.22 [1.95]
	Reading	-2.71 [2.05]	-2.73 [2.05]	-2.94 [2.05]	-1.86 [1.86]	-2.71 [3.04]	-2.73 [3.03]	-2.94 [2.99]	-1.86 [2.28]	-1.58 [1.69]	-1.58 [1.93]
	Writing	-2.07 [2.07]	-2.17 [2.02]	-2.16 [2.02]	-1.12 [1.86]	-2.07 [2.85]	-2.17 [2.70]	-2.16 [2.69]	-1.12 [2.17]	-1.95 [1.79]	-1.95 [2.07]
7	Numeracy	1.32 [2.15]	1.30 [2.12]	1.29 [2.12]	0.39 [1.92]	1.32 [3.67]	1.30 [3.48]	1.29 [3.46]	0.39 [2.45]	-3.36 [1.58] (**)	-3.36 [1.90] (*)
	Reading	-1.85 [1.51]	-1.86 [1.50]	-1.84 [1.50]	-2.41 [1.53]	-1.85 [2.78]	-1.86 [2.78]	-1.84 [2.76]	-2.41 [2.03]	-3.18 [1.84] (*)	-3.18 [1.94]
	Writing	2.56 [2.06]	2.52 [2.04]	2.29 [1.99]	2.66 [2.05]	2.56 [3.08]	2.52 [3.05]	2.29 [3.01]	2.66 [2.44]	-4.32 [1.88] (**)	-4.32 [2.01] (**)

The estimates for the average impact of the strikes on success rates are given in Table 9. Each row comes from a different regression. Column numbers refer to the corresponding equation: columns numbered (3) show the estimates from the baseline difference-in-differences equation, columns numbered (4) show the estimates with the control added for the change in the FSA in 2008. Columns numbered (5) show the estimates with the year fixed effects added and columns numbered (6) add the district fixed effects. The coefficient estimates can be interpreted as the change in the success rate - the percentage of students who meet or exceed expectations.

Table 10: The Impact of Strikes on Participation Rates

		All Available Data								Balanced Set of Schools	
		Clustered				Not Clustered				Clustered	Not Clustered
	Specification	(3)	(4)	(5)	(6)	(3)	(4)	(5)	(6)	(6)	(6)
Grade	Skill										
4	Numeracy	-1.36 [0.84]	-1.50 [0.85] (*)	-1.34 [0.84]	-1.18 [0.87]	-1.36 [1.39]	-1.50 [1.39]	-1.34 [1.40]	-1.18 [1.30]	-1.70 [0.72] (**)	-1.70 [0.96] (*)
	Reading	-0.75 [0.84]	-0.83 [0.84]	-0.73 [0.83]	-0.72 [0.84]	-0.75 [1.41]	-0.83 [1.39]	-0.73 [1.40]	-0.72 [1.31]	-1.40 [0.73] (*)	-1.40 [0.94]
	Writing	-0.68 [1.17]	-0.76 [1.17]	-0.67 [1.15]	-0.48 [1.14]	-0.68 [1.83]	-0.76 [1.79]	-0.67 [1.78]	-0.48 [1.60]	-1.40 [0.91]	-1.40 [1.02]
7	Numeracy	-1.02 [0.94]	-1.02 [0.95]	-0.93 [0.93]	-1.30 [0.95]	-1.02 [1.65]	-1.02 [1.63]	-0.93 [1.65]	-1.30 [1.36]	0.02 [0.82]	0.02 [1.16]
	Reading	-0.71 [0.89]	-0.75 [0.90]	-0.63 [0.89]	-0.73 [0.87]	-0.71 [1.58]	-0.75 [1.57]	-0.63 [1.59]	-0.73 [1.36]	-0.15 [0.79]	-0.15 [1.15]
	Writing	-0.55 [1.14]	-0.69 [1.08]	-0.55 [1.07]	-0.58 [1.05]	-0.55 [2.10]	-0.69 [2.03]	-0.55 [2.03]	-0.58 [1.71]	0.19 [0.84]	0.19 [1.17]

The estimates for the average impact of the strikes on participation rates are given in Table 10. Each row comes from a different regression. Column numbers refer to the corresponding equation: columns numbered (3) show the estimates from the baseline difference-in-differences equation, columns numbered (4) show the estimates with the control added for the change in the FSA in 2008. Columns numbered (5) show the estimates with the year fixed effects added and the columns numbered (6) add the district fixed effects. The coefficient estimates can be interpreted as the change in the participation rate.

Table 11: The Impact of Strikes on Success Rates, by Gender

			All Available Data								Balanced Set of Schools	
			Clustered				Not Clustered				Clustered	Not Clustered
		Specification	(3)	(4)	(5)	(6)	(3)	(4)	(5)	(6)	(6)	(6)
	Grade	Skill										
Female	4	Numeracy	-5.45 [2.69] (**)	-6.80 [2.69] (**)	-7.01 [2.73] (**)	-5.99 [2.43] (**)	-5.45 [2.83] (*)	-6.80 [2.65] (**)	-7.01 [2.69] (***)	-5.99 [2.32] (***)	-3.36 [2.29]	-3.36 [2.33]
		Reading	-6.66 [2.57] (**)	-6.97 [2.60] (***)	-7.18 [2.63] (***)	-6.37 [2.35] (***)	-6.66 [2.66] (**)	-6.97 [2.64] (***)	-7.18 [2.64] (***)	-6.37 [2.28] (***)	-4.58 [2.37] (*)	-4.58 [2.31] (**)
		Writing	-5.34 [3.07] (*)	-6.56 [3.02] (**)	-6.53 [3.01] (**)	-5.67 [2.82] (**)	-5.34 [3.21] (*)	-6.56 [3.09] (**)	-6.53 [3.07] (**)	-5.67 [2.80] (**)	-3.54 [2.84]	-3.54 [2.88]
	7	Numeracy	1.74 [2.58]	1.72 [2.44]	1.70 [2.41]	1.81 [2.14]	1.74 [3.51]	1.72 [3.29]	1.70 [3.30]	1.81 [2.61]	-1.09 [2.40]	-1.09 [2.88]
		Reading	-0.74 [2.76]	-0.73 [2.73]	-0.74 [2.70]	-0.82 [2.49]	-0.74 [3.50]	-0.73 [3.45]	-0.74 [3.43]	-0.82 [2.88]	-4.06 [2.88]	-4.06 [2.89]
		Writing	-1.05 [2.84]	-1.03 [2.81]	-1.07 [2.78]	-1.33 [2.62]	-1.05 [3.36]	-1.03 [3.30]	-1.07 [3.26]	-1.33 [2.84]	-5.37 [2.49] (**)	-5.37 [2.66] (**)
Male	4	Numeracy	0.94 [2.94]	0.56 [2.70]	0.51 [2.65]	-0.96 [2.14]	0.94 [3.86]	0.56 [3.61]	0.51 [3.59]	-0.96 [2.86]	-1.11 [2.35]	-1.11 [3.05]
		Reading	2.47 [2.80]	2.38 [2.73]	2.28 [2.70]	0.83 [2.31]	2.47 [3.47]	2.38 [3.44]	2.28 [3.41]	0.83 [2.71]	1.44 [2.59]	1.44 [2.91]
		Writing	0.54 [2.58]	0.21 [2.37]	0.16 [2.39]	-0.62 [2.16]	0.54 [3.33]	0.21 [3.19]	0.16 [3.18]	-0.62 [2.63]	-0.24 [2.38]	-0.24 [2.79]
	7	Numeracy	-3.83 [2.05] (*)	-3.43 [1.89] (*)	-3.29 [1.86] (*)	-3.24 [1.88] (*)	-3.83 [2.94]	-3.43 [2.73]	-3.29 [2.77]	-3.24 [2.27]	-5.54 [1.79] (***)	-5.54 [2.42] (**)
		Reading	-1.11 [2.17]	-1.04 [2.15]	-1.02 [2.14]	-0.87 [2.11]	-1.11 [2.65]	-1.04 [2.65]	-1.02 [2.68]	-0.87 [2.35]	-2.26 [1.93]	-2.26 [2.50]
		Writing	0.99 [2.13]	1.07 [2.11]	1.06 [2.09]	1.12 [2.12]	0.99 [2.77]	1.07 [2.76]	1.06 [2.78]	1.12 [2.38]	-3.17 [2.42]	-3.17 [2.60]

The estimates for the average impact of both strikes on participation rates are given in Table 11. Column numbers refer to the corresponding equation. The estimates come from an extension of the equations (3) through (6) where the gender of the students are controlled by a binary variable and the impact is estimated by interacting the gender variable with the strike variables. In particular, columns numbered (3) show the estimates from the baseline difference-in-differences equation with the addition of a binary variable indicating gender, columns numbered (4) show the estimates with the control for the change in the FSA in 2008. Columns numbered (5) show the estimates with year fixed effects and columns numbered (6) show the estimates with district fixed effects. The coefficient estimates can be interpreted as the change in the success rate - the percentage of students who meet or exceed expectations.

Table 12: The Impact of Strikes on Participation Rates, by Gender

			All Available Data								Balanced Set of Schools	
			Clustered				Not Clustered				Clustered	Not Clustered
		Specification	(3)	(4)	(5)	(6)	(3)	(4)	(5)	(6)	(6)	(6)
	Grade	Skill										
Female	4	Numeracy	-3.52 [0.99] (***)	-4.06 [1.02] (***)	-3.87 [1.02] (***)	-3.58 [0.98] (***)	-3.52 [1.28] (***)	-4.06 [1.25] (***)	-3.87 [1.28] (***)	-3.58 [1.19] (***)	-2.34 [0.84] (***)	-2.34 [1.21] (*)
		Reading	-3.21 [1.03] (***)	-3.71 [1.05] (***)	-3.53 [1.06] (***)	-3.14 [1.01] (***)	-3.21 [1.29] (**)	-3.71 [1.25] (***)	-3.53 [1.28] (***)	-3.14 [1.20] (***)	-2.17 [0.88] (**)	-2.17 [1.23] (*)
		Writing	-3.31 [1.22] (***)	-3.94 [1.25] (***)	-3.76 [1.25] (***)	-3.29 [1.20] (***)	-3.31 [1.46] (**)	-3.94 [1.40] (***)	-3.76 [1.42] (***)	-3.29 [1.36] (**)	-2.11 [1.07] (*)	-2.11 [1.39]
	7	Numeracy	-1.49 [1.24]	-1.50 [1.19]	-1.45 [1.20]	-1.35 [1.06]	-1.49 [1.97]	-1.50 [1.88]	-1.45 [1.89]	-1.35 [1.64]	-1.83 [1.03] (*)	-1.83 [1.66]
		Reading	-1.14 [0.85]	-1.10 [0.83]	-1.06 [0.83]	-1.07 [0.79]	-1.14 [1.77]	-1.10 [1.71]	-1.06 [1.71]	-1.07 [1.47]	-1.41 [0.93]	-1.41 [1.63]
		Writing	-1.93 [0.99] (*)	-1.86 [0.98] (*)	-1.81 [0.97] (*)	-1.82 [0.90] (**)	-1.93 [1.86]	-1.86 [1.77]	-1.81 [1.78]	-1.82 [1.53]	-1.76 [1.03] (*)	-1.76 [1.64]
Male	4	Numeracy	-0.51 [0.99]	-0.66 [0.98]	-0.56 [0.99]	-1.22 [0.96]	-0.51 [1.60]	-0.66 [1.59]	-0.56 [1.59]	-1.22 [1.39]	-1.04 [1.06]	-1.04 [1.40]
		Reading	-0.14 [0.99]	-0.28 [0.96]	-0.18 [0.98]	-0.89 [0.94]	-0.14 [1.56]	-0.28 [1.54]	-0.18 [1.54]	-0.89 [1.35]	-0.62 [1.07]	-0.62 [1.35]
		Writing	1.05 [1.83]	0.83 [1.76]	0.95 [1.78]	0.01 [1.50]	1.05 [2.30]	0.83 [2.25]	0.95 [2.24]	0.01 [1.88]	-0.65 [1.14]	-0.65 [1.40]
	7	Numeracy	0.26 [1.01]	0.45 [0.96]	0.45 [0.98]	0.41 [0.94]	0.26 [1.63]	0.45 [1.63]	0.45 [1.62]	0.41 [1.42]	1.79 [1.04] (*)	1.79 [1.56]
		Reading	-0.08 [1.10]	0.07 [1.07]	0.09 [1.08]	0.03 [1.05]	-0.08 [1.66]	0.07 [1.65]	0.09 [1.64]	0.03 [1.46]	1.08 [1.09]	1.08 [1.58]
		Writing	0.57 [1.05]	0.75 [0.98]	0.78 [1.00]	0.72 [0.97]	0.57 [1.71]	0.75 [1.72]	0.78 [1.70]	0.72 [1.50]	2.10 [1.06] (*)	2.10 [1.61]

The estimates for the average impact of both strikes on participation rates are given in Table 12. Column numbers refer to the corresponding equation. The estimates come from an extension of the equations 3 through 6 where the gender of the students are controlled by a binary variable and the impact is estimated by interacting the gender variable with the strike variables. In particular, columns numbered (3) show the estimates from the baseline difference-in-differences equation with the addition of a binary variable indicating gender, columns numbered (4) show the estimates with the control for the change in the FSA in 2008. Columns numbered (5) show the estimates with year fixed effects and columns numbered (6) show the estimates with district fixed effects. The coefficient estimates can be interpreted as the change in the participation rate.

Table 13: The Impact of Each Strike on Success Rates

			All Available Data								Balanced Set of Schools	
			Clustered				Not Clustered				Clustered	Not Clustered
		Specification	(3)	(4)	(5)	(6)	(3)	(4)	(5)	(6)	(6)	(6)
	Grade	Skill										
2005/2006	4	Numeracy	-1.23 [2.58]	-0.85 [2.50]	-0.80 [2.51]	2.02 [1.94]	-1.23 [2.28]	-0.85 [2.26]	-0.80 [2.27]	2.02 [2.19]	3.33 [1.83] (*)	3.33 [1.48] (**)
		Reading	-5.02 [2.31] (**)	-5.00 [2.29] (**)	-4.97 [2.30] (**)	-3.01 [1.77] (*)	-5.02 [2.37] (**)	-5.00 [2.37] (**)	-4.97 [2.38] (**)	-3.01 [1.93]	0.10 [1.81]	0.10 [1.61]
		Writing	-1.95 [2.02]	-1.56 [1.95]	-1.52 [1.96]	-0.15 [1.72]	-1.95 [1.55]	-1.56 [1.52]	-1.52 [1.52]	-0.15 [1.73]	0.98 [1.59]	0.98 [1.50]
	7	Numeracy	2.29 [2.55]	2.68 [2.56]	2.75 [2.56]	2.86 [2.31]	2.29 [3.50]	2.68 [3.48]	2.75 [3.50]	2.86 [2.47]	4.58 [2.06] (**)	4.58 [1.99] (**)
		Reading	-3.75 [2.36]	-3.70 [2.34]	-3.69 [2.35]	-3.27 [2.13]	-3.75 [2.65]	-3.70 [2.65]	-3.69 [2.66]	-3.27 [2.12]	1.46 [1.90]	1.46 [2.08]
		Writing	-1.75 [2.11]	-1.71 [2.11]	-1.85 [2.14]	-0.22 [1.79]	-1.75 [2.46]	-1.71 [2.46]	-1.85 [2.46]	-0.22 [1.98]	0.16 [1.95]	0.16 [2.14]
2014/2015	4	Numeracy	-3.04 [3.40]	-2.66 [3.44]	-2.61 [3.46]	-2.17 [3.13]	-3.04 [5.14]	-2.66 [5.13]	-2.61 [5.15]	-2.17 [3.71]	-7.03 [2.98] (**)	-7.03 [3.23] (**)
		Reading	-1.23 [3.69]	-1.22 [3.70]	-1.19 [3.71]	-0.87 [3.37]	-1.23 [5.02]	-1.22 [5.02]	-1.19 [5.04]	-0.87 [3.79]	-3.07 [2.88]	-3.07 [3.23]
		Writing	-3.14 [3.69]	-2.75 [3.71]	-2.72 [3.72]	-1.98 [3.44]	-3.14 [4.73]	-2.75 [4.72]	-2.72 [4.74]	-1.98 [3.67]	-4.54 [3.43]	-4.54 [3.54]
	7	Numeracy	-0.55 [3.48]	-0.16 [3.53]	-0.09 [3.55]	-1.94 [3.32]	-0.55 [5.67]	-0.16 [5.66]	-0.09 [5.69]	-1.94 [4.01]	-11.27 [3.27] (***)	-11.27 [2.94] (***)
		Reading	-0.21 [2.53]	-0.16 [2.54]	-0.14 [2.55]	-1.61 [2.50]	-0.21 [4.51]	-0.16 [4.51]	-0.14 [4.52]	-1.61 [3.21]	-7.88 [3.65] (**)	-7.88 [3.14] (**)
		Writing	6.22 [3.81]	6.26 [3.82]	6.12 [3.82]	5.32 [3.76]	6.22 [5.14]	6.26 [5.14]	6.12 [5.16]	5.32 [4.16]	-8.86 [3.56] (**)	-8.86 [3.22] (***)

The estimates for the separate impacts of the two strikes are given in Table 13. The estimates come from an extension of the Equation (3) where each strike year is denoted with a separate binary variable. Column numbers show the corresponding equations: columns numbered (4) show the estimates with the control for the change in the FSA in 2008. Columns numbered (5) show the estimates with year fixed effects and columns numbered (6) show the estimates with district fixed effects. The coefficient estimates can be interpreted as the change in the success rate - the percentage of students who meet or exceed expectations.

Table 14: The Impact of Each Strike on Participation Rates

			All Available Data								Balanced Set of Schools	
			Clustered				Not Clustered				Clustered	Not Clustered
		Specification	(3)	(4)	(5)	(6)	(3)	(4)	(5)	(6)	(6)	(6)
	Grade	Skill										
2005/2006	4	Numeracy	-0.75 [1.44]	-0.59 [1.43]	-0.56 [1.45]	-0.03 [1.45]	-0.75 [1.33]	-0.59 [1.33]	-0.56 [1.33]	-0.03 [1.47]	0.55 [1.15]	0.55 [1.10]
		Reading	-0.15 [1.26]	0.01 [1.26]	0.04 [1.27]	0.20 [1.32]	-0.15 [1.26]	0.01 [1.26]	0.04 [1.26]	0.20 [1.43]	0.74 [1.28]	0.74 [1.17]
		Writing	-0.27 [1.49]	-0.02 [1.47]	0.01 [1.49]	0.36 [1.50]	-0.27 [1.33]	-0.02 [1.32]	0.01 [1.33]	0.36 [1.51]	1.11 [1.39]	1.11 [1.23]
	7	Numeracy	0.41 [1.44]	0.58 [1.45]	0.63 [1.47]	0.44 [1.50]	0.41 [1.62]	0.58 [1.62]	0.63 [1.63]	0.44 [1.40]	3.29 [1.75] (*)	3.29 [1.36] (**)
		Reading	2.08 [1.68]	2.25 [1.67]	2.30 [1.69]	2.49 [1.68]	2.08 [1.69]	2.25 [1.69]	2.30 [1.69]	2.49 [1.67]	3.99 [1.79] (**)	3.99 [1.40] (***)
		Writing	-0.72 [1.85]	-0.49 [1.84]	-0.41 [1.83]	0.37 [1.71]	-0.72 [1.46]	-0.49 [1.44]	-0.41 [1.45]	0.37 [1.52]	4.41 [1.88] (**)	4.41 [1.43] (***)
2014/2015	4	Numeracy	-2.17 [1.47]	-2.02 [1.48]	-1.99 [1.48]	-2.13 [1.52]	-2.17 [2.21]	-2.02 [2.21]	-1.99 [2.22]	-2.13 [1.95]	-3.64 [1.48] (**)	-3.64 [1.43] (**)
		Reading	-1.60 [1.54]	-1.43 [1.55]	-1.40 [1.56]	-1.51 [1.60]	-1.60 [2.27]	-1.43 [2.27]	-1.40 [2.28]	-1.51 [2.01]	-3.29 [1.46] (**)	-3.29 [1.36] (**)
		Writing	-1.54 [2.17]	-1.29 [2.18]	-1.25 [2.19]	-1.20 [2.19]	-1.54 [3.00]	-1.29 [3.00]	-1.25 [3.01]	-1.20 [2.58]	-3.62 [1.62] (**)	-3.62 [1.48] (**)
	7	Numeracy	-2.62 [1.74]	-2.45 [1.73]	-2.40 [1.74]	-2.94 [1.81]	-2.62 [2.68]	-2.45 [2.68]	-2.40 [2.69]	-2.94 [2.15]	-3.22 [1.89] (*)	-3.22 [1.73] (*)
		Reading	-3.56 [1.46] (**)	-3.39 [1.46] (**)	-3.34 [1.45] (**)	-3.70 [1.50] (**)	-3.56 [2.47]	-3.39 [2.47]	-3.34 [2.48]	-3.70 [1.94] (*)	-4.32 [1.83] (**)	-4.32 [1.65] (***)
		Writing	-1.00 [2.38]	-0.76 [2.38]	-0.68 [2.40]	-1.45 [2.15]	-1.00 [3.52]	-0.76 [3.51]	-0.68 [3.53]	-1.45 [2.81]	-4.07 [1.79] (**)	-4.07 [1.67] (**)

The estimates for the separate impacts of the two strikes are given in Table 13. The estimates come from an extension of the Equation (3) where each strike year is denoted with a separate binary variable. Column numbers show the corresponding equations: columns numbered (4) show the estimates with the control for the change in the FSA in 2008. Columns numbered (5) show the estimates with year fixed effects and columns numbered (6) show the estimates with district fixed effects. The coefficient estimates can be interpreted as the change in the participation rate.

The decreases in participation rate are even smaller in magnitude, however, in one case, they are statistically significant at the 10% level of confidence. This suggest that there may have been an relatively small impact on participation rates.

While I do not estimate an impact on average, there are some of the results point a more nuanced picture. In particular, for Grade 4 female students, there are drops in both participation rates for all three subjects, and a drop in success rates for Reading. Furthermore, the strike in 2014/2015 may have been more impactful then the previous strike. Finally, the results are different when I concentrate on the the schools with larger sets of students, which have the full data for all years. This difference is not just a difference in statistical significance, but also in magnitude. I discuss these further in Section 14.[19]

14 Discussion

Overall, the results do not show evidence of an impact of the strikes on success rates and participation rates, except for Grade 4 female students. These results are visible in the first eight columns of all tables. However, restricting the analysis to a smaller set of schools with complete data shows a drop in Grade 4 participation rates and a drop in Grade 7 success rates as a result of the 2014/2015 strike. Furthermore, the 2005/2006 strike may have increased success rates for Numeracy, and may have increased participation rates for Grade 7 students. These results are on the rightmost two columns of all tables.

The difference in findings has two explanations: the first explanation is that there is an impact on all schools, but this impact is difficult to distinguish from additional statistical noise due to missing data. The balanced set of schools - with complete data - do not suffer from as much statistical noise, therefore, when I restrict the analysis to this balanced set, the impact is easier to distinguish. A second alternative explanation is that the strike mainly impacted larger schools which are more likely to be included in the balanced set. The actual explanation is probably a combination of both: for the balanced set of schools, it is not just the standard error that is lower - the magnitude of the estimated impact is also larger. This suggests that there

[19]In addition to these, I run the same regressions for the sample of aboriginal students. There are no statistically significant results. For reference, the results are included in the appendix.

may have been an additional impact on the schools with a larger number of students, which are likelier to be included in this set.[20]

The drops in participation and success rates associated with Grade 4 female students are visible in both the full data set and the smaller set of schools. Furthermore, it is robust to various specifications. I interpret this as the most robust finding of the study. Even if there is no impact of the strike on average, Grade 4 students were impacted.

The impact of a strike can show up on both participation rates and success rates. For instance, parents that anticipate a low performance may opt out of the exam. In the last two columns of Table 9, for the balanced set of schools, I estimate that the participation rates fell for Grade 4 students. At the same time, success rates fell for the Grade 7 students - see the last two columns of Table 10. A potential explanation may be that parents of Grade 4 students are more concerned about the stress of testing for their kids than the parents of Grade 7 students.

When I run the same analysis on the balanced set of schools, I find that the later strike, 2014/2015, caused a decrease in participation rates for both grades in all exams, and a decrease in success rates for Grade 7. In contrast, I estimate that the strike in 2005/2006 caused a statistically significant increase in Numeracy success rates for both grades, and it caused a statistically significant increase in the participation rates across all exams in Grade 7. There are some differences in the conditions surrounding the two strikes. One difference is the potential impact of the BCTF letter. Even though the BCTF was always opposed to standard examination, after the changes in 2008, they started to actively speak against it to parents. Every year since 2008, the BCTF sent a letter to the parents, clearing indicating their position as opposing to the FSA. By the 2014/2015 strike, this letter may have impacted parents' attitudes, and the parents may not have taken the FSA as seriously as in 2005/2006. As a consequence, they may have been less interested in having their children study for the exam, or even have them participate in the exam. This would then explain why the impact could be positive in 2005/2006, but negative in 2014/2015: In 2005/2006, the parents will have their child study, compensating for the strike, and in 2014/2015 this would no longer be the case.

[20] A potential explanation can be that schools with a larger number of students are likelier to be in low-income districts. If this is the case, then this may point to varying impact of strikes by socioeconomic class. Further analysis is required to understand the implications of this varying impact.

While we can estimate that participation decreased for some students, we do not have a way of knowing which students. If the students that are likely to not attend the exam are the students that would have had a lower academic performance, then our results with be biased downwards in absolute value. The largest estimate we have for participation rates is -1.7% for Grade 4 (See the last two columns on Table 10). Potentially, the estimates could be censoring an equivalent drop in success rates. In the worst case scenario, if all of the students that we are missing are coming from the students that would have "not met expectations", should they have taken the exam, then the Grade 4 results in Table 9 are biased by the same amount, and would likely have been statistically significant, if we had observed their performance. To assess the impact of the strike, it is important to observe both the participation and the success rates, as the impact may show up as an actual decline in academic performance, or as a decrease in participation driven by parents' expectations of low performance. In this sense, the results from Tables 9 and 10 may complement each other, as the drop in participation may be coming from parents who anticipate (correctly) that their children may perform worse. I was not able to find any research suggesting that parents in fact choose to opt out of the exam may be motivated by concerns about low performance. In fact, a recent survey suggests that parents who "opt out" are typically politically motivated, and very few are concerned about scores (Pizmony-Levy and Green Saraisky, 2016). However, results from the survey are related to the "opt out movement" and may not be applicable in the current setting. Furthermore, parents may be unwilling to admit in a survey that they are concerned about their children's academic performance. Further research is required to understand if anticipated low performance is a reason for parents to opt out of standard examinations.

15 Conclusion

I do not find a statistically significant impact of teachers' strikes on success rates or on participation rates in British Columbia on average. If there has been an impact, then this impact is indistinguishable from statistical noise. However, this does not mean that there are no impacts for particular sets of students: Most importantly, I estimate a clear negative impact on Grade 4 female students. Furthermore, the results suggest that different sets of schools can be impacted differently: the impact is more discernible and larger in magnitude in schools with a larger number of students, which are likelier to be included in the balanced sample. Finally, different strikes may have impacted students differently: most of the impact is driven by the 2014/2015 strike. Taken together, these paint a nuanced picture of strikes and academic performance: strikes can have different impacts on different groups of schools and on different genders, and this impact can change with the particular conditions surrounding the strike.

One of the main conclusions of the paper is that Grade 4 female students were impacted by the strikes. This impact is negative for all three subjects and shows on both participation rates and success rates, and statistically significant. The impact of participation is robust to the analysis in the smaller sample of schools (the balanced set) and the success rate drops do not change sign with this. This suggests that Grade 4 female students are impacted by the strike discriminately. Further study is required to understand why Grade 4 female students were impacted in particular.

Another conclusion is that the impact of a strike can depend on the circumstances surrounding a strike. All of the statistically significant negative results are driven by the 2014/2015 strike. The strike causes slightly more lost days of schooling compared to the 2005/2006. In addition, the political climate surrounding the legitimacy of the FSA examinations were different in two strikes: during the 2005/2006 strike, BCTF had not been voicing opposition to the standardized examination as strongly as in the 2014/2015. When the 2014/2015 strike took place, BCTF had been sending a letter clarifying their opposition to the FSA exams clearly for years, which hadn't been the case for the 2005/2006 strike. BCTF's opposition to the FSA exam may have changed attitudes over time towards the exam, and may have triggered parents to take the exam less seriously than in 2005/2006. Parents may not have been as motivated to have their kids compensate particularly for the exam in 2014/2015, as much as in 2005/2006, resulting in a larger and more discernible impact.

An alternative explanation for the difference in the impacts of the two strikes can be related to timing between the strike and the FSA examination. In the second of the two strikes, the students had less time to catch up until the FSA exam took place. This is in contrast to the first strike, where the time from the strike was large enough to allow for extra study. As such, even if the parents were equally motivated for their children to study for the exam, sufficient compensation may not have been feasible for the second of the two strikes.

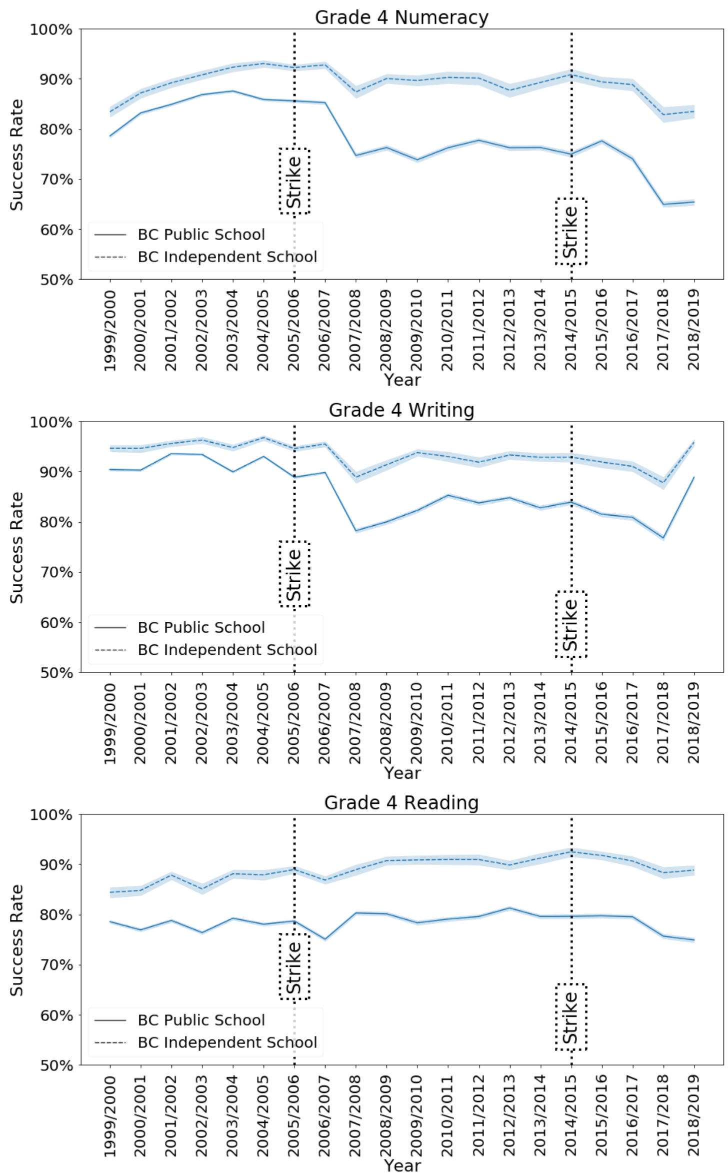
In addition to these conclusions, I make an important methodological observation. In difference-in-differences estimations, when the treatment happens with little variation across the units of observations, this can violate the i.i.d assumption in standard error estimations, and lead to over-rejection problems. There are several ways to deal with standard error estimation issues, but most researchers default to using "clustered" standard errors. However, "clustered" standard errors are not always the best tool to solve over-rejection issues (Angrist and Pischke, 2008; Abadie et al., 2010; Bester et al., 2011). The current study shows one such setting: I show in the placebo experiments that it is easy to obtain "p $<$ 5%" results as much as 40% of the time. In this case, I deal with the over-rejection problem by collapsing the unit of observations (schools) to a higher unit of observation (districts). Since previous papers in teachers' strikes literature do not discuss placebo experiments, it is possible that their results may also be suffering from over-rejection problems.[21] Simply using "clustered" standard errors is not enough, and other options, such as collapsing the data, should be used in conjunction with placebo experiments in difference-in-differences studies.

Appendix I: Figures, Student Performance

All figures show averages for all independent and all public schools in a given district. The confidence intervals show 1.97 standard deviations. The outcome variable is the success rate, i.e. the percentage of students who met or exceeded expectations in a specific exam. Dashed lines denote FSA tests taken shortly after the 2005 and 2014 strikes.

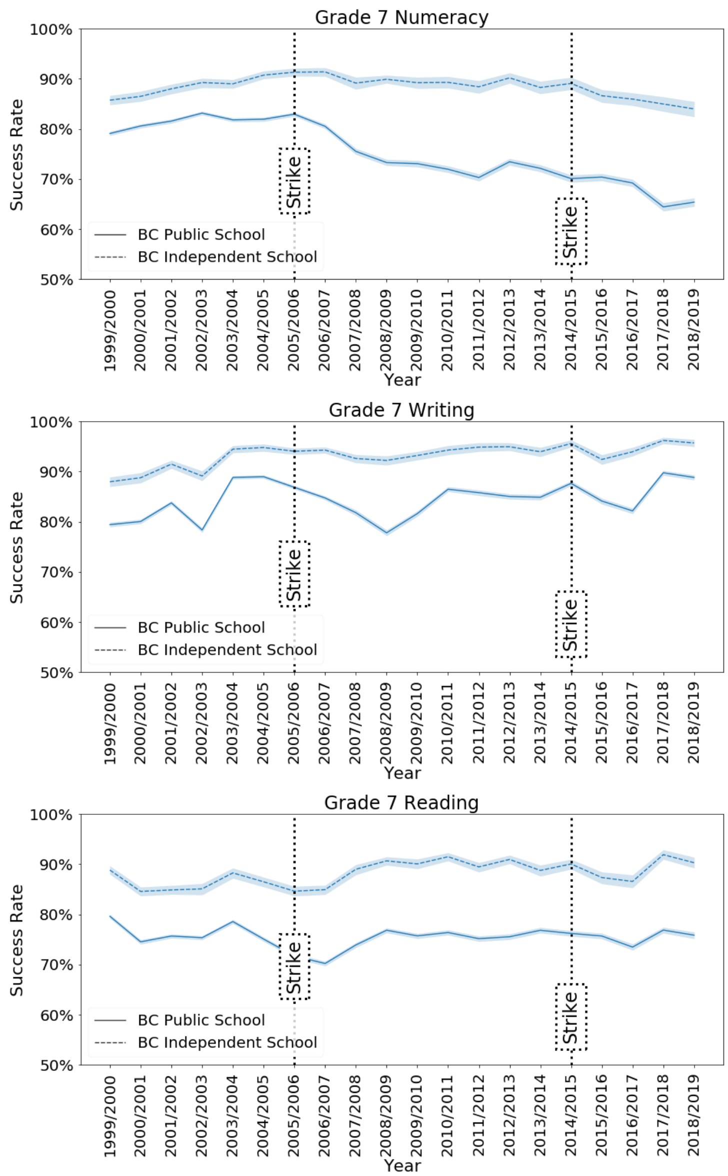
[21] For instance, in lieu of using district-fixed effects, Baker (2013) uses first differencing. However, first-differencing leads to negative serial correlation of errors (Chan et al., 1977), which in turn causes clustered standard error estimations to be lower (Cameron and Miller, 2015). It is possible that results from Baker (2013) suffer from this problem.

Figure 6: Average Success Rate for All Three Skills for Grade 4 Students



Success rate is calculated as the number of students meeting or exceeding expectations, divided by the total number of students writing the exam. The two lines represent the BC Public Schools and the BC Independent Schools. The average is calculated across schools.

Figure 7: Average Success Rate for All Three Skills for Grade 7 Students

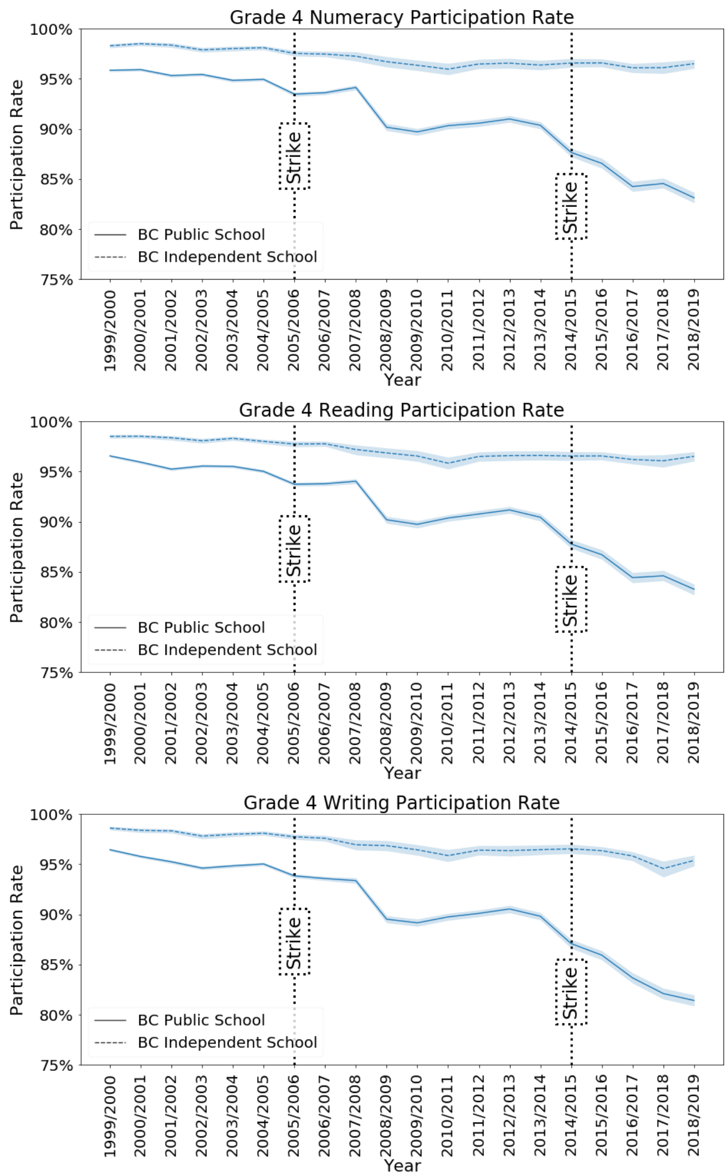


Success rate is calculated as the number of students meeting or exceeding expectations, divided by the total number of students writing the exam. The two lines represent the BC Public Schools and the BC Independent Schools. The average is calculated across schools.

Appendix II: Participation Rates

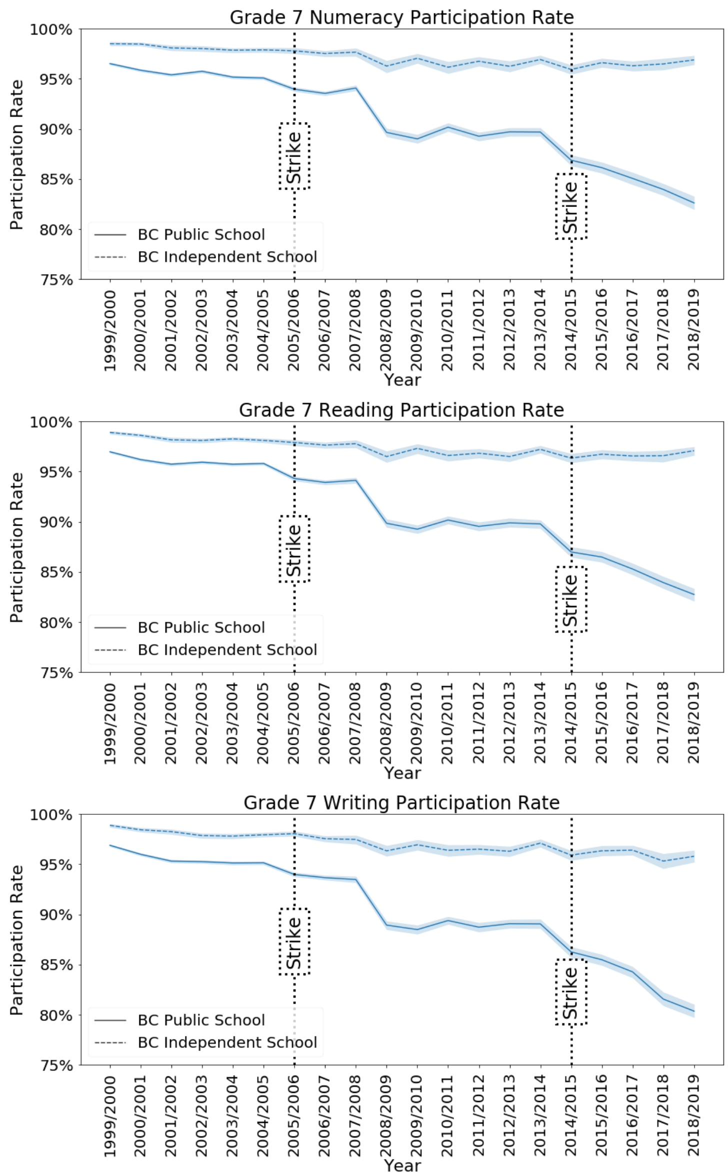
All figures show district averages for all independent and for all public schools in a given district. The confidence intervals show 1.97 standard deviations. Dashed lines denote FSA tests taken shortly after the 2005 and 2014 strikes.

Figure 8: Average Participation Rates for All Three Skills for Grade 4 Students



Participation rate is calculated as the number of students who wrote the exams divided by the number of students expected to write the exam. The average is calculated across schools.

Figure 9: Average Participation Rates for All Three Skills for Grade 7 Students



Participation rate is calculated as the number of students who wrote the exams divided by the number of students expected to write the exam. The two lines represent the BC Public Schools and the BC Independent Schools. The average is calculated across schools.

Appendix III: Placebo Estimations

Table 15 is a repetition of the placebo estimations with a balanced data set. The results are similar to those from Table 8.

Table 15: Results from Placebo Experiments, Balanced Data Set

School-level			District-level	
No Clustering	Clustering by District	2-way Clustering	No Clustering	Clustering by District
340 (42%)	372 (46%)	352 (43%)	32 (4%)	87 (11%)

Results from the placebo experiments on a balanced data set. See the Table 8 in the main body of the text for a detailed explanation.

Appendix IV: Aboriginal Students

In addition to the main results, I estimate the impact of the strikes on aboriginal students. The estimates for the success rates are in Table 16 and the estimates for participation rates are in Table 17. Most of the estimates are not statistically significant, and their sign varies between exams and grades.

Appendix V: Data Set Construction Details

I start with three data sets procured directly from the BC Government. The first data set covers the years 1999/2000 to 2006/2007. While this data set is no longer available online, it is available from the BC Ministry of Education upon request. The second and third data sets cover the years 2007/2008 to 2016/2017 and 2017/2018. to 2018/2019, respectively. As of the time of writing of this dissertation, these data sets are available for download at data.gov.bc.ca.

To merge the data sets, I use the identifier fields school year, school number, FSA skill code, grade and subpopulation (i.e. gender, English Second Language status, Aboriginal status). My two outcome variables of interest are success rates and participation rates. To obtain success rates, I use the sum of number meeting and number exceeding divided by the number of writers. For the participation rates, I divide the field number of writers by the field number of expected writers. The expressions exceeding, meeting and below in the two earlier data sets are changed to extending, on track and extending by the Ministry in the latest data set. I map these back to

Table 16: The Impact of Strikes on Success Rates of Aboriginal Students

	{}	Results, Aboriginal Students									
	Sample	Aboriginal Students									
	Balancing	All Available Data								Balanced Set of Schools	
	Clustering	By District				None				By District	None
	Specification	(3)	(4)	(5)	(6)	(3)	(4)	(5)	(6)	(6)	(6)
Grade	Skill										
4	Numeracy	-5.17 [10.73]	-10.45 [10.18]	-12.71 [10.42]	-9.50 [11.26]	-5.17 [10.72]	-10.45 [9.95]	-12.71 [10.16]	-9.50 [10.34]	-0.32 [13.69]	-0.32 [12.36]
	Reading	-7.08 [10.25]	-8.37 [10.17]	-12.85 [10.21]	-9.32 [10.55]	-7.08 [10.75]	-8.37 [10.76]	-12.85 [11.12]	-9.32 [10.85]	2.46 [11.88]	2.46 [11.34]
	Writing	-11.20 [10.39]	-15.69 [9.23] (*)	-17.58 [9.24] (*)	-13.19 [9.20]	-11.20 [9.99]	-15.69 [9.16] (*)	-17.58 [9.26] (*)	-13.19 [8.68]	-5.73 [12.48]	-5.73 [12.08]
7	Numeracy	11.94 [8.88]	10.34 [8.94]	9.09 [9.08]	6.41 [9.24]	11.94 [9.25]	10.34 [9.15]	9.09 [9.50]	6.41 [9.14]	11.88 [8.26]	11.88 [8.51]
	Reading	-2.79 [8.36]	-3.14 [8.30]	-5.50 [8.17]	-8.21 [8.62]	-2.79 [9.20]	-3.14 [9.19]	-5.50 [9.27]	-8.21 [9.30]	-6.90 [7.44]	-6.90 [8.76]
	Writing	23.72 [14.20]	21.13 [14.18]	17.07 [15.08]	16.02 [15.62]	23.72 [15.17]	21.13 [15.18]	17.07 [15.48]	16.02 [15.19]	20.65 [17.43]	20.65 [16.95]

Table 17: The Impact of Strikes on Participation Rates of Aboriginal Students

	{}	Results, Aboriginal Students' Participation									
	Sample	Aboriginal Students									
	Balancing	All Available Data								Balanced Set of Schools	
	Clustering	By District				None				By District	None
	Specification	(3)	(4)	(5)	(6)	(3)	(4)	(5)	(6)	(6)	(6)
Grade	Skill										
4	Numeracy	-1.92 [4.98]	-2.88 [4.96]	-3.01 [4.78]	-1.36 [4.65]	-1.92 [4.59]	-2.88 [4.49]	-3.01 [4.72]	-1.36 [4.52]	2.68 [6.83]	2.68 [6.82]
	Reading	2.19 [5.41]	0.82 [5.44]	1.27 [5.42]	2.63 [5.21]	2.19 [4.98]	0.82 [4.95]	1.27 [5.27]	2.63 [5.21]	6.09 [7.64]	6.09 [7.61]
	Writing	-0.52 [7.16]	-1.70 [6.90]	-2.03 [6.81]	0.47 [6.84]	-0.52 [7.71]	-1.70 [7.54]	-2.03 [7.59]	0.47 [7.08]	7.58 [11.29]	7.58 [10.72]
7	Numeracy	2.70 [5.27]	2.52 [5.20]	3.42 [5.15]	0.86 [5.11]	2.70 [5.59]	2.52 [5.67]	3.42 [5.81]	0.86 [5.88]	3.70 [4.96]	3.70 [6.39]
	Reading	-0.36 [4.34]	-1.12 [4.41]	-0.79 [4.39]	-2.39 [4.22]	-0.36 [5.90]	-1.12 [6.13]	-0.79 [6.59]	-2.39 [6.14]	-1.03 [4.22]	-1.03 [7.13]
	Writing	7.45 [9.17]	5.44 [8.90]	6.92 [8.99]	5.61 [9.48]	7.45 [9.85]	5.44 [9.83]	6.92 [10.17]	5.61 [10.18]	11.16 [11.16]	11.16 [10.52]

the older expressions, and when joined this way, I do not observe any structural breaks in the data set. Regardless, specifications 4 and 6 control for year-fixed effects. As such, if there are issues that result from a mismatch between years, these specifications will control for these issues.

In the data set, public schools are matched to their prospective districts. However, the independent schools are not. To match independent schools to the corresponding district, we use the mapping given by Fraser Institute. In the final stage of the estimation, we collapse the schools to back to the district level. This for public schools, this collapse creates a data set that is similar to the district-level data from the BC Ministry of Education. The district-level aggregation for independent schools relies on the mapping from the Fraser Institute.

There are also data irregularities. First, two independent schools are marked as belonging to a district called "DIAND", an acronym for "Department of Indian Affairs and Northern Development". I drop these two schools from the data set. Second, Len Shepherd Elementary is duplicated, so I drop the duplicate. Third, there are two dozen schools with a hyphen in their school number, result of the reorganization of these schools. I keep the first part of their school number. Finally, the Ministry of Education has reported that the figures from the year 2007/2008 suffered from data collection issues and are not reliable, so I drop this year off.

Part III
The Value of Online Scarcity Signals

Abstract

Online retailers use scarcity cues to increase sales. Many fear that these pressure tactics are meant to manipulate behavioral biases by creating a sense of urgency. At the same time, scarcity cues could also convey valuable information. We measure the value of the scarcity messages posted by Expedia to a Bayesian rational consumer. A signal reveals information on the number of seats available at the posted price. Consumers can use this information to optimally time when they purchase a ticket. The maximum increase in expected utility for a naive consumer, who does not use publicly available information, is 8 percent. For a sophisticated consumer, the increase is between 4-7 percent. Scarcity signals have a negligible impact on seller revenue and consumption.

16 Introduction

Scarcity cues and pressure tactics are widely used by online retailers to increase sales (Nagpal, 2014).[22] According to marketers and some social scientists, scarcity creates a sense of urgency, it increases desirability and gives a perceived benefit of acting quickly (Worchel et al., 1975; Lynn, 1991; Verhallen and Robben, 1994; Mullainathan and Shafir, 2013). Although many fear that sellers manipulate the psychology of consumers, scarcity messages can also deliver information that is not available otherwise. A Bayesian consumer could benefit from this information even if messages are meant to manipulate behavioral consumers subject to decision biases. This paper measures the informational value of scarcity messages in the context of air travel. Airfares can vary dramatically from day to day. Many travelers have

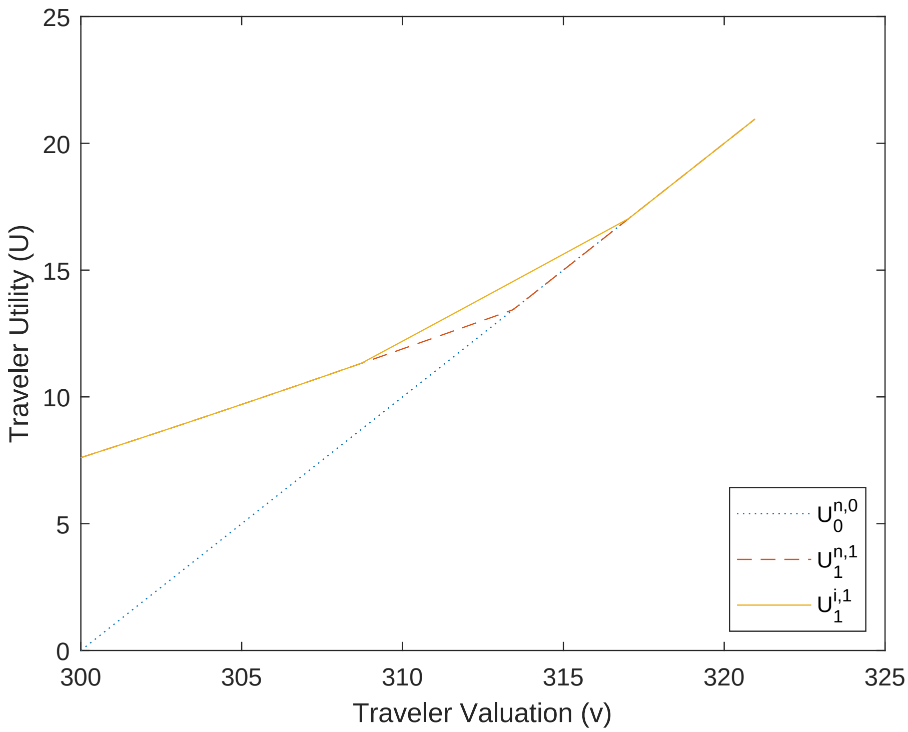
[22] For example, many Online retailers post limited-quantity messages (remaining stock left) next to selected listings. Other pressure cues include limited time offers (expiry sales) and demand interest (number of consumers browsing an item) (Aggarwal et al., 2011; Gierl and Huettl, 2010).

to choose whether to book a non-refundable ticket without knowing future fares and whether it would be wise to postpone purchase. Airlines try to influence travelers by presenting scarcity signals next to airfares. For example, fares displayed on Expedia sometimes mention that there are few seats left at the posted price.[23] We develop a Bayesian rational framework to evaluate the value of signals. This approach is supported by evidence, from similar contexts as ours, that shows that some consumers behave as rational optimizers (Li et al., 2014; Cui et al., 2016). Clearly, the rationality assumption may not apply to all travelers, but for our purpose, it is a starting point to derive empirical predictions that can be tested using only information on prices and signals. Our approach offers a relevant benchmark for the consumers who respond to messages as expected utility maximizers. Although we cannot test this assumption (because we do not observe consumer bookings), our measure delivers an upper bound for the value of signals. For our application, we collect an original dataset using a web scraping script that submits queries to the Online Travel Agent (OTA) Expedia (Edelman, 2012). As a descriptive step, we compare the distributions of price changes conditional on the signal realization; whether or not 'few seats left' is posted next to the fare. The posterior with the scarcity signal first order stochastically dominates the posterior without a scarcity signal. A scarcity signal lowers the chances that the posted fare will decrease and that it will remains constant, two pieces of information consumers care about. Showing that Expedia signals are informative is a contribution in itself and it establishes the basis for the rest of the paper. In the core of the analysis, we consider a simple one-off purchase-delay decision. The traveler can buy a ticket now or postpone her purchase decision by one week and she can do so *only once*. The only source of uncertainty is regarding what the price will be next week. This stylized scenario is consistent with recent related works (Li et al., 2014). We also present results where the consumer can delay her purchase twice and the main insights do not change. If the consumer believes that the price increases in expectation, which is often the case for airfares, her purchase decision depends on how much she values traveling relative to the current price. We obtain a threshold rule: the consumer buys at the current price if and only if her current utility, denoted $v - p_0$ where v is her valuation and

[23] The Expedia link explains "According to the data that we receive from the airline, there are very few tickets currently available at this price. While limited availability can be an indicator that he price for this flight may increase, this is not always the case."

p_0 the current price, is above a fixed positive threshold. A week later, the consumer who waited ebuys if and only if she receives a non-negative utility, that is, $v \geq p_1$, where p_1 is the updated fare.

Figure 10: Traveler's Utility as a Function of Her Valuation



Traveler's utility as a function of her valuation: (a) without waiting option (blue dots), (b) with the option of delaying purchase (red dashes), (c) with the option of delaying purchase conditional on the signal's realization (orange line). Price $p_0 = \$300$ and price p_1 is computed using the estimated distribution of weekly price change.

Figure 10 plots the traveler's utility (vertical axis) under three scenarios as a function of her valuation (horizontal axis). In the paper we typically normalize all values by the average price, but for the sake of exposition we assume here that $p_0 = \$300$, which corresponds to the average price in our sample. The figure features valuations in the range $[\$300, \$327]$ because consumers outside that range do not respond to the signal. The blue dotted line plots the consumer surplus if she cannot delay. The consumer does not care about the signal realization. Her utility is zero for valuations below \$300 (outside the range of the plot) and follows a 45 degree line for valuations above \$300. The dashed red curve plots the expected surplus if the consumer is uninformed and has the option to delay purchase. An uninformed consumer does not update her decision based on the signal's realization. The consumer with valuation 4.48% greater than p_0, that is $v = \$313$, is indifferent between buying at the current price and delaying. This consumer is located at the kink of the dashed red curve. The orange solid line plots the same surplus for the informed consumer. The two kinks in the orange curve correspond to the indifferent consumers conditional on the signal realization. Consumers with valuations far from the indifferent uninformed consumer do not benefit from the signal (the dashed red and orange lines coincide) and the increase in surplus from the signal is maximum for the indifferent consumer. Figure 10 also reveals that for most travelers the gains from waiting (the difference between the red and blue curves) is significantly higher than the gains from the signal (the difference between the blue and orange curve). An important contribution of this paper is to show that the increase in consumer surplus from the signal can be obtained by estimating non-parametrically the price distributions conditional on the signal realization. This method to compute consumer surpluses generalizes to travelers who conditions their decisions on publicly available information in addition to the signal. The signal increases the expected utility of an unsophisticated traveler, who does not condition her decision on any publicly observable information, by at most 8 percent. This corresponds to the percentage utility gain for traveler $v = \$313$ in Figure 10. We also compute the value of information for a sophisticated consumer, who uses public information in addition to the signal to predict future fares (Mantin and Gillen, 2011). For a traveler who conditions her decision on the number of days remaining till departure, the increase in expected utility is between 4 and 7 percent. Finally, we compute the impact of scarcity signals on seller revenue and on the number of tickets sold. To do so, we assume that there is a uniform distribution of consumer valuation in the neighborhood of the indifferent consumer. Scarcity signals have a small negative impact on seller revenue for some subsamples of the data and little impact on the number of tickets sold. This surprising result calls for more work to understand the supply side of signals. A natural extension would be to add non-Bayesian travelers who respond to the signal in ways that systematically benefit the supplier. This work is related to several strands of literature. The model touches upon the literatures on price discrimination with information revelation (Lewis and Sappington, 1994) and

Bayesian persuasion (Gentzkow and Kamenica, 2011).[24] The literature on scarcity signals is mixed. Scarcity theory in psychology and marketing argues that signals are largely used to exploit consumer biases (Brock, 1968; Aggarwal et al., 2011; Aguirre-Rodriguez, 2013; DellaVigna and Gentzkow, 2009; Mullainathan and Shafir, 2013). At the same time, Cui et al. (2016) offer convincing evidence that consumers respond, as assumed in this work, rationally and strategically to real-time Online information.[25] Finally, the empirical application is related to the airline literature which is reviewed in the next section. The rest of this paper is organized as follows. Section 3 presents a model of consumer decision making under price uncertainty, derives a measure of the value of information and computes the impact of the signal on seller revenues and consumption. The following section presents the data and descriptive statistics. Section 20 presents our main results and the last section concludes.

17 Online Travel Booking and Scarcity Signals

The broader context for this study is travel booking and air travel demand. There is a large literature on revenue management but the models used in empirical studies make simplistic assumptions about demand (McAfee and Te Velde, 2006; Escobari, 2012; Sweeting and Sweeney, 2015). Consumers are largely myopic: they arrive randomly and do not anticipate future fares.[26] Such simplicity, which is necessary to manage the complexity inherent with the inter-temporal trade-offs associated with the allocation of a fixed and perishable capacity, leaves the demand side largely unexplored. Demand studies of consumer bookings are rare due to the absence of publicly available data. A notable exception that had access to a proprietary booking dataset is Li et al. (2014). They report that up to 19 percent of consumers strategically delay booking based on expectations about future prices.

[24]Strategic information revelation has been studied empirically in the context of Buy/Sell recommendations by financial analysts (Stickel, 1995).

[25]Using Amazon data from a natural experiment with lightning deals, they conclude that "customers not only learn from real-time availability information, but also rationally use observable product attributes to moderate their inferences about the deal's quality. This finding supports the fundamental assumption of consumers' strategical and rational reaction to inventory information in the literature."

[26]Theoretical models account for some consumer strategizing (Dana, 1998; Deneckere and Peck, 2012).

In practice, most travelers make many decisions to book a plane ticket. They decide when to start a search, whether to search directly at the airline's branded site or at an OTA, what queries to make, what flights to look at and whether to visit a meta-search engine (e.g. Kayak) that relies on big data analytics to show a variety of price comparisons. It is widely accepted that finding a cheap fare for a given itinerary has a lot to do with timing. Delaying purchasing a ticket can be profitable, especially 3 weeks prior to departure and earlier because fare drops are not uncommon (Bilotkach and Rupp, 2011). Fare drops could be caused by slow sales (Escobari, 2012), temporary promotions, competitive pressure (Gerardi and Shapiro, 2009), or other reasons... Anecdotal evidence suggests that some consumers actively search for low fares. They compare prices across sellers, sign up for fare alerts, and experiment with diverse searches on multiple days.[27] Hopper.com reports that most customers purchase a ticket within two weeks of their initial search. Beyond these casual observations, we are not aware of systematic empirical research on how consumers search for airfares.

This paper makes a step toward understanding the benefits from delaying purchase and the role played by scarcity signals in that decision. We take a normative approach in that we leverage the power of the rationality assumption to compute the value of delaying purchase and the value of scarcity signals. With the exception of Li et al. (2014), we are not aware of any work that has looked at the option value of delay in the context of airfares.[28] This is surprising given the importance of non-refundable bookings in revenue management. Our normative benchmark should be complemented with analysis of booking data to assess the cost borne by the consumers who deviate from the rationality assumption.

To keep matters simple, we model here only the gains associated with purchasing a ticket at a lower fare. Some consumers also delay because they are uncertain about their traveling plans. As time goes by, they become more confident about their traveling needs. For these consumers, there is an additional option value of waiting. We leave this issue aside and assume that consumers have a given valuation, keeping in mind that valuation uncertainty can easily be added to the analysis. We consider simple scenarios where the consumer can delay by one week once or twice because this is realistic in our

[27] *Fare alerts* are email notices sent to subscribers when ticket prices plunges or when it is a good time to purchase a ticket.

[28] There is much theoretical research on the real option of delaying purchase with fluctuating prices Ho et al. (1998).

context. The analysis could be extended to include multiple delays.

18 A Model of Consumer Response to Scarcity Signals

A vendor sells tickets to travelers. For the moment, we focus on the decision of a single traveler with willingness to pay $v > 0$. In section 18.3 we consider a population of travelers with heterogeneous valuations in order to compute the aggregate effect of the signal. The timeline is summarized in Figure 11: the traveler can buy at the current price p_0 or wait for next period's price, $p_1 = (1 + r)p_0$, where the growth rate, $r \in [-1, \infty)$, is a random variable distributed with integrable CDF $F^n()$.[29] $F^n()$ denotes the consumer's prior without a signal. When the vendor sends a scarcity signal, the consumer uses the signal realization, which can be (b)ad or (g)ood, to update her prior. In the Expedia application, a bad signal will be interpreted as having the mention 'few seats left' posted next to the price and a bad signal as having no mention. The bad realization occurs with probability τ_b. The posteriors are integrable functions denoted $F^b(r)$ and $F^g(r)$.

Bayes rule imposes

$$F^n(r) = (1 - \tau_b)F^g(r) + \tau_b F^b(r). \tag{7}$$

For example, the good signal could imply a better distribution of growth rate (lower fares) in the sense of first order stochastic dominance, $F^g(r) \geq F^b(r)$. Note, however, that this assumption is not required in the analysis. Let Er^s, for $s \in \{n, g, b\}$ denote the mean growth rates under the prior and posteriors. The consumer is risk neutral, Bayesian optimizer and does not discount. The model addresses the following issues: (a) For which value of v does the consumer buy early? (b) How does the signal realization influence this decision? (c) What is the utility increase associated with the signal? (d) What is the change in supplier revenues and units sold?

We take the price and signal policies as given. Modeling these supply side policies is beyond the scope of this paper. What is important for this work is for these policies to be stable. Thus, a rational consumer correctly updates her prior F^n to posterior F^g or F^b depending on the signal realization. We

[29]The flight is not available the following period with probability $1 - \lim_{x\to\infty} F^s(x)$.

illustrate some of the results with two examples that deliver benchmark close form solutions:

> *Example 1:* The signal delivers information on the probability that the price remains constant. The distribution of price growth rate has a mass probability at zero and this mass is smaller under the bad realization than under the good one. The distributions of growth rates are such that $F_1^n(r) = F_1^b(r)$ for $r < 0$ and $F_1^b(r) = F_1^n(r) - (1 - \tau_b)x$ for $r \geq 0$.
>
> *Example 2:* The signal shifts the cumulative distribution function by a constant. The conditional posterior are horizontal shifts of the prior: $F_2^b(r) = F_2^n(r - x)$. The mean growth rate under the bad signal is equal to the mean in growth rate under the prior plus x.[30]

18.1 Informational Value of Scarcity Signals

The consumer's utility from buying in current period is $U_0(v) = Max(v - p_0, 0)$. Next period, the consumer purchases when $v \geq p_1$, that is, for price returns $r \leq \frac{v}{p_0} - 1$. Define the functions $\mathbf{r}(v) \triangleq \frac{v}{p_0} - 1$ and $\mathbf{v}(r) \triangleq p_0(1 + r)$ to facilitate going back and forth from valuation to equivalent return. The expected utility from waiting given belief F^s is $E(Max(v - p_1, 0)|s) = p_0 \int_{-1}^{\mathbf{r}(v)} F^s(r)dr$.[31] If the mean growth rate is non-positive, $Er^s \leq 0$, we have $E(Max(v - p_1, 0)|s) \geq v - E(p_1|s) \geq v - p_0$ for any v, and the consumer weakly prefers to wait. Otherwise, there exists a solution v^s to the indifference condition

$$E(Max(v^s - p_1, 0)|s) = v^s - p_0.$$

This is the valuation of the consumer indifferent between buying and waiting. We also call that consumer the indifferent consumer. The indifference condition is rewritten $\int_{-1}^{\mathbf{r}(v^s)} F^s(r)dr = \mathbf{r}(v^s)$. We define $\rho^s = \infty$ if $Er^s \leq 0$. Otherwise, ρ^s is the solution to

$$\rho^s = \int_{-1}^{\rho^s} F^s(r)dr \tag{8}$$

[30] Using equation (7), we obtain $F_1^g(r) = F_1^n(r) + \tau_b x$ for $r \geq 0$ and $F_2^g(r) = \frac{F_2^n(r) - \tau_b F_2^n(r-x)}{1-\tau_b}$.

[31] We have $E(Max(v - p_1, 0)|s) = vF^s(\mathbf{r}(v)) - p_0 \int_{-1}^{\mathbf{r}(v)} (1 + r)dF^s(r) = p_0 \int_{-1}^{\mathbf{r}(v)} F^s(r)dr$.

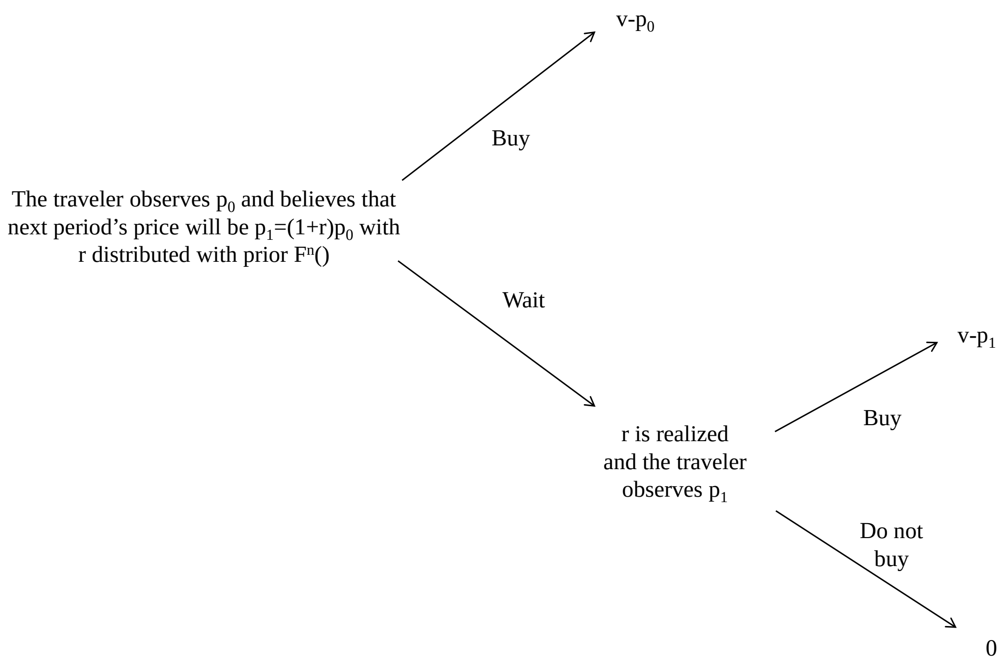


Figure 11: Decision Timeline

Decision timeline when the traveler does not receive a signal. With a signal, the consumer starts with a posterior which is conditional on the signal's realization: $F^b()$ with probability τ_b and $F^g()$ with probability $1 - \tau_b$. The price p_1 is drawn according the posterior and the decision timeline remains otherwise the same.

Lemma 1. *There exist a unique a triplet (ρ^b, ρ^n, ρ^g) such that $Min(\rho^g, \rho^b) \leq \rho^n \leq Max(\rho^g, \rho^b)$. When consumer v has belief $F^s()$, she waits if $v \in [0, \mathbf{v}(\rho^s))$ and buys early if $v \in (\mathbf{v}(\rho^s), \infty)$. Her expected utility is:*

$$U_1^s(v) = \begin{cases} p_0 \int_{-1}^{\mathbf{r}(v)} F^s(r)dr, & \textit{if } v \in [0, \mathbf{v}(\rho^s)] \\ v - p_0, & \textit{if } v \in [\mathbf{v}(\rho^s), \infty]. \end{cases} \tag{9}$$

Lemma 1 says that (ρ^b, ρ^g) lie on each side of ρ^n. A sufficient condition for $\rho^b \leq \rho^g$ is first order stochastic dominance (FOSD): $F^g(r) \geq F^b(r)$ for all r. In the rest of the analysis, we label the two states such that $\rho^b \leq \rho^g$, which is a matter of convention. Under this usage, we obtain the intuitive outcome that the bad signal triggers some consumers to change their decision from 'wait' to 'buy' and the good signal triggers some consumers to change their decision the other way around:

Proposition 1. *(a) Consumer $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$ waits without signals. With scarcity signals, she switches to buy when the signal is bad. (b) Consumer $v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]$ buys early without a signal. With scarcity signals, she switches to wait when the signal is good. (c) Consumer $v \notin [\mathbf{v}(\rho^b), \mathbf{v}(\rho^g)]$ does the same with and without a signal.*

The introduction of scarcity signals changes both the decision to wait (timing of purchase) and the decision to purchase (a consumer who waits may not buy in period one). Consumer v's expected utility when she receives a signal is

$$U_1^i(v) = \tau_b U_1^b(v) + (1-\tau_b) U_1^g(v).$$

Consumer v's gain from the signal is $\Delta U(v) = U_1^i(v) - U_1^n(v)$. We show in the Appendix that

$$\Delta U(v) = \begin{cases} \tau_b \left(v - p_0 - p_0 \int_{-1}^{\mathbf{r}(v)} F^b(r)dr \right), & \text{if } v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)] \\ (1-\tau_b) \left(p_0 \int_{-1}^{\mathbf{r}(v)} F^g(r)dr - (v - p_0) \right), & \text{if } v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)] \end{cases} \tag{10}$$

with $\Delta U(v) = 0$ for $v \notin (\mathbf{v}(\rho^b), \mathbf{v}(\rho^g))$. We say that a signal has no value if it does not improve the consumer's decision independently of her valuation v.

Corollary 1. *A signal has no value if and only if $\rho^g = \rho^n = \rho^b$.*

The signal has no value when the indifferent consumer is the same independently of the signal realization. A random signal, for example, has no value: We have $F^g() = F^b()$ and the Corollary applies. Another special case where the signal has no value to any consumer happens when $\rho^g = \rho^b = \infty$ which is equivalent to $Er^b \leq 0$ (this implies the condition in the Corollary). The consumer always waits. The signal is worthless but could still be statistically informative (for example, if $Er^b \neq Er^g$). Interestingly, a signal that helps predict availability ($\lim_\infty F^g(r) > \lim_\infty F^b(r)$) is not valuable if the condition in Corollary 1 holds. On the contrary, and somewhat counter to intuition, the signal can be valuable even when prices decrease on average ($Er^n < 0$).

Corollary 2. *The consumer with value* $\mathbf{v}(\rho^n)$ *receives the highest utility gain from the signal.*

Using identity (10), we obtain $\Delta U(\mathbf{v}(\rho^n)) = p_0\tau_b(1-\tau_b)H(\rho^n)$ where $H(r) \triangleq \int_{-1}^{r}\left(F^g(y) - F^b(y)\right)dy$. We define the value of the signal, $I \triangleq \frac{\Delta U(\mathbf{v}(\rho^n))}{U^n(\mathbf{v}(\rho^n))}$ as the relative utility change to consumer $\mathbf{v}(\rho^n)$. As explained above, only a consumer with a valuation $v \in (\mathbf{v}(\rho^b), \mathbf{v}(\rho^g))$ benefits (in expectation) from the signal. Expression I is the value of information to the indifferent consumer. It is an upper bound on the value of information across all consumers. After replacement, we have

$$I = \frac{\tau_b(1-\tau_b)}{\rho^n}H(\rho^n). \tag{11}$$

As expected, we have $I = 0$ when the condition in Corollary 1 holds ($\rho^b = \rho^g$ implies $H(\rho^n) = 0$). The value of information has the following properties: It is independent of p_0. It increases, ceteris paribus, as there is more uncertainty about the signal realization ($\tau_b(1-\tau_b)$ large), as the consumer has lower threshold (ρ^n small) and as the signal shifts the posterior further apart ($H()$ large).

Take example 1. Since $H_1(r) = rx$ for $r > 0$, the value of information simplifies to $I_1 = (1-\tau_b)\tau_b x$. The value of information is independent of the prior $F_1^n()$ and of threshold ρ^n. It increases as the signal shifts the probability of no price change by a larger amount (x large). For example 2 we have $H_2(r) = \frac{1}{1-\tau_b}\int_{r-x}^{r} F^n(y)dy \approx \frac{x}{1-\tau_b}F_2^n(r)$, where the approximation holds for x small, and $I \approx \tau_b x \frac{F_2(\rho^n)}{\rho^n}$. The value of information increases as the signal shifts the distributions of growth rate further appart (x large). It is proportional to $\tau_b x$: the consumer $\mathbf{v}(\rho^n)$ cares only about the product of the probability that the bad realization be drawn and the impact of the bad realization on the posterior distribution.[32]

[32] Note that this holds only for consumer $\mathbf{v}(\rho^n)$. Holding constant $\tau_b x$, the consumers with valuation below $\mathbf{v}(\rho^n)$ prefers a signal with low τ_b. The opposite holds for consumer $v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]$. Take the case of consumer $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$. Rewrite $\Delta U(v) = \tau_b\left(v - p_0 - p_0\int_{-1}^{\mathbf{r}(v)} F^n(r)dr\right) + \tau_b p_0 \int_{-1}^{\mathbf{r}(v)}(F^g(r) - F^n(r))\,dr$. The second term is approximated by $\tau_b p_0 x F(\mathbf{r}(v))$ which is proportional to the product $\tau_b x$. The first term, however, decreases with τ_b since $v - p_0 - p_0\int_{-1}^{\mathbf{r}(v)} F^n(r)dr < 0$ for $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$.

18.2 Supplier Revenues and Consumption

The model has only two periods and a single consumer. Within this restricted framework, one can look at changes in expected (static) consumer revenue and consumption. A bigger picture would include profits and welfare. But investigating the impact of scarcity signals on these outcomes requires modeling dynamic trade-offs that are beyond the scope of this work.[33] Denote by $\Delta R(v)$ the difference in revenue, received from the consumer with valuation v, with and without a signal. $\Delta R(v) = 0$ for $v \notin [\mathbf{v}(\rho^b), \mathbf{v}(\rho^g)]$ and

$$\Delta R(v) = \begin{cases} p_0 \tau_b \left(1 - \int_{-1}^{\mathbf{r}(v)} (1+r) dF^b(r)\right), & \text{if } v \in (\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)) \\ -p_0(1-\tau_b)\left(1 - \int_{-1}^{\mathbf{r}(v)} (1+r) dF^g(r)\right), & \text{if } v \in (\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)). \end{cases} \tag{12}$$

Take the top line in the above equation. Traveler $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$ waits without the signal. The supplier earns $p_0 \int_{-1}^{\mathbf{r}(v)} (1+r) dF^n(r)$. With a signal the traveler buys early when the realization is bad and waits otherwise. The supplier earns p_0 in the former case and $p_0 \int_{-1}^{\mathbf{r}(v)} (1+r) dF^g(r)$ in the latter one. The expected supplier revenues with a signal are $\tau_b p_0 + (1-\tau_b) p_0 \int_{-1}^{\mathbf{r}(v)} (1+r) dF^g(r)$. Taking the difference between the two revenues gives the top expression in equation (12). The revenues for $v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]$ are computed similarly. Lemma 2 in the Appendix shows that the function $\Delta R(v)$ is equal to zero up to $\mathbf{v}(\rho^b)$, at which point it jumps to a positive value, decreases up to $\mathbf{v}(\rho^n)$ where it drops to a negative value, then increases up to $\mathbf{v}(\rho^g)$ where it is still negative and where it finally jumps back to zero. The supplier loses from travelers with valuation $v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]$. She gains from travelers with valuation $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$ under a condition that holds in our application.

The signal also changes the expected probability of purchase. This is important to the seller because inventory is central to revenue management. Denote by $\Delta C(v)$ the difference in consumption by consumer with valuation v, with and without a signal.

$$\Delta C(v) = \begin{cases} \tau_b (1 - F^b(\mathbf{r}(v)), & \text{if } v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)] \\ -(1-\tau_b)(1 - F^g(\mathbf{r}(v)), & \text{if } v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]. \end{cases} \tag{13}$$

[33] In the absence of congestion and with no variable cost, change in welfare is equal to change in consumer surplus. Under dynamic revenue management, however, the supplier sells capacity to consumers who arrive continuously till the departure date, and there is an opportunity cost of capacity.

The signal changes the composition of consumers who end up traveling: A consumer with valuation $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$ is more likely to consume and the opposite holds for a consumer with valuation above $v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]$.[34]

18.3 Consumer Aggregation

Next, we look at the impact of the signal on overall revenue. We assume that v has CDF $G(v)$. Aggregating across consumers, the expected change in supplier revenues is $\Delta\overline{R} = \int_{\mathbf{v}(\rho^b)}^{\mathbf{v}(\rho^g)} \Delta R(v) dG(v)$ and consumption (quantity sold) $\Delta\overline{C} = \int_{\mathbf{v}(\rho^b)}^{\mathbf{v}(\rho^g)} C(v) dG(v)$. The sign and magnitude of $\Delta\overline{R}$ and $\Delta\overline{C}$ depend on the four primitives $(\tau_b, F^n(), F^b(), G())$. It is not possible to evaluate these expressions in the absence of information about $G()$. One can make progress, however, by looking at 'small changes' in signals.

Take the family of binary signals that are generated through a linear combination of the prior F^n and a perturbation S, as in, $F(r, x) = F^n(r) + xS(r)$. The prior's cumulative distribution is $F^n(r) = F(r, 0)$. The bad signal occurs with probability τ_b and the associated posterior cumulative distribution is $F^b(r|x) = F(r, -x(1 - \tau_b))$. Equation (7) says that the good signal's posterior is $F^g(r|x) = F(r, x\tau_b)$, where the posteriors are integrable CDFs.

A signal with $x = 0$ does not convey any information. As the value of x increases so does the weight put on S. We denote $\Delta\overline{R}(x) = R(x) - R(0)$ the changes in expected revenues associated to signal x. $\Delta\overline{C}(x)$ is similarly defined.

Proposition 2. *For small x, the signal has no first order impact on revenue and consumption ($\Delta\overline{R}'(0) = \Delta\overline{C}'(0) = 0$). Revenue increases ($\Delta\overline{R}''(0) > 0$) if and only if $\frac{f(\rho^n)}{1-F(\rho^n)} - \frac{g'(v^n)}{g(v^n)} p_0 > \frac{2}{1+\rho^n}$. Consumption increases ($\Delta\overline{C}''(0) > 0$) if and only if $\frac{f(\rho^n)}{1-F(\rho^n)} - \frac{g'(v^n)}{g(v^n)} p_0 > 0$.*

Proposition 2 is related to the Bayesian persuasion literature. Gentzkow and Kamenica (2011) consider a sender who can inform a single receiver. They characterize the optimal multi-dimensional signal that satisfies Bayes

[34] This would be important in a three (or more) periods extension because the signal influences the set of consumers who buy at p_0 or p_1 and those who remain in the pool of potential buyers in the third period. Everything else equal, the seller is better off with a pool of high valuation travelers.

plausibility (equation 7). Our application has multiple receivers. Each receiver's private type is her valuation v. The optimal signal depends on the type distribution $G(v)$ which is unknown. To get around this problem, we consider small changes in posteriors that influence only the receivers in a small neighborhood of the indifferent consumer (type v^n).[35] Proposition 2 delivers testable conditions that rest on minimum structure. When $g'(v^n) = 0$, the signal increases consumption. It also increases revenue if and only if $\frac{f(\rho^n)}{1-F(\rho^n)} > \frac{2}{1+\rho^n}$. In general, the signal is more likely to increase revenue and consumption when there are more consumers to the left of the indifferent consumer than to the right ($g' < 0$).[36] This is because the seller always looses from the consumers on the right of the indifferent consumer and benefits (at least for some) of the consumers on the left.

Note that the characteristics of the signal $(\tau_b, S())$ do not influence whether the signal is profitable or not. It influences the scale of the impact.[37] A signal is more profitable if $\tau_b = .5$ and if $S()$ puts more weight on low growth rate realizations.

19 Data and Descriptive Statistics

The Expedia signals are framed in term of the number of available seats at the posted price.[38] The model takes the signal, and posterior distributions, as given. This is reasonable since the focus here is on the consumer decision problem.

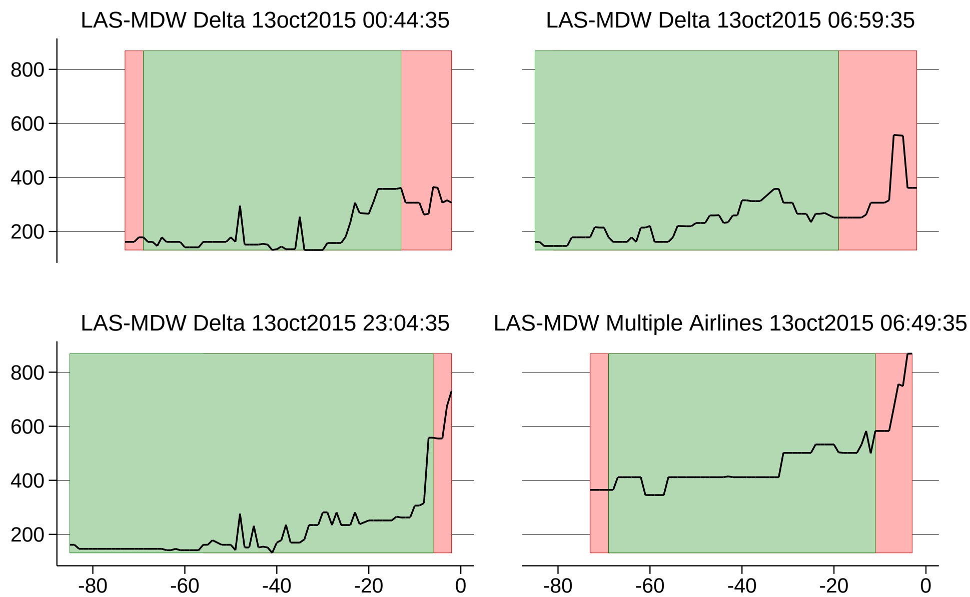
[35] We also restrict to binary signals. This is not restrictive when signals are coarse which holds in our application.

[36] The seller's revenue in the absence of signal, $R(v) = p_0(1 - G(p_0))$, is concave when $-\frac{g'(v^n)}{g(v^n)} p_0 < 2$, which does not exclude the possibility that $g'(v^n) < 0$.

[37] The Appendix, shows that $\Delta\overline{R}''(0)$ and $\Delta\overline{C}''(0)$ are proportional to $\tau_b(1 - \tau_b)\left(\int_{-1}^{\rho^n} S(r)dr\right)^2$.

[38] This is due to the way airlines revenue management systems work: A fixed number of seats is made available at a given fare and the system updates the fare when few seats remain (Lazarev, 2013). That being said, airlines and/or Expedia do not have to reveal the information about the number of seat left at the current price. They could strategically manipulate the information sent to consumers.

Figure 12: Flight Price and Signal Realization



Flight price and signal realization: An illustration. The four panels report four of the flights available for a given Expedia querie (LAS-MDW departure on October 13, 2015). Each panel reports non-availability (blank), the value of the signal (good signal is shaded green and bad red) and price (black line) as a function of DiA (horizontal axis). For each flight, DiA runs from about 80 days prior to departure (-80 on the left) till departure (0 on the right). The vertical axis reports the price in dollar.

19.1 Data Collection

We use a web-scraping script to collect data on airfares and signals. Many sellers send on a daily basis scarcity signals for a large number of travel itineraries. We select a small subset of sellers, routes, and travel dates. As with past research the sampling is constrained by the time horizon and restrictions on query processing (Edelman, 2012). We end up running daily queries for travels plans that take place at most 100 days in the future. Our dataset is similar to past studies using Internet airfares (Bilotkach and Rupp, 2011; Escobari, 2012; McAfee and Te Velde, 2006) with the shared caveat that what will be learned is sample-specific. Following Escobari (2010) and Bilotkach et al. (2010), we use Expedia which is one of the largest OTA worldwide. We conduct a number of specific searches, or *travel queries*, for one-way trips. A query comprises a route and departure date. The methodology used to collect the data is described in the appendix (Section 24). It may be that Expedia employs cookies to track visitors, and returns a personalized price for someone who returns to the site. Although our scraper accepts cookies from Expedia, it also delete the cookies after each query. Therefore, Expedia cannot identify the scraper through cookies. The price we collect is the price that Expedia would return a fresh query. The issue of supply side personalized signals is further discussed in the Appendix (Section 24). The travel queries span 10 routes (city pairs) and 22 departure dates. We selected routes with a single non-stop carrier and significant gains from purchase timing according to hopper.com. The selected routes rank low on the FAA measures of competition. For these routes, we expect sellers to have more information about future fare changes because price randomness associated with competitive dynamics is less important. Many of our routes are the same as the sample of monopoly routes used by Bilotkach and Rupp (2011). A travel query may return a large number of *flight* options. We collect the prices and signals for each option displayed. The signal is a dummy variable that is equal to one if a scarcity message is posted. We conduct travel queries each day between the 19th of July and the 26th of October, 2015.[39] We started with a given set of departure date and added new ones as existing ones expire. Denote *day-in-advance* (DiA) the number of days between the booking day and the departure day. Due to the sampling methodology, DiA is about evenly distributed between 1 to 100 days. Figure 12 presents the basic nature of the data. The figure plots the price and signal realization as a function of DiA for a selected set of flights corresponding to a given query (the panels correspond to different flights). Although the signal rarely varies from day to day for a given query, we see much variation in the signal value across flight options for the same query.

19.2 Descriptive Statistics

Table 18 presents summary statistics on the main variables. The price increases on average by 6 percent over the next 7 days. As expected, the average price growth increases with the length of the window used to com-

[39] Fare sales typically last for a few days. Daily price collection minimizes the probability of 'missing' fare sales that are only available to travelers who check fares on a daily basis.

pute changes (1 versus 7 or 14 days). There is a small chance of scarcity signal ($\tau_b = .33$). Looking at the 7 day window, the signal shifts the posteriors to 12 percent under the bad realization and to 4 percent under the good one. The top panel on Figure 13 plots the distribution of price change conditional on the signal. The signal shifts the posterior distribution by significant amounts: The good signal CDF first-order stochastically dominates the bad signal CDF. Moreover, the probability that the price stays the same (jump at $r_7 = 0$) is higher with the good signal. There is also a jump at $r_7 = \infty$ (recall that we coded an unavailable flight option as an infinite price increase). The probability of non-availability is higher with the bad signal. All the evidence point to the same conclusion that the signal is informative. That being said, the two conditional distributions have the same support. A good signal does not mean that the price will not increase and a bad signal does not mean that a large price decrease is not possible. The distributions presented on the top panel on Figures 13 are averages over all DiA. One would like to make sure that the patterns observed on this figure remain for subsamples of DiA where airfares are stable. The concern is that low DiA could be associated with more frequent bad signals and higher price growth. The bottom panel (Figure 13) reproduces the top panel but only for DiA greater than 56 days (more than 8 weeks prior to departure). The main patterns found on the top panel remain although slightly attenuated. We conclude that the signal contains information that is not solely about the changes that take place in the last few weeks before departure. Table 19 reports key quantiles of the distributions of price returns broken down by week. Recall that the signal has no value when $Er^b < 0$. This is never the case. The two distributions are ordered by FOSD with some exception (in weeks 2-7, the first decile is weakly smaller in the bad state). Thus, Figure 13 conceals heterogeneity that could be important when we compute the value of information. This is relevant because a violation of FOSD can imply negative value of information (which means that the consumer should do the opposite from what the signal advices). Figure 13 reveals that the main difference between two CDFs is the size of the jump at $r = 0$: Under the good signal there is a greater probability of no price change. Thus, Example 1 may not be a bad approximation of what the signal does to the two posteriors. Using the formula for I specific to Example 1, $I = (1 - \tau_b)\tau_b x$, we plug the values $\tau_b = .33$ from Table 18 and approximation $x = .2$ from Figure 13, to obtain the value $I = .042$. This rough approximation says that the signal increases the consumer utility by 4.2 percent.

Table 18: Signals, Prices and Availability

Variable	Obs	Mean	Std. Dev.	P25	P50	P75
Signal (1: BAD, 0: GOOD)	539506	.33	.47	0	0	1
Days in Advance	539506	47.66	26.91	24	47	71
Price	539506	299.91	160.07	165.1	274.6	409.1
Price | BAD	180071	293.58	148.93	179.2	266.6	366.2
Price | GOOD	359435	303.08	165.28	162.1	281.6	409.1
r_{t+1}	455910	.02	.19	0	0	0
$r_{t+1}\|$ BAD	141828	.04	.22	0	0	.03
$r_{t+1}\|$ GOOD	314082	0	.16	0	0	0
r_{t+7}	345129	.06	.32	-.01	0	.08
$r_{t+7}\|$ BAD	98109	.12	.36	-.01	.01	.2
$r_{t+7}\|$ GOOD	247020	.04	.29	-.01	0	.02
r_{t+14}	266187	.11	.42	-.04	0	.16
$r_{t+14}\|$ BAD	71179	.18	.48	-.04	.06	.27
$r_{t+14}\|$ GOOD	195008	.08	.39	-.04	0	.1

20 Results

Given the close connection between the model and data, the empirical analysis uses simple statistics: The empirical prior and the two posteriors ($F^n()$, $F^b()$, $F^g()$) are computed to estimate non-parametrically the values of ρ^g, ρ^n, ρ^g and I. We estimate these values using the empirical distributions and report equal-tail confidence intervals that are computed using case re-sampling bootstrapping.[40] In the computations presented in the Tables, we normalize all expressions assuming $p_0 = 1$ which implies $\mathbf{v}(r) = 1 + r$. The value of v is expressed in units of p_0. Since the average ticket price in the sample is close to \$300, we assume $p_0 = \$300$ in Figures 10, 15, 16 and 17 and when we report results in monetary terms. As Li et al. (2014), we initially assume that a traveler postpones her decision only once and by one week. For robustness sake, we subsequently consider sequential delays. We compute the gains from conditioning this decision on the Expedia signal. We initially treat all flight options returned for a query as unique products. Then, we assume that the traveler cares only about the cheapest option.

[40] We re-compute all statistics reported for each bootstrap sample. The bootstrap samples depend on the original data subsample used to compute the statistic. The data subsamples are: entire sample, decomposition by DiA, DiA28-56XCarrier, DiA28-56XRoute.

20.1 Baseline Value of Information

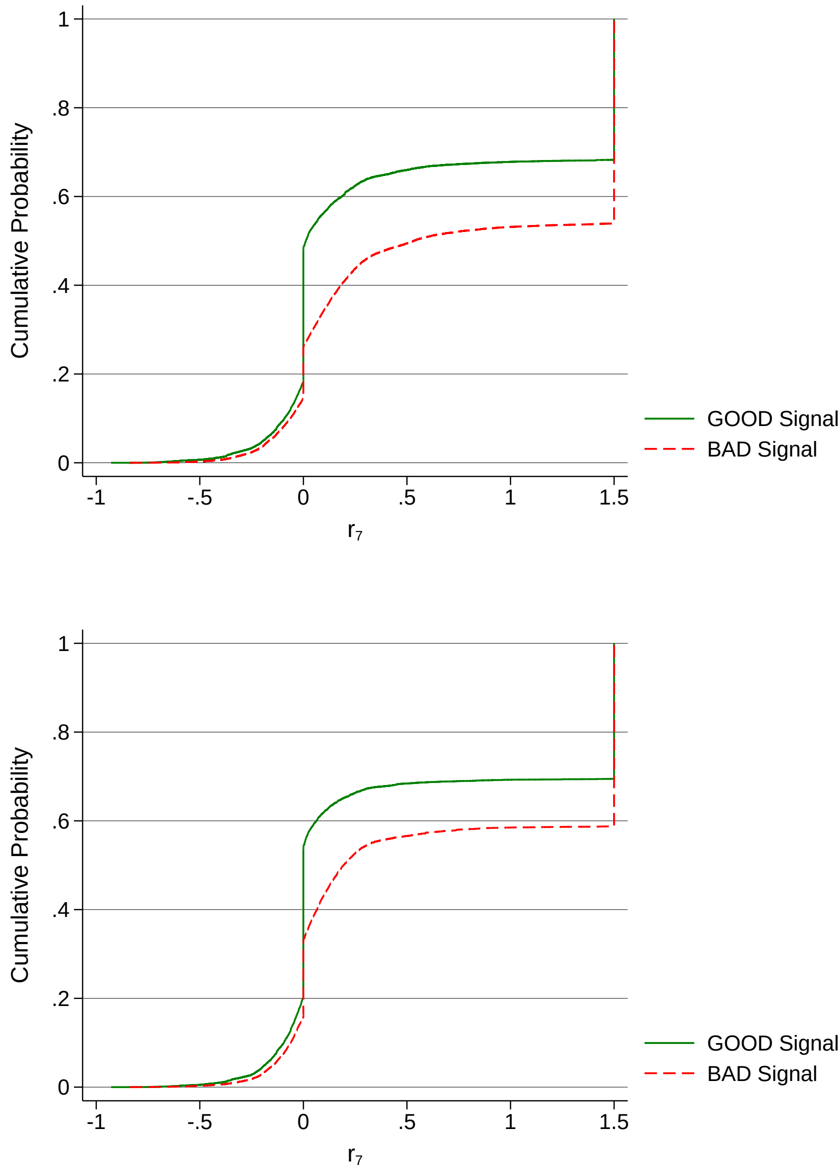
We compute $(\rho^n, \rho^b, \rho^g, I)$ for an unsophisticated consumer who does not condition her decision to buy/wait on public information. In order to eyeball ρ^s, Figure 14 plots the function $f_1(r) = \int_{-1}^{r} F^s(x)dx$ and the forty-five degree line. According to equation (8), ρ^s is found where these two intercept. Repeating this for the two posteriors (conditional on the signal realization), the top line of Table 20 reports the values of (ρ^n, ρ^b, ρ^g) for the entire sample. The consumers who value traveling between 2.9 and 4.5 percent more than the price of the ticket would wait without a signal but prefer to buy early when the signal is bad. Instead, consumers in the range 4.5 to 5.7 percent would buy early without a signal and change their decision to wait when the signal is good. Stated in dollar amounts, the consumers with a valuation in the interval $[\$308.7, \$317.1]$ respond to the signal and the remaining consumers do not. The consumers with a valuation below \$308.7 always wait and those with a valuation above \$317.1 always buy. We use equation (11) to compute $I = \frac{\tau_b(1-\tau_b)}{\rho^n} H(\rho^n)$, which is reported in column 6 of Table 20. The signal increases the utility of the consumer with valuation $\mathbf{v}(\rho^n)$ by about 8.3 percent. This is not a negligible amount. It is higher than the approximation presented in the previous Section. That approximation took into account only the change in probability that the price remain constant. The higher figure presented in Table 20 says that the signal does not only help the traveler to predict events when prices remain constant ($r = 0$) but also when prices are more likely to decrease (on Figure 13, the good posterior lays above the bad one for $r < 0$). Figure 15 plots the percentage utility change for the consumers who respond to the signal. The consumer who benefits the most correspond to the indifferent consumer used to compute the value of I. The value of information is positive, peaks at ρ^n, and has a tent shape. The average gain amongst the consumers who respond to the signal, $\frac{\Delta \overline{U}}{\rho^g - \rho^b}$, is reported in Table 20 as 0.00186 corresponding to a dollar value of \$.56. The average percentage utility increase amongst the consumers who respond to the signal is $\frac{\Delta \overline{U}}{\overline{U}} = 4\%$. The percentage utility increase is large relative to the absolute utility increase because the consumers who respond to the signal receive a small surplus in the absence of the signal. One may argue that the positive value of information we found could have just happened by chance. To make sure that this is not the case we go back to Corollary 1 to construct a *placebo test*. The Corollary says that a random signal should have no value. We draw a thousand replications of a random signal (a vector of binary draws with the same τ_b as in the Expedia sample). Corollary 1 predicts that $\rho^g = \rho^n = \rho^b$. This is indeed the case. The averages are exactly equal (.0448) with very small 95 percent confidence intervals (range at most .001). To make sure, we also compute the value of information. It is very small on average ($I = .0001\%$) and the 95 percent confidence interval over the thousand draws is also small, $[-.002, .002]$. The values of information computed from the Expedia sample (reported in Table 20) fall outside this interval.

20.2 Extensions

Consumer sophistication. The signal is correlated with public information that is also correlated with price changes. For example, Expedia scarcity signal are more common close to the departure date (low DiA) which is when prices are also more likely to increase. A sophisticated consumer, who conditions her decision on DiA, may not benefit from the signal if DiA is a sufficient statistic for the signal. Similarly, a sophisticated consumer may condition her posterior on route, airline, or other publicly observable variables. In order to investigate whether a sophisticated consumer still benefits from the information in the signal, we report the value of I after controlling for a set of conditioning variables. Table 20 reports the value of I for three subsets of DiA: less than 28 days, 28 to 56 days, and more than 56 days. The value are respectively 3.7, 7.4 and 6.7. The value of the signal is of the same magnitude and still significant. The traveler benefits most from signals sent 5 to 8 weeks in advance. The value of information is lowest within four weeks of departure. This is because it is more difficult to predict prices close to the departure date (the three distributions $F^s()$ are closer to one another). The information value of the signal also remains positive and significant after conditioning on airline or routes. The value of I varies across airlines in the range $6.3 - 11.6$. The variation across routes is in the range $4.4 - 8.5$. These differences could be because airfare are more difficult to predict in some routes, because of differences in airline policies, or because competitive conditions vary across routes.

Synthetic signal. To put the reported values of I into perspective, we conduct the following thought experiment. Take a world without signal and consider two travelers: one use DiA to condition her purchasing decision and the other doesn't. We compute how much the sophisticated traveler, who understands that the distribution of price returns depends on DiA,

Figure 13: Distribution of the Price Change



Distribution of Expedia percentage price change over 7 day period as a function of Expedia signal realization, F^n and F^b (top figure is all DiA; bottom one is $DIA \geq 56$).

gains relative to an unsophisticated one, who does not base her decision on DiA. This is like creating a *synthetic* DiA signal. Take the case where the synthetic signal conditions the distribution of returns on $DiA \leq 28$ and $DiA > 28$. The motivation is that a consumer who delays purchasing a ticket a month in advance knows that prices are more likely to increase than if she does so more than a month in advance. The utility gain from becoming sophisticated (learn about the distribution of returns conditional on DiA) is $I = \frac{\tau_b(1-\tau_b)}{\rho^n} H(\rho^n)$ where ρ^n is the indifferent unsophisticated traveler (.448), τ_b is the fraction of observations in the sample with $DiA \leq 28$ (29 percent), $F^n()$ is the distribution of price return in the entire sample, and $H()$ is computed using for $F^g()$ the distribution of price return for $DiA > 28$. In this counter-factual thought experiment, we obtain $I = 10.4\%$ which is a little more than the baseline value of the Expedia signal (8.3%). It is important to note, however, that the Expedia signal contains information that is not contained in the synthetic signal. This is because a consumer who conditions her decision on DiA still benefits from the signal (recall the values $3.7-6.7\%$ reported above). Thus, a sophisticated traveler still benefits from using the signal.

Signal definition. An Expedia signal reveals that there is a limited number, typically between 1 and 5, of seats left at the posted price (see Table 27 in the Appendix). Since the model is based on binary signals, we have defined the two states as no signal versus any number of seats left. For the sake of robustness, we discuss how the results change when we define the binary states differently: a 'bad' signal occurs only when there is exactly one seat left. For this new definition of the signal, we find that the value of information in the entire sample drops to 3.5% and varies between $2.4-3.3\%$ across the three DiA windows. The value of information is smaller when scarcity is defined more narrowly. Although consumers gain from knowing that there is only one seat left, they gain even more when they know that there is between one and five seats left.

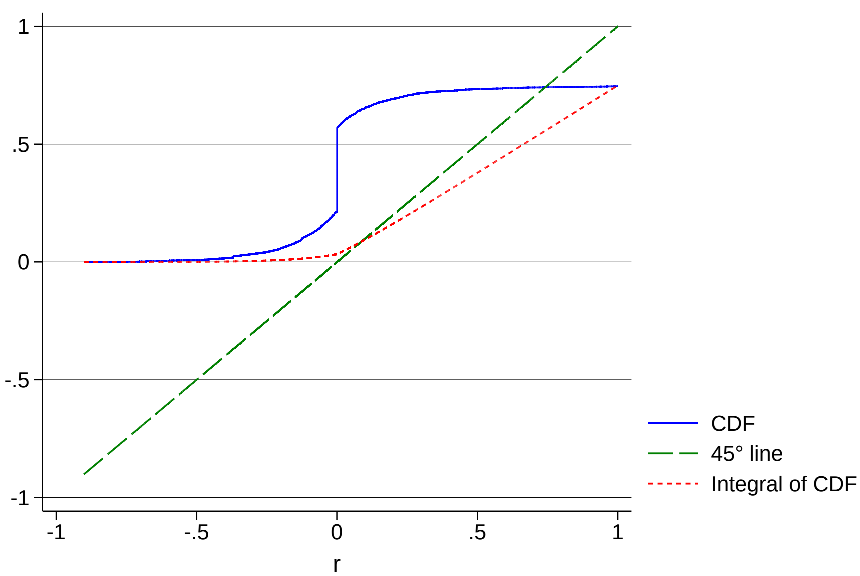
Flight substitution. A query returns on average 178 different fare options in our sample. Consumers may not care equally about the options returned for a given query. Using all flight options is reasonable if differentiation (in term of departure time or airline loyalty, for example) is important. The assumption is that each consumer is interested in a specific flight option. An alternative way to proceed, which we present now as a robustness check, is to assume that all flight options are perfect substitute. Under this scenario, all travelers care only about the cheapest option. Accordingly, we compute the value of information using the much smaller sample composed of the lowest airfare per query. The value of information is 5.9% in this new sample and varies between $3.9 - 7\%$ across the three DiA windows. We conclude that the value of information presented in Table 20 is not driven by outlier fares that are rarely purchased. Expedia signals are valuable even when we take the lowest fare per query.

Sequential delays. The analysis assumes that the traveler receives one signal and can delay purchase once by one week. In practice, the consumer can delay purchase multiple times, and learn information from multiple signal realizations. We investigate how these considerations change the results. The appendix (Section 24) extends the model to three dates to evaluate: (a) the benefit of sequential delays, and (b) the value of receiving multiple signals. This extension is for the sake of robustness and also to compare the value of the signals with the value from delaying consumption. Figure 17 replicates Figures 10 when the consumer can sequentially delay twice. To be consistent with Figure 10, we use the entire sample of observations for which we have three consecutive prices. Figure 17 shows that the key findings from the one delay case carry through to two delays: (a) the gains from information relative to no information (difference between orange and dash red curves) is small relative to the gains from sequential delays relative to no delay (difference between dash red curve and blue line), (b) the indifferent consumer benefits the most from the signal, (c) consumers with high and low surpluses do not benefit or benefit little from the signal. Some differences between the two Figures are worth mentioning. Consumers with low initial surplus now still benefit from information. To explain, take the consumer $v = p_0$. This consumer initially waits independently of the signal realization. But it is possible for the price to decrease by an amount such that this consumer becomes the indifferent consumer in the intermediate date $\frac{v-p_1}{p_1} = \rho^n$. For her second decision, the consumer now cares about the second signal. By extension, any consumer values information in the intermediate period for a small interval of intermediate prices. Obviously these intermediate prices are unlikely to happen when the consumer has a low initial surplus. This explains why the value of information decreases with the relative consumer surplus. Another difference is that the consumer surplus without information has two kinks. The new kink is due to the fact that the price does not change

($r = 0$) with a mass probability.[41]

Risk Aversion. Assume the traveler is risk averse with utility function $u(x)$ which is increasing and concave. For ease of comparison with the risk neutral case, we assume $u(0) = 0$ and $u'(0) = 1$. The traveler's current utility is $U_0^{n,0}(v) = u(Max(v - p_0, 0))$ and her expected utility from waiting given belief F^s (computed in Section 18) is now $U_1^{s,1}(v) = \int_{-1}^{\mathbf{r}(v)} u(v - p_0(1+r))dF^s(r)$. Risk aversion decreases both the utility from purchasing right away and the expected utility from waiting. There exist a unique threshold ρ^s (Lemma 1) if u'' is constant. (See Appendix 24). For the sake of conciseness, we reproduce Figure 10 for the CARA utility case for a risk averse traveler with CARA coefficient .05. See Figure 18. The three curves $U_0^{n,0}(v)$, $U_1^{n,1}(v)$ and $U_1^{i,1}(v)$ shift down but the general patterns observed in Figure 10 remain.

Figure 14: Computation of ρ^n



Computation of ρ^n: Functions $f_1(r) = \int_{-1}^{r} F^n(x)dx$ and $f_0(r) = r$ are computed using entire sample. ρ^n is such that $f_1(\rho) = f_0(\rho)$.

[41] Consumers with high relative surplus always buy early. This explains the second kink. The first kind happens because the first period growth rate distribution $F()$ has a unique probability mass at $r = 0$. With a fixed probability, the consumer surplus has the shape of Figure 10. The consumer surplus for the other draws of the first period growth rate also have a kink but these kinks are smoothed out.

Table 19: r_{t+7} by Week in Advance

Variable	Obs	Mean	Std. Dev.	P10	P25	P50	P75	P90
				Entire Sample				
r_{t+7}	357894	.06	.31	-.15	-.01	0	.08	.29
$r_{t+7}\|B$	101152	.12	.36	-.15	-.01	.01	.2	.47
$r_{t+7}\|G$	256742	.04	.29	-.15	-.01	0	.02	.23
				Week 1				
r_{t+7}	11474	.3	.56	-.13	0	.2	.49	.84
$r_{t+7}\|B$	4642	.36	.66	-.21	0	.21	.56	1.03
$r_{t+7}\|G$	6832	.25	.47	-.04	0	.19	.36	.68
				Week 2				
r_{t+7}	27221	.24	.49	-.13	0	.14	.36	.68
$r_{t+7}\|B$	11286	.26	.49	-.17	0	.16	.43	.77
$r_{t+7}\|G$	15935	.23	.49	-.05	0	.13	.32	.61
				Week 3				
r_{t+7}	29653	.15	.41	-.15	0	.02	.23	.52
$r_{t+7}\|B$	11397	.18	.39	-.17	0	.1	.29	.57
$r_{t+7}\|G$	18256	.13	.43	-.13	0	0	.18	.44
				Week 4				
r_{t+7}	30376	.08	.35	-.18	-.02	0	.12	.35
$r_{t+7}\|B$	10601	.12	.35	-.18	-.03	.03	.22	.47
$r_{t+7}\|G$	19775	.06	.35	-.17	-.01	0	.04	.27
				Week 5				
r_{t+7}	30084	.03	.29	-.19	-.04	0	.05	.24
$r_{t+7}\|B$	9393	.08	.32	-.17	-.05	0	.16	.35
$r_{t+7}\|G$	20691	.01	.27	-.2	-.04	0	0	.17
				Week 6				
r_{t+7}	29740	.03	.25	-.15	-.02	0	.04	.22
$r_{t+7}\|B$	8956	.07	.26	-.15	-.03	0	.14	.31
$r_{t+7}\|G$	20784	.01	.24	-.15	-.02	0	0	.15
				Week 7				
r_{t+7}	30133	.02	.22	-.14	-.02	0	.03	.2
$r_{t+7}\|B$	8044	.06	.24	-.15	-.02	0	.13	.28
$r_{t+7}\|G$	22089	.01	.21	-.14	-.02	0	0	.14
				Week 8				
r_{t+7}	30384	.02	.22	-.14	-.01	0	.03	.19
$r_{t+7}\|B$	7368	.06	.23	-.12	-.01	0	.12	.26
$r_{t+7}\|G$	23016	.01	.22	-.14	-.02	0	0	.14
				Week 9				
r_{t+7}	29557	.02	.23	-.14	-.03	0	.01	.18
$r_{t+7}\|B$	6813	.06	.27	-.13	-.02	0	.11	.26
$r_{t+7}\|G$	22744	0	.21	-.15	-.03	0	0	.13
				Week 10				
r_{t+7}	26336	.03	.24	-.14	-.02	0	.01	.17
$r_{t+7}\|B$	5555	.07	.29	-.12	0	0	.12	.28
$r_{t+7}\|G$	20781	.01	.23	-.14	-.02	0	0	.13

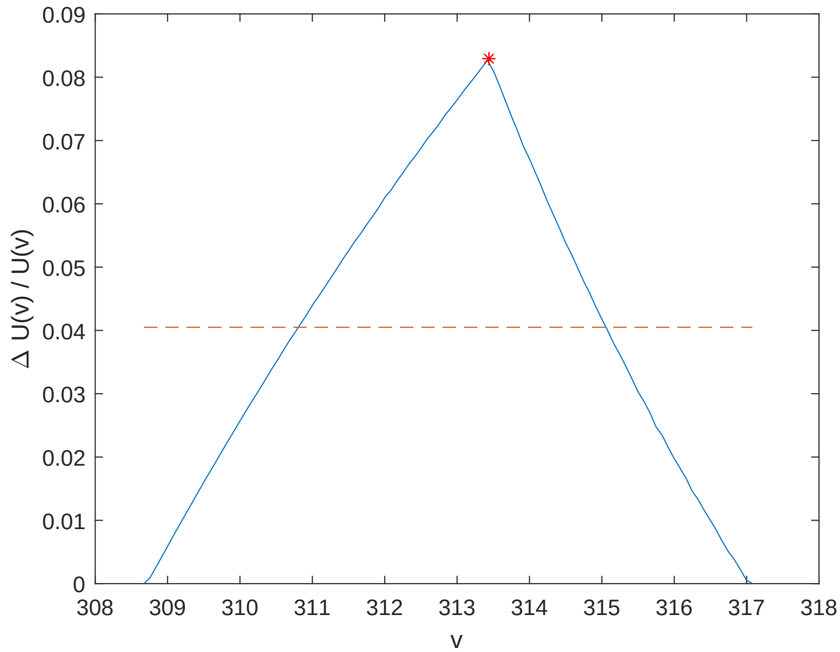
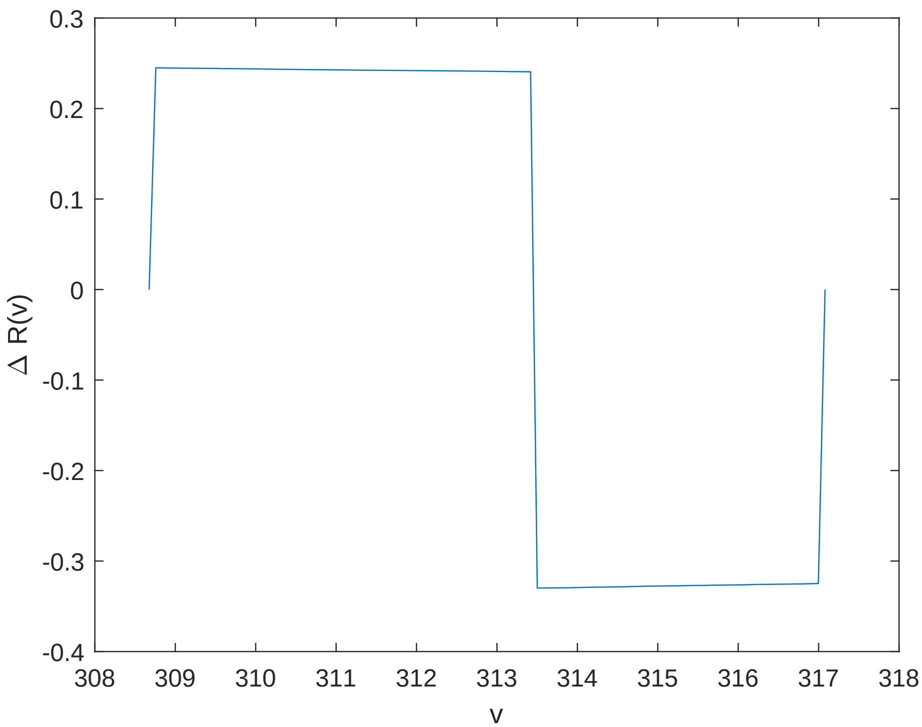


Figure 15: Utility Increase Due to Signal

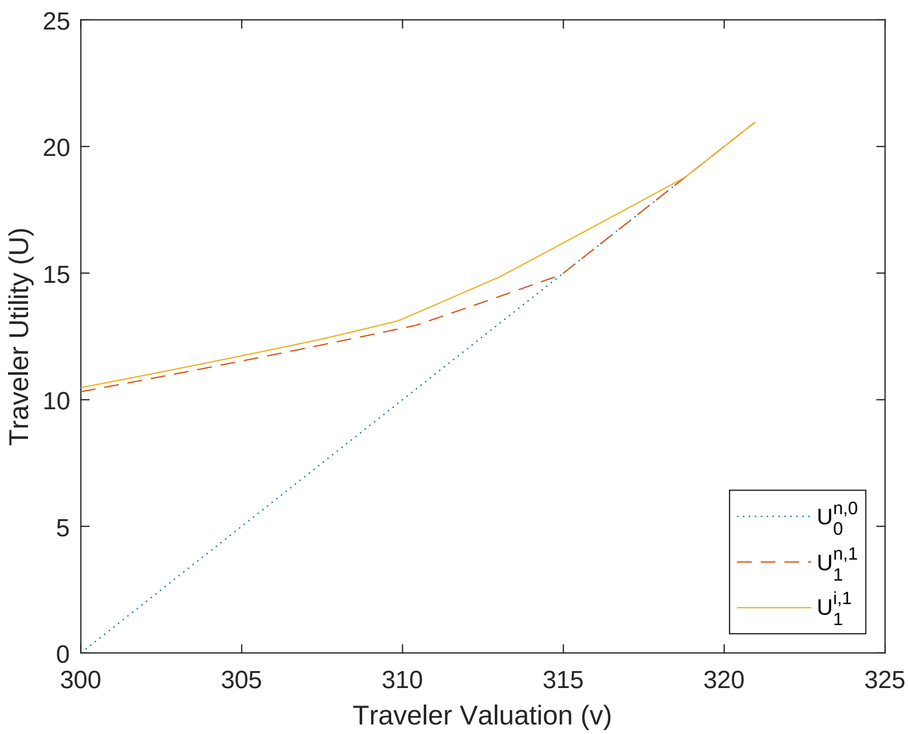
Solid blue line shows percentage utility increase $\frac{\Delta U(\mathbf{v}(r))}{U(\mathbf{v}(r))}$. The red dashed line plots the average across all $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^g)]$. The red star presents I, the maximum percentage utility increase.

Figure 16: Change in Revenue $R(v)$



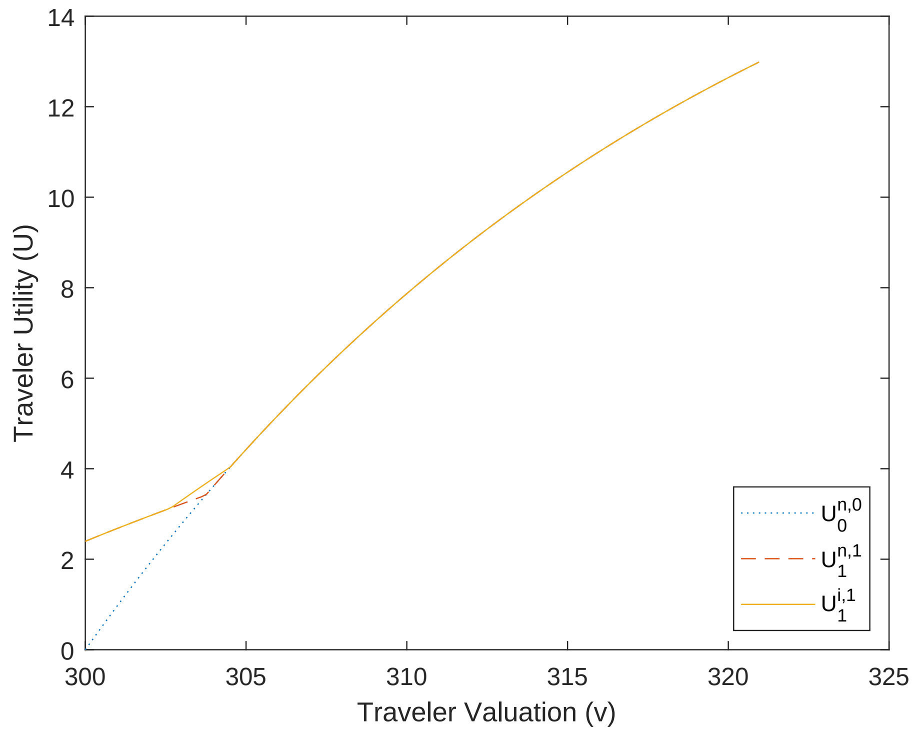
Change in Revenue $R(v)$, measured as a fraction of p_0

Figure 17: Two-Stage Extension



Expected utilities (a) without waiting option (blue dashes), (b) with the option of delaying purchase twice (red dashes), (c) with the option of delaying purchase twice and two signals (orange line).

Figure 18: Extension: CARA Utility Function



The traveler has CARA utility function $u(x) = \frac{1-exp(-CARAx)}{CARA}$ with coefficient $CARA = .05$. Expected utilities (a) without waiting option (blue dashes), (b) with the option of delaying purchase twice (red dashes), (c) with the option of delaying purchase conditional on the signal's realization (orange line).

Table 20: Main Results - Utility

Utility	ρ^u	ρ^b	ρ^g	τ_b	I	$\frac{\Delta \overline{U}}{\rho^g - \rho^b}$	$\frac{\Delta \overline{U}}{\overline{U}}$	N
All	0.0448	0.0291	0.0568	0.334	0.0829	0.00186	0.0405	539506
	(0.0437, 0.0458)	(0.0281, 0.0299)	(0.0553, 0.0583)		(0.0795, 0.0877)	(0.00177, 0.00198)	(0.0389, 0.0427)	
$DiA \leq 28$	0.0234	0.0212	0.0257	0.469	0.0367	0.000436	0.0181	153842
	(0.0224, 0.0243)	(0.0202, 0.0222)	(0.0243, 0.0272)		(0.0243, 0.049)	(0.000298, 0.000589)	(0.0124, 0.0242)	
$28 \leq DiA \leq 56$	0.0672	0.0418	0.0875	0.327	0.0743	0.00249	0.0364	165325
	(0.0649, 0.0697)	(0.0396, 0.0438)	(0.0837, 0.0911)		(0.0694, 0.0793)	(0.00228, 0.00268)	(0.034, 0.0388)	
$56 \leq DiA \leq 84$	0.0569	0.0316	0.0702	0.25	0.0675	0.00192	0.0348	155047
	(0.0547, 0.0587)	(0.0295, 0.0338)	(0.0672, 0.0729)		(0.0623, 0.0736)	(0.00175, 0.0021)	(0.032, 0.038)	
Delta	0.0607	0.04	0.0793	0.362	0.0824	0.00249	0.0395	157598
	(0.0589, 0.0622)	(0.0382, 0.0416)	(0.076, 0.0824)		(0.0755, 0.0882)	(0.00228, 0.00272)	(0.0366, 0.0423)	
American Airlines	0.0544	0.0306	0.0773	0.371	0.116	0.00317	0.0548	112925
	(0.0515, 0.0573)	(0.0284, 0.0326)	(0.0726, 0.0821)		(0.107, 0.125)	(0.00286, 0.00347)	(0.0514, 0.0588)	
Multiple Airlines	0.0208	0.0156	0.0243	0.339	0.0632	0.000658	0.0311	149417
	(0.0201, 0.0215)	(0.0147, 0.0164)	(0.0233, 0.0254)		(0.054, 0.0729)	(0.000559, 0.000763)	(0.0265, 0.0358)	
US Airways	0.0509	0.0357	0.0613	0.318	0.0681	0.00171	0.0332	48562
	(0.0467, 0.0552)	(0.032, 0.0394)	(0.0553, 0.0665)		(0.0555, 0.0816)	(0.00133, 0.00208)	(0.0273, 0.0398)	
United	0.0639	0.0253	0.0853	0.258	0.107	0.00342	0.055	39220
	(0.0601, 0.0678)	(0.0225, 0.0286)	(0.0795, 0.0916)		(0.096, 0.114)	(0.00302, 0.00379)	(0.0502, 0.0594)	
MIA-BOS, $28 \leq DiA \leq 56$	0.0663	0.0475	0.0873	0.434	0.0778	0.00251	0.0357	21792
	(0.0611, 0.072)	(0.0425, 0.0525)	(0.0774, 0.0966)		(0.0624, 0.0944)	(0.00193, 0.00315)	(0.0292, 0.0428)	
MIA-DFW, $28 \leq DiA \leq 56$	0.104	0.0608	0.137	0.345	0.0842	0.00428	0.0406	20522
	(0.0952, 0.113)	(0.0516, 0.0684)	(0.123, 0.15)		(0.0682, 0.0982)	(0.00339, 0.00506)	(0.0336, 0.0474)	
SNA-DFW, $28 \leq DiA \leq 56$	0.0813	0.0289	0.111	0.199	0.0854	0.00331	0.0432	15595
	(0.0693, 0.0919)	(0.0226, 0.0336)	(0.0936, 0.127)		(0.0753, 0.0974)	(0.00273, 0.00402)	(0.0383, 0.0498)	
PHX-MDW, $28 \leq DiA \leq 56$	0.0506	0.0403	0.0585	0.362	0.0436	0.00108	0.021	13451
	(0.0449, 0.0547)	(0.0337, 0.0461)	(0.0519, 0.0641)		(0.0235, 0.0618)	(0.000547, 0.00154)	(0.0111, 0.03)	

Table 21: Main Results - Revenue

Revenue	$\Delta R(v(\rho^b))$	$\Delta R(v(\rho^{u+}))$	$\frac{f(\rho^u)}{1-F(\rho^u)} - \frac{2}{1+\rho^u}$	$\frac{\Delta \overline{R}}{\rho^g - \rho^b}$	$\frac{\Delta \overline{R}}{\overline{R}}$	N
All	0.245	-0.33	-1.91	-0.0024	-0.0035	539506
	(0.244, 0.246)	(-0.333, -0.327)	(-1.92, -1.91)	(-0.0037, 0.004)	(-0.0053, 0.0057)	
$DiA \leq 28$	0.403	-0.393	-1.95	-0.0016	-0.0026	153842
	(0.401, 0.405)	(-0.398, -0.39)	(-1.96, -1.95)	(-0.0075, 0.0074)	(-0.012, 0.012)	
$28 \leq DiA \leq 56$	0.209	-0.272	-1.87	-0.0075	-0.0097	165325
	(0.206, 0.211)	(-0.277, -0.268)	(-1.88, -1.87)	(-0.014, -0.0062)	(-0.018, -0.008)	
$56 \leq DiA \leq 84$	0.16	-0.311	-1.89	-0.0062	-0.0088	155047
	(0.157, 0.162)	(-0.316, -0.305)	(-1.9, -1.89)	(-0.012, -0.0046)	(-0.017, -0.0065)	
Delta	0.255	-0.295	-1.89	-0.0028	-0.0038	157598
	(0.253, 0.258)	(-0.299, -0.29)	(-1.89, -1.88)	(-0.0071, 0.0052)	(-0.0096, 0.0069)	
American Airlines	0.278	-0.303	-1.9	-0.0068	-0.0093	112925
	(0.275, 0.282)	(-0.311, -0.296)	(-1.9, -1.89)	(-0.015, 0.0012)	(-0.021, 0.0016)	
Multiple Airlines	0.258	-0.389	-1.96	-0.00073	-0.0012	149417
	(0.256, 0.261)	(-0.395, -0.384)	(-1.96, -1.96)	(-0.0078, 0.0069)	(-0.013, 0.011)	
US Airways	0.238	-0.358	-1.9	-0.0069	-0.01	48562
	(0.233, 0.242)	(-0.37, -0.347)	(-1.91, -1.9)	(-0.02, 0.0008)	(-0.029, 0.0011)	
United	0.186	-0.355	-1.88	-0.0091	-0.013	39220
	(0.181, 0.19)	(-0.364, -0.346)	(-1.89, -1.87)	(-0.02, -0.002)	(-0.029, -0.0032)	
MIA-BOS, $28 \leq DiA \leq 56$	0.29	-0.271	-1.88	-0.0074	-0.0096	21792
	(0.283, 0.299)	(-0.281, -0.26)	(-1.88, -1.87)	(-0.019, 0.0033)	(-0.025, 0.0042)	
MIA-DFW, $28 \leq DiA \leq 56$	0.222	-0.311	-1.81	-0.012	-0.016	20522
	(0.215, 0.229)	(-0.322, -0.301)	(-1.83, -1.8)	(-0.022, -0.0025)	(-0.03, -0.0039)	
SNA-DFW, $28 \leq DiA \leq 56$	0.139	-0.261	-1.85	-0.0064	-0.0083	15595
	(0.133, 0.144)	(-0.274, -0.247)	(-1.87, -1.83)	(-0.016, 0.0036)	(-0.021, 0.0045)	
PHX-MDW, $28 \leq DiA \leq 56$	0.225	-0.305	-1.9	-0.0035	-0.0048	13451
	(0.214, 0.235)	(-0.318, -0.291)	(-1.91, -1.9)	(-0.013, 0.01)	(-0.018, 0.014)	

Table 22: Main Results - Consumption

Consumption	$\Delta\overline{C}$	$\frac{\Delta\overline{C}}{\rho^g-\rho^b}$	$\overline{C}$	$\frac{\Delta\overline{C}}{\overline{C}}$	N
All	-2.6e-06	-9.5e-05	0.019	-0.00014	539506
	(-8.1e-05, 8.9e-05)	(-0.0029, 0.0032)	(0.018, 0.02)	(-0.0043, 0.0046)	
$DiA \leq 28$	2.8e-05	0.0063	0.0028	0.01	153842
	(1.1e-05, 0.00011)	(0.0022, 0.023)	(0.0019, 0.0038)	(0.0034, 0.038)	
$28 \leq DiA \leq 56$	-5.4e-06	-0.00012	0.034	-0.00016	165325
	(-0.00016, 0.00016)	(-0.0035, 0.0036)	(0.031, 0.037)	(-0.0047, 0.0048)	
$56 \leq DiA \leq 84$	-3.3e-05	-0.00086	0.027	-0.0012	155047
	(-0.00018, 0.00012)	(-0.0048, 0.0033)	(0.025, 0.03)	(-0.0067, 0.0047)	
Delta	2e-05	0.00051	0.028	0.0007	157598
	(-0.00014, 0.00031)	(-0.0039, 0.0077)	(0.026, 0.031)	(-0.0053, 0.011)	
American Airlines	5.8e-05	0.0012	0.033	0.0017	112925
	(-0.00025, 0.00054)	(-0.005, 0.012)	(0.03, 0.037)	(-0.007, 0.016)	
Multiple Airlines	-2.3e-07	-2.6e-05	0.0054	-4.2e-05	149417
	(-5.4e-05, 5.5e-05)	(-0.0059, 0.0062)	(0.0046, 0.0062)	(-0.0096, 0.0099)	
US Airways	0.00012	0.0045	0.017	0.0069	48562
	(-0.00019, 0.00078)	(-0.0055, 0.03)	(0.012, 0.02)	(-0.0083, 0.046)	
United	0.0003	0.0049	0.04	0.0074	39220
	(-4.9e-05, 0.0011)	(-0.00087, 0.018)	(0.035, 0.045)	(-0.0015, 0.027)	
MIA-BOS, $28 \leq DiA \leq 56$	-0.00025	-0.0063	0.03	-0.0083	21792
	(-0.0015, 0.00027)	(-0.038, 0.0047)	(0.021, 0.038)	(-0.051, 0.0061)	
MIA-DFW, $28 \leq DiA \leq 56$	0.00026	0.0034	0.055	0.0047	20522
	(-0.00034, 0.0023)	(-0.0054, 0.03)	(0.043, 0.065)	(-0.0075, 0.041)	
SNA-DFW, $28 \leq DiA \leq 56$	-0.00043	-0.0052	0.062	-0.0069	15595
	(-0.0024, 0.00097)	(-0.028, 0.012)	(0.048, 0.076)	(-0.038, 0.016)	
PHX-MDW, $28 \leq DiA \leq 56$	-4e-05	-0.0022	0.013	-0.0031	13451
	(-0.00072, 0.00036)	(-0.036, 0.025)	(0.0067, 0.018)	(-0.052, 0.035)	

20.3 Revenue and Consumption

The impact of the signal on revenue and consumption depends on the consumer's willingness to pay. To illustrate, Figure 16 plots the change in revenues, $\Delta R(v)$ from equation 12, computed using the entire sample. (a) Revenue increases from ρ^b till ρ^n ($\Delta R(v) > 0$) and decreases from ρ^n till ρ^g ($\Delta R(v) < 0$), with a large drop at ρ^n. (b) Although the signal has a large impact the revenue per individual type, this impact largely cancels out once averaged across all consumers. These predictions are consistent with the theoretical properties of the revenue function (see Lemma 2 in Appendix). To document the overall impact of the signal on revenue, we report the arithmetic average of revenue and consumption across all consumers (assuming a uniform distribution of valuation $g(v) = 1$).[42] The main findings for the change in revenue are reported in Table 21:

1. The change in revenue for individual travelers is large and significant. Column 1, for example, reports for consumer $\mathbf{v}(\rho^b)$ a positive increase in revenue, $\Delta R(\mathbf{v}(\rho^b)) = .245$, corresponding to a \$73.5 increase for a \$300 ticket. The largest decrease in revenue occurs for consumer $\mathbf{v}(\rho^{n+})$ with $\Delta R(\mathbf{v}(\rho^{n+})) = -.33$ (column 2).

2. Column 5 reports the change in per-consumer revenue (averaged across all consumers who respond to the signal and measured as a fraction of p_0), $\frac{\Delta \overline{R}}{\rho^g - \rho^b}$, while column 6 reports the percentage increase in revenues for the consumers who respond, $\frac{\Delta \overline{R}}{R}$. The numbers are small and non-significant for most subsamples. In four subsamples, the numbers are negative and significant although the effect is small (a decrease of about 3\$ per consumer corresponding to a percentage decrease in revenue of 1%).

3. Column 4 reports the value of $\frac{f(\rho^n)}{1-F(\rho^n)} - \frac{2}{1+\rho^n}$. According to Proposition 2, profits should decrease when this expression is negative.[43] The value is indeed negative in all but two cases. As expected, the subsamples for which the changes in revenue in column 5 and 6 are statistically significant are entirely consistent with the prediction reported in column 4.

[42]The formula for this expression is derived in the appendix.

[43]This expression has to be greater than $\frac{g'(v^n)}{g(v^n)} p_0$ for the use of signal to be profitable. This term, however, cancels under the uniform assumption ($g'(v) = 0$).

Overall, the results suggest that the signal has a small and negative impact on revenue (for some subsamples). This may appear surprising until one acknowledges that revenue is not everything that matters to the seller. To start, the signal also changes consumption. A lower consumption is good for the seller because it increases remaining inventory. Table 22, however, shows that the signal has little impact on consumption:

1. The average per consumer change in consumption, $\frac{\Delta \overline{C}}{\rho^g - \rho^b}$, is reported in column 3. It is small in magnitude and significant for only two subsamples. It is small because the mass of consumers who postpone purchase under a good signal is very close to the mass of consumer who anticipate purchase under a bad signal.

2. The percentage change in consumption, $\frac{\Delta \overline{C}}{C}$, is reported in column 4. It is very small in magnitude and insignificant in all but one subsample.

A final consideration, in addition to revenue and consumption, is that the signal also changes how consumer sort between those who purchase early and those postpone. The consumers who change from 'buy' to 'wait' have a higher valuation than those who change their decisions the other way around (valuations $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$ versus $[\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]$ respectively). With a scarcity signal, a more attractive set of consumers remains potential buyers (see footnote 34). The seller can earn more from her remaining inventory. Thus, a scarcity signal can be beneficial even if does not change revenue or the number of units sold. These dynamic considerations go beyond the scope of what can be done with these data.

21 Conclusion

This paper computes the value of scarcity signals to a Bayesian risk neutral consumer in the context of the Online travel industry. We find that scarcity signals can be valuable to both an unsophisticated traveler (who does not condition her decision on publicly available information) and to a sophisticated one. That being said, scarcity signals benefit only a small range of consumers and even for these consumers the signals have a very small impact on consumer welfare, of the order of a few dollars for a ticket that costs on average \$300. But signals can have a significant proportional impact on expected utility because it influences travelers who would not receive much surplus in the absence of signals. For most consumers the gains from delaying purchase is much larger than the additional gains that can be obtained by conditioning this decision (to delay purchase) on the realization of the scarcity signal. This paper takes a first step toward measuring the benefits of delaying travel booking. The bigger picture is to study the inefficiencies associated with travel booking and this includes the time wasted searching, sub-optimal choice due to pre-mature commitment, no-shows and cancellations, to name just a few issues... We also find that scarcity signals have little impact on average revenue and average consumption. This is because the revenue increase from the consumers who anticipate a purchase (buy when the signal is bad) are balanced by the losses from the consumers who do the opposite (wait when the signal is good). The finding that scarcity signals do not affect revenue under a Bayesian benchmark is somewhat puzzling and calls for more research. It could be that the use of signal is the equilibrium outcome of a game played between suppliers (something not modeled here), and as a group, suppliers would be better-off without signals. Alternatively, it could be that some consumers deviate from the rational benchmark and that we end up under estimating the impact of signals. We briefly discuss the issue of behavioral consumers in the Appendix (Section 24).

22 Appendix: Notations

In the result section, we typically normalize all expression assuming $p_0 = 1$, $\mathbf{v}(r) = 1 + r$ and $\mathbf{r}(v) = v - 1$. The value of v is thus expressed in units of p_0. Averages are computed over the individuals who respond to the signal, $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^g)]$, and using arithmetic average ($g(v) = 1$). We obtain:[44]

Using equations 10, 12 and 13, we obtain $\Delta\overline{U}$, $\Delta\overline{R}$, $\Delta\overline{C}$. For example, $\Delta\overline{U} = \tau_b \int_{\rho^b}^{\rho^n} \left(r - \int_{-1}^{r} F^b(x)dx \right) dr + (1 - \tau_b) \int_{\rho^n}^{\rho^g} \left(\int_{-1}^{r} F^g(x)dx - r \right) dr.$

23 Appendix: Proofs

Proof of Lemma 1 The function $G^s(x) = x - \int_{-1}^{x} F^s(r)dr$ is strictly increasing on the support of F^s. We have $G^s(-1) < 0$ and applying integration by parts $\lim_\infty G^s(x) = \lim_\infty \left(x + \int_{-1}^{x} rdF^s(r) - xF^s(x) \right) = \lim_\infty x\,(1 - F^s(x)) +$

[44] Note that $\int_{-1}^{r}(1 + x)dF(x) = (1 + r)F(r) - \int_{-1}^{r} F(x)dx$.

Table 23: Notations

	Single Consumer
$g(v), G(v)$	Consumer valuation PDF and cdf
$s \in \{n, b, g\}$	State (no signal, bad, good)
$p_1 = (1 + r^s)p_0$	Prices (period 0 and 1)
Er^s	Mean growth rate
$F^s()$	Distributions of growth rate (eq. 7)
$H(r)$	$\int_{-1}^{r}(F^g(y) - F^b(y))dy$
$\mathbf{v}(r)$, $\mathbf{r}(v)$	$\mathbf{v}(r) = p_0(1 + r)$, $\mathbf{r}(v) = \frac{v}{p_0} - 1$
ρ^s	Marginal consumer (eq. 8)
$U_0(v)$	Utility of consumer v if she does not delay purchase
$U_1^s(v)$	Expected utility of consumer v given belief F^s (eq. 9)
$U_1^i(v)$	$\tau_b U^b(v) + (1 - \tau_b)U^g(v)$
I	Maximum value of information (eq. 11)
$R(v)$, $C(v)$	Revenue and consumption from consumer v under prior ($s = n$)
$\Delta U(v)$, $\Delta R(v)$, $\Delta C(v)$	Impact of signal on individual outcomes (see eq. 10, 12 and 13)
	Consumer Aggregation
$\overline{U}$, $\overline{R}$, $\overline{C}$	$\overline{U} = \int_{\mathbf{v}(\rho^b)}^{\mathbf{v}(\rho^g)} U^n(v)dv$ and same for $\overline{R}$, $\overline{C}$
$\Delta\overline{U}$, $\Delta\overline{R}$, $\Delta\overline{C}$	Average impact across all v (e.g. $\Delta\overline{U} = \int_{\mathbf{v}(\rho^b)}^{\mathbf{v}(\rho^g)} \Delta U(v)dv$)
$F(r, y)$, $S(r)$, $F^s(r\|x)$	Small signal (Proposition 2)
$\Delta\overline{U}(x)$, $\Delta\overline{R}(x)$, $\Delta\overline{C}(x)$	Utility, Revenue Consumption with small signal

Table 24: Utility, Revenue, Consumption

	$v \leq \mathbf{v}(\rho^n)$	$v \geq \mathbf{v}(\rho^n)$	Consumer Aggregation	
$U^n(v)$	$\int_{-1}^{v-1} F^n(x)dx$	$v - 1$	$\overline{U}$	$\int_{\rho^b}^{\rho^n} \left(\int_{-1}^{r} F^n(x)dx\right) dr + \int_{\rho^n}^{\rho^g} rdr$
$R(v)$	$\int_{-1}^{v-1}(1 + r)dF^n(r)$	1	$\overline{R}$	$\int_{\rho^b}^{\rho^n} \left((1 + r)F^n(r) - \int_{-1}^{r} F^n(x)dx\right) dr + \rho^g - \rho^n$
$C(v)$	$F(v - 1)$	1	$\overline{C}$	$\int_{\rho^b}^{\rho^n} F^n(r)dr + \rho^g - \rho^n$

Er^s. If $Er^s > 0$, equation (8) has a unique solution ρ^s and this solution is non-negative since $G^s(0) \leq 0$. Next, we prove by contradiction that $Min(\rho^g, \rho^b) \leq \rho^n \leq Max(\rho^g, \rho^b)$. Assume, for example, that $Min(\rho^g, \rho^b) > \rho^n$. The following three obsevations (a) $\rho^n < \rho^g$, (b) $G^g(x)$ strictly increasing, and (c) $G^g(\rho^g) = 0$ imply that $G^g(\rho^n) < 0$. The same reasoning implies that $G^b(\rho^n) < 0$. Thus $\tau_b G^b(\rho^n) + (1 - \tau^b)G^g(\rho^n) < 0$ or $\rho^n < \int_{-1}^{\rho^n} F^n(r)dr$. A contradiction. The same logic proves the other inequality. Finally, $v - p_0 - E(Max(v - p_1, 0)|s) = p_0 G^s(\mathbf{r}(v))$. Thus, consumer v stricly prefers to buy early when her belief is $F^s()$ if and only if $G^s(\mathbf{r}(v)) > 0$, or $v \in (\mathbf{v}(\rho^s), \infty)$. $\square$

Proof of Proposition 1: (a) For $v \in [0, \mathbf{v}(\rho^b))$ we have $E(Max(v - p_1, 0)|s) > v - p_0$. Thus, the consumer prefers to wait with or without signal. (b) For $v \in (\mathbf{v}(\rho^b), \mathbf{v}(\rho^n))$ we have $E(Max(v - p_1, 0)|n) > v - p_0$, $E(Max(v - p_1, 0)|g) > v - p_0$ and $E(Max(v - p_1, 0)|b) < v - p_0$. Thus, the consumer buys early only if the signal is bad. (c) For $v \in (\mathbf{v}(\rho^n), \mathbf{v}(\rho^g))$ we have $v - p_0 > E(Max(v - p_1, 0)|n)$, $v - p_0 > E(Max(v - p_1, 0)|b)$ and $E(Max(v - p_1, 0)|g) > v - p_0$. Thus, the consumer waits only if the signal is good. (d) For $v > \mathbf{v}(\rho^g)$ we have $E(Max(v - p_1, 0)|s) < v - p_0$. Thus, the consumer prefers to buy early with or without signal. $\square$

Derivation of equation (10) for $\Delta U(v)$: Consumer $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$ waits without a signal. Her expected utility is $p_0 \int_{-1}^{\mathbf{r}(v)} F^n(r)dr$. With the signal, she wait when the realization is good. Her expected utility is $p_0 \int_{-1}^{\mathbf{r}(v)} F^g(r)dr$. She buys early when the realization is bad and receive utility $v - p_0$. Taking expectation, we obtain:

$$\Delta U(v) = \tau_b(v - p_0) + (1 - \tau_b)p_0 \int_{-1}^{\mathbf{r}(v)} F^g(r)dr - p_0 \int_{-1}^{\mathbf{r}(v)} F^n(r)dr$$

for $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$. Using identity (7), we obtain the top part in equation (10). The same reasoning applies to the bottom part of equation (10) corresponding to $v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]$. From Lemma 1 and the convention used for labeling states (b,g), we have $p_0 \int_{-1}^{\mathbf{r}(v)} F^n(r)dr > v - p_0 > p_0 \int_{-1}^{\mathbf{r}(v)} F^b(r)dr$ for $v \in (\mathbf{v}(\rho^b), \mathbf{v}(\rho^n))$. This implies $\Delta U(v) > 0$. Similarly, we have $p_0 \int_{-1}^{\mathbf{r}(v)} F^g(r)dr > v - p_0 > p_0 \int_{-1}^{\mathbf{r}(v)} F^n(r)dr$ for $v \in (\mathbf{v}(\rho^n), \mathbf{v}(\rho^g))$, and again $\Delta U(v) > 0$.

Proof of Corollary 2: $\frac{\partial}{\partial v}\Delta U(v) = \tau_b(1-F^b(\mathbf{r}(v))) \geq 0$ for $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$ and $\frac{\partial}{\partial v}\Delta U(v) = (1-\tau_b)(F^g(\mathbf{r}(v)) - 1) \leq 0$ for $v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]$. Thus, $\Delta U(v)$ reaches a maximum at $\mathbf{v}(\rho^n)$. $\square$

Properties of the Revenue Function:

Using the identity $\int_{-1}^{\rho}(1+r)dF(r) = (1+\rho)F(\rho) - \int_{-1}^{\rho} F(r)dr$, we can rewrite equation (12) as

$$\begin{cases} p_0\tau_b\left(1-(1+\mathbf{r}(v))F^b(\mathbf{r}(v)) + \int_{-1}^{\mathbf{r}(v)} F^b(r)dr\right), & \text{if } v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)] \\ -p_0(1-\tau_b)\left(1-(1+\mathbf{r}(v))F^g(\mathbf{r}(v)) + \int_{-1}^{\mathbf{r}(v)} F^g(r)dr\right), & \text{if } v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]. \end{cases} \tag{14}$$

The supplier's profit function has the following properties.

Lemma 2. *(a)* $\Delta R'(v) < 0$ *for* $v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)]$. $\Delta R'(v) > 0$ *for* $v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]$. *(b)* $\Delta R(\mathbf{v}(\rho^b)) = p_0\tau_b(1+\rho^b)(1-F^b(\rho^b)) > 0$, $\Delta R(\mathbf{v}(\rho^g)) = -p_0(1-\tau_b)(1+\rho^g)(1-F^g(\rho^g)) < 0$. $\lim_{\rho\to\rho^{n+}} \Delta R(\mathbf{v}((\rho)) < 0$.
$\lim_{\rho\to\rho^{n+}} \Delta R(\mathbf{v}((\rho)) - \lim_{\rho\to\rho^{n-}} \Delta R(\mathbf{v}((\rho)) = -p_0(1+\rho^n)(1-F^n(\rho^n)) < 0$.
$\lim_{\rho\to\rho^{n-}} \Delta R(\mathbf{v}((\rho)) > 0$ *if and only if* $(1+\rho^b)(1-F^b(\rho^n)) > \int_{\rho^b}^{\rho^n}(F^b(\rho^n) - F^b(x))dx$. *(c)* $\Delta R(v) < 0$ *for* $v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]$.

Note that the Lemma implies
$\Delta R(\mathbf{v}(\rho^b)) = Max_v\Delta R(v)$ and $\lim_{\rho\to\rho^{n+}} \Delta R(\mathbf{v}((\rho)) = Min_v\Delta R(v)$.
The jump at $\mathbf{v}(\rho^b)$ is equal to $\Delta R(\mathbf{v}(\rho^b)) = p_0\tau_b(1+\rho^b)(1-F^b(\rho^b))$.
Proof of Lemma 2 (a) Take derivatives with respect to v in equation (12)

$$\Delta R'(v) = \begin{cases} -p_0\tau_b(1+\mathbf{r}(v))f^b(\mathbf{r}(v)) < 0, & \text{if } v \in [\mathbf{v}(\rho^b), \mathbf{v}(\rho^n)] \\ p_0(1-\tau_b)(1+\mathbf{r}(v))f^g(\mathbf{r}(v)) > 0, & \text{if } v \in [\mathbf{v}(\rho^n), \mathbf{v}(\rho^g)]. \end{cases}$$

(b) From equation (14),
we have $\Delta R(\mathbf{v}(\rho^b)) = p_0\tau_b\left(1-(1+\rho^b)F^b(\rho^b) + \int_{-1}^{\rho^b} F^b(r)dr\right)$ and since $\int_{-1}^{\rho^b} F^b(r)dr = \rho^b$, we obtain,
$\Delta R(\mathbf{v}(\rho^b)) = p_0\tau_b(1+\rho^b)(1-F^b(\rho^b))$. Applying the same logic one obtains $\Delta R(\mathbf{v}(\rho^g))$. Evaluate equation (14) at ρ^{n+}, add ρ^n and subtract $\int_{-1}^{\rho^n} F^n(r)dr$ (recall $\rho^n = \int_{-1}^{\rho^n} F^n(r)dr$), to obtain
$\lim_{\rho\to\rho^{n+}} \Delta R(\mathbf{v}((\rho)) = -p_0(1-\tau_b)\left((1+\rho^n)(1-F^g(\rho^n)) + \int_{-1}^{\rho^n}(F^g(r) - F^n(r))dr\right) < 0$
where the inequality follows from $F^g(r) > F^n(r)$. Evaluate equation (14) at

ρ^{n-}, $\lim_{\rho\to\rho^{n-}} \Delta R(\mathbf{v}((\rho)) = p_0\tau_b \left(1-(1+\rho^n)F^b(\rho^n)+\int_{-1}^{\rho^n} F^b(r))dr\right)$. But since $\rho^b = \int_{-1}^{\rho^b} F^b(r))dr$, we obtain
$\lim_{\rho\to\rho^{n-}} \Delta R(\mathbf{v}((\rho)) = p_0\tau_b \left(1+\rho^b-(1+\rho^n)F^b(\rho^n)+\int_{\rho^b}^{\rho^n} F^b(r))dr\right)$ or
$\lim_{\rho\to\rho^{n-}} \Delta R(\mathbf{v}((\rho)) = p_0\tau_b \left((1+\rho^b)(1-F^b(\rho^n))-\int_{\rho^b}^{\rho^n} (F^b(\rho^n)-F^b(r))dr\right)$.
And $\lim_{\rho\to\rho^{n-}} \Delta R(\mathbf{v}((\rho)) > 0$ is equivalent to $(1+\rho^b)(1-F^b(\rho^n)) > \int_{\rho^b}^{\rho^n} (F^b(\rho^n)-F^b(r))dr$. Although one can construct examples such that this condition is violated, this does not happen in our application because $\rho^n-\rho^b$ is small relative to $1-F^b(\rho^n)$. (c) $\Delta R(v)<0$ for $v\in[\mathbf{v}(\rho^n),\mathbf{v}(\rho^g)]$ follows from $\Delta R(v)$ monotone in $[\mathbf{v}(\rho^n),\mathbf{v}(\rho^g)]$, $\Delta R(\mathbf{v}(\rho^g))<0$ and $\lim_{\rho\to\rho^{n+}} \Delta R(\mathbf{v}((\rho))<0$. □
Proof of Proposition 2: Let $\rho(y)$ denote the indifferent consumer when

the cumulative distribution of price returns is $F(r,y)$. Specifically, $\rho(y)$ is defined by equation (8) after replacing $F^s(r)$ with $F(r,y)$. For ease of notation, we denote $\rho^n=\rho(0)$ and $v^n=\mathbf{v}(\rho(0))$. When the cumulative distribution of price returns is $F(r,y)$, the supplier revenues are

$$R(y) = p_0\int_0^{\mathbf{v}(\rho(y))}\left(\int_{-1}^{\mathbf{r}(v)}(1+r)dF(r,y)\right)dG(v)+p_0\left(1-G(\mathbf{v}(\rho(y)))\right).$$

This is because a traveler with a valuation below $\mathbf{v}(\rho(y))$ waits, and subsequently purchases if the price is below her valuation. A traveler with a valuation above $\mathbf{v}(\rho(y))$ buys early. Expected revenues with signal x are $\tau_b R(-x(1-\tau_b))+(1-\tau_b)R(x\tau_b)$ and

$$\Delta\overline{R}(x) = \tau_b R(-x(1-\tau_b))+(1-\tau_b)R(x\tau_b)-R(0).$$

We have $\Delta\overline{R}'(x) = \tau_b(1-\tau_b)(-R'(-x(1-\tau_b))+R'(x\tau_b))$ and $\Delta\overline{R}'(0)=0$. There is no first-order impact of the signal on profits. We have $\Delta\overline{R}''(x) = \tau_b(1-\tau_b)((1-\tau_b)R''(-x(1-\tau_b))+\tau_b R''(x\tau_b))$ and $\Delta\overline{R}''(0)=\tau_b(1-\tau_b)R''(0)$. Denote $R(y)=p_0K(\rho(y),y)$. Differentiating twice with respect to y gives $R''(y)=p_0\frac{d^2K}{dy^2}(y)$ with

$$\frac{dK}{dy} = K_1\rho' + K_2$$
$$\frac{d^2K}{dy^2} = K_1\rho'' + K_{11}\left(\rho'\right)^2 + K_{22} + 2K_{12}\rho'$$

where K_1, for example, denotes the derivative of K with respect to its first argument. We have

$$\begin{aligned}
K_1(\rho, y) &= p_0 g(\mathbf{v}(\rho)) \left(\int_{-1}^{\rho} (1+r) d(F(r,y)) - 1 \right) \\
K_2(\rho, y) &= \int_0^{\mathbf{v}(\rho)} \int_{-1}^{\mathbf{r}(v)} (1+r) dS(r) dG(v) \\
K_{11}(\rho, y) &= p_0 g(\mathbf{v}(\rho))(1+\rho) \left(f(\rho, y) - \frac{g'(\mathbf{v}(\rho))}{g(\mathbf{v}(\rho))} p_0 (1 - F(\rho, y) \right) \\
K_{22}(\rho, y) &= 0 \\
K_{12}(\rho, y) &= p_0 g(\mathbf{v}(\rho)) \left((1+\rho) S(\rho) - \int_{-1}^{\rho} S(r) dr \right)
\end{aligned}$$

where the last expression uses the identity $\int_{-1}^{\rho}(1+r)dS(r) = (1+\rho)S(\rho) - \int_{-1}^{\rho} S(r)dr$ is obtained using integration by part. Next, take $\int_{-1}^{\rho}(1+r)d(F(r,y)) - 1$ in K_1. Apply the same integration by part identity, and evaluate at $\rho = \rho(y)$ to obtain $\int_{-1}^{\rho(y)}(1+r)dF(r,y) - 1 = -(1+\rho(y))(1 - F(\rho(y), y))$ and consequently

$$K_1(\rho(y), y) = -p_0 g(\mathbf{v}(\rho(y))(1+\rho(y))(1 - F(r,y)).$$

Using equation (8), we evaluate:

$$\begin{aligned}
\rho'(0) &= \frac{\int_{-1}^{\rho(y)} S(r)dr}{1 - F(\rho, 0)} \\
\rho''(0) &= \frac{2S(r) \int_{-1}^{\rho(y)} S(r)dr}{(1 - F(\rho, 0))^2}
\end{aligned} \tag{15}$$

Replacing the expressions for K_1, K_{11}, K_{22}, K_{12}, ρ' and ρ'' gives:

$$\left.\frac{d^2K}{dy^2}\right|_{y=0} = p_0(1+\rho^n) g(v^n) \frac{\left(\int_{-1}^{\rho^n} S(r)dr\right)^2}{1 - F^n(\rho^n)} \left(\frac{f(\rho^n)}{1 - F^n(\rho^n)} - \frac{g'(v^n)}{g(v^n)} p_0 - \frac{2}{1+\rho^n} \right)$$

from which we conclude

$$\Delta \overline{R}''(0) = p_0^2 (1+\rho^n) g(v^n) \tau_b (1-\tau_b) \frac{\left(\int_{-1}^{\rho^n} S(r)dr\right)^2}{1 - F^n(\rho^n)} \left(\frac{f(\rho^n)}{1 - F^n(\rho^n)} - \frac{g'(v^n)}{g(v^n)} p_0 - \frac{2}{1+\rho^n} \right)$$

A similar argument applies for consumption where $R(y)$ is replaced with $C(y) = \int_0^{\mathbf{v}(\rho(y))} F(\mathbf{r}(v)) dG(v) + 1 - G(\mathbf{v}(\rho(y)))$. We obtain $\Delta \overline{C}''(0) = \tau_b(1 - \tau_b) C''(0)$.

24 Appendix: Extensions

This appendix briefly discusses four extensions: (a) sequential booking decisions, (b) risk aversion, (c) loss averse and behavioral traveler, (d) endogenous signals (cookies). *Sequential Booking Decisions.* The traveler can buy

at date zero, one or two. Without loss of generality, let $p_0 = 1$, $p_1 = 1+r_1$ and $p_2 = (1+r_1)(1+r_2)$. The random variable r_1 (resp. r_2) is distributed with CDF $F_1^{s_1}()$ when the first signal realization in s_1 (resp. $F_2^{s_1,s_2}(.|r_1)$ when the two signal realizations are s_1 and s_2). $F_1(r_1)$ and $F_2(r_2|r_1)$ denote the unconditional CDF defined similarly as in equation (7). All expected utilities are measured at date zero. $U_2^n(v)$ is the date zero expected utility if the traveler has no information and can delay purchase twice. U_2^i is the traveler's date zero utility if she can delay twice and receives a new signals each time. We have:
$U_2^n(v) = Max\left(v-1, \int_{-1}^{\infty} Max\left(v-(1+r_1), \int_{-1}^{\infty} Max\left(v-(1+r_1)(1+r_2), 0\right) dF_2(r_2|r_1)\right) dF_1(r_1)\right).$
This is because a traveler who has not bought in date zero or one, buys in period two if $v-(1+r_1)(1+r_2) > 0$. This corresponds to the third Max operator. If the traveler has not bought in date zero, she anticipates in period one to receive $\int_{-1}^{\infty} Max\left(v-(1+r_1)(1+r_2), 0\right) dF_2(r_2|r_1)$ if she waits. She buys in date one if $v-(1+r_1) > \int_{-1}^{\infty} Max\left(v-(1+r_1)(1+r_2), 0\right) dF_2(r_2|r_1)$. This explains the second Max operator. In date zero, she can buy or delay (first Max operator). When the traveler receives signals, the decisions are conditional on the signals' realizations and we obtain:

$$U_2^i(v) = \sum_{s_1=b,g} Pr(s_1) Max\left(v-1, \sum_{s_2=b,g} Pr(s_2|s_1) \int_{-1}^{\infty} Max\left(v-(1+r_1), \int_{-1}^{\infty} Max\left(v-(1+r_1)(1+r_2), 0\right) dF_2^{s_1,s_2}(r_2|r_1)\right) dF_1^{s_1}(r_1)\right).$$

$U_2^i(v) - U_2^n(v)$ is the value of information under sequential booking. It is positive and should be compared with $\Delta U(v)$ which is the value of information with one period.

Risk Aversion. We adapt Lemma 1's proof of the existence of a unique ρ^s to the case of risk aversion. We have $U_0^{n,0}(p_0) - U_1^{s,1}(p_0) = u(0) - \int_{-1}^{\mathbf{r}(v)} u(v - p_0(1+r))dF^s(r) < 0$. For large valuations v, $U_0^{n,0}(v) - U_1^{s,1}(v)$ is approximated by $u(v-p_0) - Eu(v - p_0(1+r))$ which is positive since $Er^s > 0$ and $u()$ concave. Thus the equation $U_0^{n,0}(v) = U_1^{s,1}(v)$ has at least one solution and an odd number of solutions. We conclude by showing that $\frac{d}{dv}\left(U_0^{n,0}(v) - U_1^{s,1}(v)\right)$ changes sign at most once, since this implies that there are at most 2 solutions. Under the assumption that u'' is constant, we have $\frac{d}{dv}U_1^{s,1}(v) = p_0 u'' \int_{-1}^{\mathbf{r}(v)} F^s(r)dr + u'(0)F(\mathbf{r}(v))$ and $\frac{d}{dv}U_0^{n,0}(v) = u'(0) +$

$p_0 u''\mathbf{r}(v)$. Thus, $\frac{d}{dv}\left(U_0^{n,0}(v) - U_1^{s,1}(v)\right) = p_0 u''\left(\int_{-1}^{\mathbf{r}(v)}(1 - F^s(r))dr - 1\right) + u'(0)(1 - F^s(\mathbf{r}(v)))$ which is decreasing in v. But $\frac{d}{dv}\left(U_0^{n,0}(0) - U_1^{s,1}(0)\right) = -p_0 u'' + u'(0) > 0$ implies that the function changes sign at most once. Note that the proof does not follow when u'' is not constant because we cannot show that $\frac{d}{dv}\left(U_0^{n,0}(v) - U_1^{s,1}(v)\right)$ is decreasing in general.[45]

Behavioral Consumer. Some travelers may not be Bayesian expected utility maximizers. Consumers could deviate from the standard paradigm in several ways. They may be loss averse, have regrets (Nasiry and Popescu, 2012), be subject to attentional shifts triggered by scarcity cues (Mullainathan and Shafir, 2013), or nurture other biases. A general treatment of behavioral consumers is beyond the scope of this paper. We briefly touch upon the case of loss averse consumers to demonstrate that generalizing the analysis can raise non-trivial issues. The central premise of loss aversion is that consumers have a reference transaction and treat differently gains and losses evaluated in comparison to this reference. It is up to the researcher to define the reference transaction and the literature offers multiple approaches. There are several candidate reference transactions in our application but two stand out: purchasing early (utility $v - p_0$) and purchasing late (utility $EMax(v - p_1, 0)$). Another consideration specific to our application is that the signal realization could changes the reference transaction. A bad signal may put more salience on purchasing early and the opposite holds for a good signal. This demonstrates that defining the reference point of a loss averse consumer raises non-trivial issues.

Endogenous Signals. As explained in the main text, the results hold for a traveler who deletes Expedia cookies after each query. In practice, Expedia may use cookies to record past searches and condition prices and signals on this information. Signals are endogenous because they depend on consumer behavior. Expedia may also use the IP address of the server from where a search originates, the type of device used to make the search (PC or mobile phone), and other query attributes... Although we could not find any evidence for endogenous signals, this possibility deserves a mention. Extending

[45] $\frac{d}{dv}\left(U_0^{n,0}(v) - U_1^{s,1}(v)\right) = p_0\left(\int_0^{\mathbf{r}(v)} u''(v - p_0(1+r))dr - \int_{-1}^{\mathbf{r}(v)} u''(v - p_0(1+r))F^s(r))dr\right) + u'(0)(1 - F^s(\mathbf{r}(v)))$. The term in bracket is not necessarily decreasing in v.

the analysis to cookies and server location, for example, would require considering two types of traveler: (1) Unsophisticated travelers conduct searches from the same server and do not delete cookies. (2) Sophisticated travelers use proxy servers and run multiple searches with different ways to handle cookies. Clearly, sophisticated travelers may be able to form more informative posteriors about future prices. That being said, as long as travelers remain Bayesian utility maximizers, they use their posterior based on the information they have acquired and the analysis carries through.

Appendix: Expedia Dataset

We have collected prices and scarcity signals for one-way travel from the Expedia website. Our scraper accepts cookies from Expedia and deletes the cookies after each query. A *route* is a pair composed of an origin and a destination airport. We use the standard three-letter designations for airports. A *query* is a pair composed of a *route* and a *departure date*. Each query is submitted on different *booking dates* and returns a number of flight options. Each *flight* is identified by a query, a departure and arrival time, the number of layovers and the carrier(s). Therefore, the nesting goes from route to query to flight. For each flight, we collect the price and signal. We construct the variable days-in-advance DiA as the number of days between the departure and booking dates. *Route Selection:* We selected routes on the basis of four criteria: (a) Routes with a dominant carrier. (b) Busy routes in term of passenger. (c) Routes previously selected in airline literature. (d) Routes with large potential price savings. We implemented these selection criteria as follows. To find routes operated by a dominant carrier, we use the T-100 data bank from the Bureau of Transportation Statistics web site (www.transtats.bts.gov). The T-100 data bank includes the number of airlines per route, as well as total number of passengers. All of our ten routes are operated by a single airline. Although there is a single carrier offering a direct flight, it is possible to combine two or more flights to make the city-pair travel. Within this set of routes, we selected the top 25 subset in terms of number of passengers transported. A similar approach was previously used by Bilotkach and Rupp (2011). Indeed, many of our routes are the same as theirs and also as Bilotkach and Pejcinovska (2007). Finally, 7 of our 10 routes have destinations listed as "domestic destinations" with "largest potential savings" by Hopper.com.[46] For these routes, signals are likely to have a greater value. *Sampling:* We run queries daily for about 100 days, starting

on July 19th and ending on October 26th, 2015. The first departure date is the 10th of August, which is approximately three weeks from our first booking date (19th of July). The following 11 departures dates are 8 days apart (18th of August, 26th of August, and so on... till November 5th). A flight expires when its departure date has past. When this happens, a new query is automatically added. Using this rolling window sampling method, we end up with 22 departure dates that cover fairly evenly the seven days of the week. Combined with 10 routes, this adds to 220 distinct queries. For some queries (those with a departure date between Octobeer 26th and November 5th), we obtain time series (price, signal) for each flight that cover 100 days. The time series are shorter for the rest of the queries. Table 25 reports the observation count broken down by route and airline. The more busy routes offer more flight options and end up with more observations. There are 9 main airlines serving the 10 routes. The flights observations have between one and 14 weeks in advance, and because of the rolling departure date sampling methodology, the number of observations is roughly evenly distributed with about 40K observations per week in advance. The number of observations is about 9K for the first departure date, peaks to about 47K on the 10th departure date and then decreases to about 2K on the last one (22nd departure date). This is because the early and late departure dates are queried less frequently. Table 26 shows that there are on average 179 flights displayed for a given query made on a given booking date. This average varies a little by route. Table 27 shows that the signal can take five positive values: 1, 2,.., 5 seats left at a given price. The frequency of sending signals with high values is slightly lower. In most of the empirical analysis we assume that the signal takes a binary value: 0 if Expedia does not report a number of seats left at the posted price) and 1 otherwise.

[46] Hopper.com is an aggregator website that collects price data and reports the best days in advance to purchase tickets.

Table 25: Observation Count

Route	N
AUS-DAL	12635
CLT-FLL	80099
DFW-SNA	56752
LAS-MDW	48524
MCO-MDW	54519
MDW-LAS	57053
MIA-BOS	72854
MIA-DFW	62791
PHX-MDW	43616
SNA-DFW	52539
Total	541382

Source: expedia_v1.0.dta

Airline	N
Alaska Airlines	23534
American Airline	112925
Delta	157598
Frontier Airline	1734
JetBlue Airways	4158
Multiple Airline	149417
Sun Country Airl	142
US Airways	48562
United	39220
Virgin America	4092
Total	541382

Source: expedia_v1.0.dta

Table 26: Number of flights per queryXbooking date (mean/min/max)

route	mean	min	max
AUS-DAL	24.0979	9	42
CLT-FLL	187.0916	105	288
DFW-SNA	173.3658	56	304
LAS-MDW	175.6805	25	289
MCO-MDW	179.5969	45	289
MDW-LAS	179.6123	35	282
MIA-BOS	210.2532	76	372
MIA-DFW	203.59	99	349
PHX-MDW	148.1207	26	237
SNA-DFW	163.4948	66	278
Total	178.9122	9	372

Table 27: Observation count for each signal realization (number of seat left at posted price)

Signal Realization	**N**
0	359435
1	45611
2	39744
3	35083
4	31978
5	27655
Total	539506

References

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505, June 2010. ISSN 0162-1459, 1537-274X. doi: 10.1198/jasa.2009.ap08746. URL http://www.tandfonline.com/doi/abs/10.1198/jasa.2009.ap08746.

Praveen Aggarwal, Sung Youl Jun, and Jong Ho Huh. Scarcity messages. *Journal of Advertising*, 40(3):19–30, 2011.

Alexandra Aguirre-Rodriguez. The effect of consumer persuasion knowledge on scarcity appeal persuasiveness. *Journal of Advertising*, 42(4):371–379, 2013.

Anna Aizer. The Gender Wage Gap and Domestic Violence. *Am Econ Rev*, 100(4):1847–1859, September 2010. ISSN 0002-8282. doi: 10.1257/aer.100.4.1847. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4123456/. 00128.

Paul D. Allison. Using Panel Data to Estimate the Effects of Events. *Sociological Methods & Research*, 23(2):174–199, November 1994. ISSN 0049-1241. doi: 10.1177/0049124194023002002. URL https://doi.org/10.1177/0049124194023002002.

D. Mark Anderson and Mary Beth Walker. Does Shortening the School Week Impact Student Performance? Evidence from the Four-Day School Week. *Education Finance and Policy*, 10(3):314–349, July 2015. ISSN 1557-3060, 1557-3079.

Joshua D Angrist and Jorn-Steffen Pischke. Mostly Harmless Econometrics: An Empiricist's Companion. page 290, 2008.

Esteban M. Aucejo and Teresa Foy Romano. Assessing the effect of school days and absences on test score performance. *Economics of Education Review*, 55:70–87, December 2016. ISSN 02727757. doi: 10.1016/j.econedurev.2016.08.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0272775715301655.

Michael Baker. Industrial actions in schools: strikes and student achievement. *Canadian Journal of Economics/Revue canadienne d'économique*, 46(3): 1014–1036, 2013.

Badi H. Baltagi and Dan Levin. Cigarette taxation: Raising revenues and reducing consumption. *Structural Change and Economic Dynamics*, 3(2): 321–335, December 1992. ISSN 0954-349X. doi: 10.1016/0954-349X(92)90010-4. URL http://www.sciencedirect.com/science/article/pii/0954349X92900104.

BCTF. BCTF Online Museum > Bargaining, Jun 2016a. URL https://bctf.ca/history/rooms/BargainingDetail.aspx?id=39526. [Online; accessed 6. Jun. 2020].

BCTF. The story of the 2005 BC teachers' strike, 2016b. URL https://bctf.ca/uploadedFiles/HistoryMuseum/Rooms/Bargaining/I_Am_the_BCTF.pdf.

BCTF. About us, Jun 2016c. URL https://www.bctf.ca/AboutUs.aspx. [Online; accessed 6. Jun. 2020].

Richard A. Berk, Alec Campbell, Ruth Klap, and Bruce Western. The Deterrent Effect of Arrest in Incidents of Domestic Violence: A Bayesian Analysis of Four Field Experiments. *American Sociological Review*, 57 (5):698–708, 1992. ISSN 0003-1224. doi: 10.2307/2095923. URL http://www.jstor.org/stable/2095923. 00255.

Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates. 2003. URL http://economics.mit.edu/files/750. 00000.

C. Alan Bester, Timothy G. Conley, and Christian B. Hansen. Inference with dependent data using cluster covariance estimators. *Journal of Econometrics*, 165(2):137–151, December 2011. ISSN 03044076. doi: 10.1016/j.jeconom.2011.01.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0304407611000431.

Volodymyr Bilotkach and Marija Pejcinovska. Distribution of airline tickets: a tale of two market structures. *Available at SSRN 1031747*, 2007.

Volodymyr Bilotkach and Nicholas G Rupp. A guide to booking airline tickets online. *Available at SSRN 1966729*, 2011.

Province of British Columbia. Foundation Skills Assessment (FSA) - Province of British Columbia, Mar 2020. URL https://www2.gov.bc.ca/gov/content/education-training/k-12/administration/program-management/assessment/foundation-skills-assessment. [Online; accessed 26. May 2020].

Timothy C Brock. Implications of commodity theory for value change. *Psychological foundations of attitudes*, 1:243–275, 1968.

Eric J. Brunner and Tim Squires. The bargaining power of teachers' unions and the allocation of school resources. *Journal of Urban Economics*, 76:15–27, July 2013. ISSN 00941190. doi: 10.1016/j.jue.2013.01.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0094119013000120.

Shawn D. Bushway. The Impact of an Arrest on the Job Stability of Young White American Men. *Journal of Research in Crime and Delinquency*, 35 (4):454–479, November 1998. ISSN 0022-4278, 1552-731X. doi: 10.1177/0022427898035004005. URL http://jrc.sagepub.com/content/35/4/454. 00114.

Eve S. Buzawa and Carl G. Buzawa. *Do Arrests and Restraining Orders Work?* SAGE Publications, March 1996. ISBN 978-1-4522-4805-9. Google-Books-ID: 8wp1AwAAQBAJ.

William E. Caldwell and Loretta M. Jeffreys. The Effect of Teacher Strikes on Student Achievement: New Evidence. *Government Union Review*, 4 (1):40–58, 1983.

William E. Caldwell and Michael D. Moskalski. The Effect of School District Strikes on Student Achievement. *Government Union Review*, 2(4):3–14, 1981.

A. Colin Cameron and Douglas L. Miller. A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372, 2015. URL http://jhr.uwpress.org/content/50/2/317.short. 00451.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90(3):414–427, July 2008. ISSN 0034-6535. doi: 10.1162/rest.90.3.414. URL http://dx.doi.org/10.1162/rest.90.3.414.

David Card and Alan B Krueger. Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. page 41, 1992.

Robert Carini, Brian Powell, and Lala Carr Steelman. Do Teacher Unions Hinder Educational Performance?: Lessons Learned from State SAT and ACT Scores. *Harvard Educational Review*, 70(4):437–467, December 2000. ISSN 0017-8055. doi: 10.17763/haer.70.4.w17t1201442683k6. URL https://www.hepgjournals.org/doi/abs/10.17763/haer.70.4.w17t1201442683k6.

Jay G. Chambers. The impact of collective bargaining for teachers on resource allocation in public school districts. *Journal of Urban Economics*, 4(3):324–339, July 1977. ISSN 00941190. doi: 10.1016/0094-1190(77)90015-8. URL https://linkinghub.elsevier.com/retrieve/pii/0094119077900158.

K. Hung Chan, Jack C. Hayya, and J. Keith Ord. A Note on Trend Removal Methods: The Case of Polynomial Regression versus Variate Differencing. *Econometrica*, 45(3):737, April 1977. ISSN 00129682. doi: 10.2307/1911686. URL http://www.jstor.org/stable/1911686?origin=crossref.

Meda Chesney-Lind. CRIMINALIZING VICTIMIZATION: THE UNINTENDED CONSEQUENCES OF PRO-ARREST POLICIES FOR GIRLS AND WOMEN. *Criminology <html_ent glyph="@amp;" ascii="&"/> Public Policy*, 2(1):81–90, November 2002. ISSN 1538-6473, 1745-9133. doi: 10.1111/j.1745-9133.2002.tb00108.x. URL http://doi.wiley.com/10.1111/j.1745-9133.2002.tb00108.x.

Harrell Chesson, Paul Harrison, and William J. Kassler. Sex under the Influence: The Effect of Alcohol Policy on Sexually Transmitted Disease Rates in the United States. *The Journal of Law and Economics*, 43(1):215–238, April 2000. ISSN 0022-2186. doi: 10.1086/467453. URL http://www.journals.uchicago.edu/doi/abs/10.1086/467453.

Pascal Courty and Sinan Ozel. The value of online scarcity signals. *Information Economics and Policy*, 46:23–40, March 2019. ISSN 0167-6245. doi: 10.1016/j.infoecopol.2018.12.003. URL http://www.sciencedirect.com/science/article/pii/S0167624517301257.

Joshua M. Cowen and Katharine O. Strunk. The impact of teachers' unions on educational outcomes: What we know and what we need to learn. *Economics of Education Review*, 48:208–223, October 2015. ISSN 02727757. doi: 10.1016/j.econedurev.2015.02.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0272775715000242.

Ruomeng Cui, Dennis J Zhang, and Achal Bassamboo. Learning from inventory availability information: Field evidence from amazon. 2016.

James D Dana, Jr. Advance-purchase discounts and price discrimination in competitive markets. *Journal of Political Economy*, 106(2):395–422, 1998.

Stefano DellaVigna and Matthew Gentzkow. Persuasion: Empirical Evidence. Working Paper 15298, National Bureau of Economic Research, August 2009. URL http://www.nber.org/papers/w15298. 00220.

Raymond Deneckere and James Peck. Dynamic competition with random demand and costless search: A theory of price posting. *Econometrica*, 80 (3):1185–1247, 2012.

Matthew Desmond and Nicol Valdez. Unpolicing the urban poor consequences of third-party policing for inner-city women. *American Sociological Review*, page 0003122412470829, 2012.

A. K. Dills. Alcohol Prohibition and Cirrhosis. *American Law and Economics Association*, 6(2):285–318, August 2004. ISSN 1465-7260. doi: 10.1093/aler/ahh003. URL https://academic.oup.com/aler/article-lookup/doi/10.1093/aler/ahh003.

Laura Dugan. Domestic Violence Legislation: Exploring Its Impact on the Likelihood of Domestic Violence, Police Involvement, and Arrest*. *Criminology & Public Policy*, 2(2):283–312, March 2003. ISSN 1745-9133. doi: 10.1111/j.1745-9133.2003.tb00126.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1745-9133.2003.tb00126.x/abstract. 00000.

Laura Dugan, Daniel S. Nagin, and Richard Rosenfeld. Exposure Reduction or Retaliation? The Effects of Domestic Violence Resources on Intimate-Partner Homicide. *Law & Society Review*, 37(1): 169–198, 2003. URL http://onlinelibrary.wiley.com/doi/10.1111/1540-5893.3701005/full. 00175.

Malcolm M. Duplantis, Timothy D. Chandler, and Terry G. Geske. The growth and impact of teachers' unions in states without collective bargaining legislation. *Economics of Education Review*, 14(2):167–178, June 1995. ISSN 02727757. doi: 10.1016/0272-7757(95)90396-P. URL https://linkinghub.elsevier.com/retrieve/pii/027277579590396P.

Randall W. Eberts and Joe A. Stone. TEACHER UNIONS AND THE COST OF PUBLIC EDUCATION. *Economic Inquiry*, 24(4):631–643, October 1986. ISSN 00952583, 14657295. doi: 10.1111/j.1465-7295.1986.tb01838.x. URL http://doi.wiley.com/10.1111/j.1465-7295.1986.tb01838.x.

Benjamin Edelman. Using internet data for economic research. *The Journal of Economic Perspectives*, pages 189–206, 2012.

Diego Escobari. Dynamic pricing, advance sales and aggregate demand learning in airlines. *The Journal of Industrial Economics*, 60(4):697–724, 2012.

Amy Farmer and Jill Tiefenthaler. Explaining the Recent Decline in Domestic Violence. *Contemporary Economic Policy*, 21(2):158–172, April 2003. ISSN 1465-7287. doi: 10.1093/cep/byg002. URL http://onlinelibrary.wiley.com/doi/10.1093/cep/byg002/abstract. 00089.

Barbara Fedders. Lobbying for mandatory-arrest policies: Race, class, and the politics of the battered women's movement. *NYU Rev. L. & Soc. Change*, 23:281, 1997. 00074.

Martin Fischer, Martin Karlsson, Therese Nilsson, and Nina Schwarz. The Long-Term Effects of Long Terms. page 76, 2017.

Richard B. Freeman and Casey Ichniowski, editors. *When public sector workers unionize*. A National Bureau of Economic Research project report. University of Chicago Press, Chicago, 1988. ISBN 978-0-226-26166-9.

Victoria Frye, Mary Haviland, and Valli Rajah. Dual Arrest and Other Unintended Consequences of Mandatory Arrest in New York City: A Brief

Report. *Journal of Family Violence*, 22(6):397–405, July 2007. ISSN 0885-7482, 1573-2851. doi: 10.1007/s10896-007-9094-y. URL http://link.springer.com/10.1007/s10896-007-9094-y. 00041.

Daniel G. Gallagher. Teacher Negotiations, School District Expenditures, and Taxation Levels. *Educational Administration Quarterly*, 15(1):67–82, February 1979. ISSN 0013-161X. doi: 10.1177/0013131X7901500105. URL https://doi.org/10.1177/0013131X7901500105.

Matthew Gentzkow and Emir Kamenica. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

Kristopher S Gerardi and Adam Hale Shapiro. Does competition reduce price dispersion? new evidence from the airline industry. *Journal of Political Economy*, 117(1):1–37, 2009.

Heribert Gierl and Verena Huettl. Are scarce products always more attractive? the interaction of different types of scarcity signals with products' suitability for conspicuous consumption. *International Journal of Research in Marketing*, 27(3):225–235, 2010.

David E. Giles. Interpreting Dummy Variables in Semi-logarithmic Regression Models: Exact Distributional Results. 2011. URL https://www.researchgate.net/profile/David_Giles3/publication/228954943_Interpreting_Dummy_Variables_in_Semi-Logarithmic_Regression_Models_Exact_Distributional_Results/links/02e7e529a1cec79fe6000000.pdf. 00000.

Benjamin Hansen. School Year Length and Student Performance: Quasi-Experimental Evidence. *SSRN Electronic Journal*, 2011. ISSN 1556-5068. doi: 10.2139/ssrn.2269846. URL http://www.ssrn.com/abstract=2269846.

Hart, W. L. *Attorney General's Task Force on Family Violence: final report, September 1984*. The Task Force, Washington, D.C., 1984. URL https//catalog.hathitrust.org/Record/000645353.

Demise Herd. Ideology, history and changing models of liver cirrhosis epidemiology. *Addiction*, 87(8):1113–1126, August 1992. ISSN 0965-2140, 1360-0443. doi: 10.1111/j.1360-0443.1992.tb01998.x. URL http://doi.wiley.com/10.1111/j.1360-0443.1992.tb01998.x.

David Hirschel, Eve Buzawa, April Pattavina, and Don Faggiani. Domestic Violence and Mandatory Arrest Laws: To What Extent Do They Influence Police Arrest Decisions? *The Journal of Criminal Law and Criminology (1973-)*, 98(1):255–298, October 2007a. ISSN 0091-4169. URL http://www.jstor.org/stable/40042852. 00073.

Teck-Hua Ho, Christopher S Tang, and David R Bell. Rational shopping behavior and the option value of variable pricing. *Management Science*, 44(12-part-2):S145–S160, 1998.

C. M. Hoxby. How Teachers' Unions Affect Education Production. *The Quarterly Journal of Economics*, 111(3):671–718, August 1996. ISSN 0033-5533, 1531-4650. doi: 10.2307/2946669. URL https://academic.oup.com/qje/article-lookup/doi/10.2307/2946669.

Mathias Huebener and Jan Marcus. Compressing instruction time into fewer years of schooling and the impact on student performance. *Economics of Education Review*, 58:1–14, June 2017. ISSN 02727757. doi: 10.1016/j.econedurev.2017.03.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0272775716304186.

Drew Humphries. No Easy Answers: Public Policy, Criminal Justice, and Domestic Violence. *Criminology & Public Policy*, 2(1):91–96, November 2002. ISSN 1745-9133. doi: 10.1111/j.1745-9133.2002.tb00109.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1745-9133.2002.tb00109.x/abstract. 00028.

B. Hunter and J. Borland. Effect of Arrest on Indigenous Employment Prospects, The. *BOCSAR NSW Crime and Justice Bulletins*, page 8, 1999. URL http://search.informit.com.au/documentSummary;dn=916095124847041;res=IELHSS. 00028.

Radha Iyengar. Does the certainty of arrest reduce domestic violence? Evidence from mandatory and recommended arrest laws. *Journal of Public Economics*, 93(1-2):85–98, February 2009. ISSN 00472727. doi: 10.1016/j.jpubeco.2008.09.006. URL http://linkinghub.elsevier.com/retrieve/pii/S0047272708001345. 00055.

David Jaume and Alexander Willen. The Long-Run Effects of Teacher Strikes: Evidence from Argentina. *Journal of Labor Economics*, 37

(4):1097–1139, October 2019. ISSN 0734-306X, 1537-5307. doi: 10.1086/703134. URL https://www.journals.uchicago.edu/doi/10.1086/703134.

David R. Johnson. Do Strikes and Work-to-Rule Campaigns Change Elementary School Assessment Results? *Canadian Public Policy / Analyse de Politiques*, 37(4):479–494, 2011. URL http://www.jstor.org/stable/23074917.

Lynn Langton, Marcus Berzofsky, Christopher P. Krebs, and Hope Smiley-McDonald. Victimizations not reported to the police, 2006-2010. Technical report, US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, 2012. URL http://www.dgfpi.de/tl_files/pdf/medien/2012-08-20_NCVS_USA_Victimizations-not-reported-dot-Police_2006-2010.pdf.

John Lazarev. The welfare effects of intertemporal price discrimination: an empirical analysis of airline pricing in us monopoly markets. *New York University*, 2013.

Jong-Wha Lee and Robert J Barro. Schooling Quality in a Cross-Section of Countries. page 24, 2001.

Tracy R Lewis and David EM Sappington. Supplying information to facilitate price discrimination. *International Economic Review*, pages 309–327, 1994.

Jun Li, Nelson Granados, and Serguei Netessine. Are consumers strategic? structural estimation from the air-travel industry. *Management Science*, 60(9):2114–2137, 2014.

Kimberly A Lochner, Ichiro Kawachi, Robert T Brennan, and Stephen L Buka. Social capital and neighborhood mortality rates in Chicago. *Social Science & Medicine*, 56(8):1797–1805, April 2003. ISSN 0277-9536. doi: 10.1016/S0277-9536(02)00177-6. URL http://www.sciencedirect.com/science/article/pii/S0277953602001776. 00486.

Johnathan Lott and Lawrence W. Kenny. State teacher union strength and student achievement. *Economics of Education Review*, 35:93–103, August 2013. ISSN 02727757. doi: 10.1016/j.econedurev.2013.03.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0272775713000496.

Michael F. Lovenheim and Alexander Willen. The Long-Run Effects of Teacher Collective Bargaining. *American Economic Journal: Economic Policy*, 11(3):292–324, August 2019. ISSN 1945-7731, 1945-774X. doi: 10.1257/pol.20170570. URL https://pubs.aeaweb.org/doi/10.1257/pol.20170570.

MichaelÂ F. Lovenheim. The Effect of Teachers' Unions on Education Production: Evidence from Union Election Certifications in Three Midwestern States. *Journal of Labor Economics*, 27(4):525–587, October 2009. ISSN 0734-306X, 1537-5307. doi: 10.1086/605653. URL https://www.journals.uchicago.edu/doi/10.1086/605653.

Michael Lynn. Scarcity effects on value: A quantitative review of the commodity theory literature. *Psychology & Marketing*, 8(1):43–57, 1991. URL http://onlinelibrary.wiley.com/doi/10.1002/mar.4220080105/full. 00377.

James Macinko, Frederico C Guanais, and Maria de FÃ¡tima Marinho de Souza. Evaluation of the impact of the Family Health Program on infant mortality in Brazil, 1990-2002. *Journal of Epidemiology and Community Health (1979-)*, 60(1):13–19, 2006. URL http://www.jstor.org/stable/40795067.

James G. MacKinnon. How cluster-robust inference is changing applied econometrics. *Canadian Journal of Economics/Revue canadienne d'Ã©conomique*, 52(3):851–881, August 2019. ISSN 0008-4085, 1540-5982. doi: 10.1111/caje.12388. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/caje.12388.

Benny Mantin and David Gillen. The hidden information content of price movements. *European Journal of Operational Research*, 211(2):385–393, 2011.

Dave E. Marcotte and Steven W. Hemelt. Unscheduled School Closings and Student Performance. *Education Finance and Policy*, 3(3):316–338, July 2008. ISSN 1557-3060, 1557-3079. doi: 10.1162/edfp.2008.3.3.316. URL http://www.mitpressjournals.org/doi/10.1162/edfp.2008.3.3.316.

Christopher Maxwell, Joel H. Garner, and Jeffrey A. Fagan. *The effects of arrest on intimate partner violence: New evidence from the Spouse Assault Replication Program*. US Department of Justice, Office of Justice Programs, National Institute of Justice, 2001. URL http://niccsa.org/uploads/file/f815aa02a0804639a53bb4a800952235/EFFECTSOFARRESTONINTIMATEPARTNERVIOLENCE.pdf. 00195.

Christopher D. Maxwell, Joel H. Garner, and Jeffrey A. Fagan. The Preventive Effects of Arrest on Intimate Partner Violence: Research, Policy and Theory*. *Criminology & Public Policy*, 2(1):51–80, November 2002. ISSN 1745-9133. doi: 10.1111/j.1745-9133.2002.tb00107.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1745-9133.2002.tb00107.x/abstract. 00000.

R Preston McAfee and Vera Te Velde. Dynamic pricing in the airline industry. *forthcoming in Handbook on Economics and Information Systems, Ed: TJ Hendershott, Elsevier*, 2006.

Susan L. Miller. Unintended Side Effects of Pro-Arrest Policies and Their Race and Class Implications for Battered Women: A Cautionary Note. *Criminal Justice Policy Review*, 3(3):299–317, October 1989. ISSN 0887-4034. doi: 10.1177/088740348900300305. URL https://doi.org/10.1177/088740348900300305. Publisher: SAGE Publications Inc.

Jeffrey A Miron and Jeffrey Zwiebel. Alcohol Consumption During Prohibition. Working Paper 3675, National Bureau of Economic Research, April 1991. URL http://www.nber.org/papers/w3675. Series: Working Paper Series.

Sendhil Mullainathan and Eldar Shafir. *Scarcity: Why having too little means so much*. Macmillan, 2013.

Mohita Nagpal. How to use urgency and scarcity principles to increase ecommerce sales. *https://vwo.com/blog/use-urgency-scarcity-increase-ecommerce-sales/*, 2014.

Javad Nasiry and Ioana Popescu. Advance selling when consumers regret. *Management Science*, 58(6):1160–1177, 2012.

Devah Pager. The Mark of a Criminal Record. *American Journal of Sociology*, page 39, 2003.

Alexandra Pavlidakis. Mandatory Arrest: Past Its Prime. *SANTA CLARA LAW REVIEW*, page 37, 2009.

Jorn Steffen Pischke. The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Years. *The Economic Journal*, 117(523):1216–1242, October 2007. ISSN 0013-0133, 1468-0297. doi: 10.1111/j.1468-0297.2007.02080.x. URL https://academic.oup.com/ej/article/117/523/1216-1242/5086553.

Oren Pizmony-Levy and Nancy Green Saraisky. Who Opts Out and Why? Results from a national survey on opting out of standardized tests. 2016. doi: 10.7916/D8K074GW. URL https://doi.org/10.7916/D8K074GW.

Helen Raptis and Thomas Fleming. LARGE SCALE ASSESSMENT OUTCOMES IN BRITISH COLUMBIA, 1876-1999. *Canadian Journal of Education*, page 32, 2006.

Charles A. Register and Paul W. Grimes. Collective bargaining, teachers, and student achievement. *Journal of Labor Research*, 12(2):99–109, June 1991. ISSN 0195-3613, 1936-4768. doi: 10.1007/BF02685375. URL http://link.springer.com/10.1007/BF02685375.

Heather Rose and Jon Sonstelie. School board politics, school district size, and the bargaining power of teachers' unions. *Journal of Urban Economics*, 67(3):438–450, May 2010. ISSN 00941190. doi: 10.1016/j.jue.2010.01.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0094119010000021.

Miriam H. Ruttenberg. Feminist Critique of Mandatory Arrest: An Analysis of Race and Gender in Domestic Violence Policy, A. *Am. UJ Gender & L.*, 2:171, 1994. 00000.

Lawrence W. Sherman and Richard A. Berk. The Specific Deterrent Effects of Arrest for Domestic Assault. *American Sociological Review*, 49(2):261, April 1984. ISSN 00031224. doi: 10.2307/2095575. URL http://www.jstor.org/stable/2095575?origin=crossref. 01421.

Lawrence W. Sherman and Heather M. Harris. Increased death rates of domestic violence victims from arresting vs. warning suspects in the Milwaukee Domestic Violence Experiment (MilDVE). *J Exp Criminol*, 11(1):1–20, May 2014. ISSN 1573-3750, 1572-8315. doi: 10.

1007/s11292-014-9203-x. URL http://link.springer.com/article/10.1007/s11292-014-9203-x. 00007.

Lawrence W. Sherman, Janell D. Schmidt, Dennis P. Rogan, Douglas A. Smith, Patrick R. Gartin, Ellen G. Cohn, J. Collins, and Anthony R. Bacich. The Variable Effects of Arrest on Criminal Careers: The Milwaukee Domestic Violence Experiment. *The Journal of Criminal Law and Criminology (1973-)*, 83(1):137, 1992a. ISSN 00914169. doi: 10.2307/1143827. URL http://www.jstor.org/stable/1143827?origin=crossref.

Marielle Simon, Kadriye Ercikan, and Michel Rousseau. *Improving large-scale assessment in education: Theory, issues, and practice*. Routledge, 2012.

David P. Sims. Strategic responses to school accountability measures: It's all in the timing. *Economics of Education Review*, 27(1):58–68, February 2008. ISSN 02727757. doi: 10.1016/j.econedurev.2006.05.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0272775706001373.

Betsey Stevenson and Justin Wolfers. Bargaining in the Shadow of the Law: Divorce Laws and Family Distress. *The Quarterly Journal of Economics*, 121(1):267–288, February 2006. ISSN 0033-5533. URL http://www.jstor.org/stable/25098790. 00256.

Scott E. Stickel. The anatomy of the performance of buy and sell recommendations. *Financial Analysts Journal*, 51(5):25–39, 1995. URL http://www.cfapubs.org/doi/abs/10.2469/faj.v51.n5.1933.

Andrew Sweeting and Kane Sweeney. Staggered vs. simultaneous price setting with an application to an online market. Technical report, Working paper, University of Maryland, College Park, MD, 2015.

Yuliya Talmazan. B.C. teachers' strike: The timeline. *Global News*, Sep 2014. URL https://globalnews.ca/news/1534959/b-c-teachers-strike-the-timeline.

Kenneth Wm. Thornicroft. Teacher Strikes and Student Achievement: Evidence from Ohio. *Journal of Collective Negotiations in the Public Sector*, 23(1):1–1, March 1994. ISSN 0047-2301. doi: 10.2190/65X6-ETU0-3MQ5-27Q9.

Theo MM Verhallen and Henry SJ Robben. Scarcity and preference: An experiment on unavailability and product evaluation. *Journal of economic psychology*, 15(2):315–331, 1994.

Dinand Webbink and Michèle Belot. The lost generation: The effect of teacher strikes on students. evidence from belgium. 2006.

Dinand Webbink and Michèle Belot. Do Teacher Strikes Harm Educational Attainment of Students? November 2010. URL https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9914.2010.00494.x.

C. R. Winegarden and Paula M. Bracy. Demographic Consequences of Maternal-Leave Programs in Industrial Countries: Evidence from Fixed-Effects Models. *Southern Economic Journal*, 61(4):1020, April 1995. ISSN 00384038. doi: 10.2307/1060738. URL https://www.jstor.org/stable/1060738?origin=crossref.

John Winters. Teacher Salaries and Teacher Unions: A Spatial Econometric Approach - John V. Winters, 2011, 2011. URL https://journals.sagepub.com/doi/10.1177/001979391106400406.

Stephen Worchel, Jerry Lee, and Akanbi Adewole. Effects of supply and demand on ratings of object value. *Journal of Personality and Social Psychology*, 32(5):906, 1975.

A. M. Zeoli, A. Norris, and H. Brenner. A Summary and Analysis of Warrantless Arrest Statutes for Domestic Violence in the United States. *Journal of Interpersonal Violence*, 26(14):2811–2833, September 2011. ISSN 0886-2605, 1552-6518. doi: 10.1177/0886260510390945. URL http://jiv.sagepub.com/cgi/doi/10.1177/0886260510390945. 00013.

Michael A. Zigarelli. Dispute resolution mechanisms and teacher bargaining outcomes. *Journal of Labor Research*, 17(1):135–148, March 1996. ISSN 0195-3613, 1936-4768. doi: 10.1007/BF02685789. URL http://link.springer.com/10.1007/BF02685789.

Perry A. Zirkel. The Academic Effects of Teacher Strikes. *Journal of Collective Negotiations in the Public Sector*, 21(2):1–1, June 1992. ISSN 0047-2301. doi: 10.2190/GYNQ-CJPH-AAKR-EELV.

Joan Zorza. The Criminal Law of Misdemeanor Domestic Violence, 1970-1990. *The Journal of Criminal Law and Criminology (1973-)*, 83(1):46, 1992. ISSN 00914169. doi: 10.2307/1143824. URL http://www.jstor.org/stable/1143824?origin=crossref. 00327.

Harris L. Zwerling. Pennsylvania teachers' strikes and academic performance. *Journal of Collective Negotiations*, 32(2):151–172, 2008.