ESSAYS IN PUBLIC AND LABOR ECONOMICS.

by

KENDALL HOUGHTON

A DISSERTATION

Presented to the Department of Economics
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2021

DISSERTATION APPROVAL PAGE

Student: Kendall Houghton

Title: Essays in Public and Labor Economics.

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Economics by:

Benjamin Hansen	Co-chair
Caroline Weber	Co-chair
Grant McDermott	Core Member
Laura Leete	Institutional Representative

and

Andrew Karduna	Interim Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2021

DISSERTATION ABSTRACT

Kendall Houghton

Doctor of Philosophy

Department of Economics

June 2021

Title: Essays in Public and Labor Economics.

This dissertation considers three topics in public and labor economics. Chapter I introduces the work. In Chapter II, I consider the gender wage gap in the United States by evaluating the differences in work timing and elasticity between men and women. Chapter III evaluates the interaction between Supplemental Nutrition Assistance Program (SNAP) benefit disbursement and drug-related fatalities. Chapter IV provides a general empirical test of tax invariance (TIV). In Chapter V, I conclude the work.

CURRICULUM VITAE

NAME OF AUTHOR: Kendall Houghton

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:
University of Oregon, Eugene, OR
University of Montana, Missoula, MT

DEGREES AWARDED:
Doctor of Philosophy, Economics, 2021, University of Oregon
Master of Science, Economics, 2016, University of Oregon
Bachelor of Arts, Economics, 2013, University of Montana

AREAS OF SPECIAL INTEREST:
Labor Economics
Public Economics

PROFESSIONAL EXPERIENCE:

Researcher, Institute on Ecosystems, Montana State University, 2014 - 2015

Researcher, Environmental and Natural Resource Economics Lab, University of Montana, 2013 - 2015

Research Intern, Institute on Ecosystems, University of Montana, 2012 - 2012

Research Intern, Governor Schweitzer's Energy Program, University of Montana, 2011 - 2011

GRANTS, AWARDS AND HONORS:

Outstanding Graduate Teaching, Economics Department, University of Oregon, 2019

General University Scholarship, University of Oregon, 2019

Kleinsorge Summer Research Award, University of Oregon, 2019

Kimble First-Year Teaching Award, University of Oregon, 2018

M. Gregg Smith Fellowship, College of Arts and Sciences, University of Oregon, 2018

Distinctive Scholar Award, Economics Department, University of Oregon, 2015

Kleinsorge Incoming Student Award, Economics Department, University of Oregon, 2015

Graduation with High Honors, University of Montana, 2013

Outstanding Senior in Economics, University of Montana, 2013

Positive Externality in Economics, Economics Department, University of Montana, 2013

Presidential Leadership Scholarship, University of Montana, 2009 - 2013

Western Undergraduate Exchange, University of Montana, 2009 - 2013

Watkins Undergraduate Research Scholarship, University of Montana, 2012

Research Experience for Undergraduates Program, National Science Foundation, 2012

PUBLICATIONS:

"Public Opinion about Management Strategies for a Low-Profile Species across Multiple Jurisdictions: Whitebark pine in the Northern Rockies" *People and Nature*, 2020 (with Eric D. Raile, Helen T. Naughton, Lena Wooldridge, Courtney Kellogg, Michael P. Wallner, and Elizabeth A. Shanahan)

"How much are US households prepared to pay to manage and protect whitebark pine (Pinus albicaulis Engelm.)?" *Forestry: An International Journal of Forest Research*. 2018 (with Eric D. Raile, Elizabeth A. Shanahan, Helen T. Naughton, and Michael P. Wallner)

"Trade and Sustainability: The Impact of the International Tropical Timber Agreements on Exports," *International Environmental Agreements: Politics, Law and Economics*. 2017 (with Helen T. Naughton)

"International Environmental Agreement Effectiveness: A Review of Empirical Studies," Research Handbooks in Comparative Law and Economics, Chapter 18. 2016 (with Helen T. Naughton)

ACKNOWLEDGEMENTS

I thank the members of my committee for their guidance in completing this work. I thank Caroline Weber for providing wise research counsel, remarkable econometric expertise, and consistently gracious mentorship. I am grateful to Benjamin Hansen for bringing clear eyes on the contribution of my papers and encouragement to push further. I thank Grant McDermott for sharing both his expertise and enthusiasm on data science. I would like to acknowledge Jeremy Piger, Shankha Chakraborty, Teri Rowe, and Sharon Kaplan for their support through this process. I thank Helen Naughton for encouraging my interest in economics.

I thank my family, friends, and community. I thank my mom, Lisa, and my dad, Dennis.

To everyone. You are what an economist looks like.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

This work contributes to our understanding of human behavior and public policy in the United States.

In Chapter II, I contribute to work on the gender wage gap in the United States. The gender wage gap in the United States is persistent and especially pronounced at the top of the distribution. Recent worker surveys suggest this gap partly driven by a difference in average work hours, even between men and women employed full-time. This paper examines gender differences in work timing and elasticity using hourly data on tech worker activity. I find both genders work outside the traditional work week, but men work more than women on nights and weekends — times when children are typically present in the home. To isolate the impact of children at home, I examine how work activity varies in response to unexpected winter weather public school closures. Women respond to these unexpected breaks in childcare by reducing work activity by 34%. Male work activity does not respond to these unexpected breaks. These results are consistent with the emerging theory that men and women in high-wage professions are working different amounts and suggest asymmetric childcare responsibilities could be a reason for the difference.

Chapter III evaluates the interaction between Supplemental Nutrition Assistance Program (SNAP) benefit disbursement and drug-related fatal automobile collisions. Distributing SNAP benefits on days other than the first of the month, adding an additional income shock to the monthly calendar, increases the number of drug-related fatal automobile collisions by 1.21 percent. A one-percentage point increase in the share of SNAP benefits distributed on a day leads to a .2 percent increase in the number of drug-related fatal automobile collisions. This estimation utilizes a novel dataset of variation in SNAP distribution dates across states, and switches in distribution date regimes within states over time to identify the causal relationship.

Chapter IV provides a general empirical test of tax invariance (TIV). When a 25 percent tax remitted by manufacturers was eliminated in Washington state and the retail excise tax was simultaneously increased from 25 to 37 percent—a shift intended to be revenue-neutral—TIV did not hold. Manufacturers kept two-thirds of their tax savings instead of passing all their savings through to retail firms via lower prices as predicted by TIV. One-third of the retail tax increase was passed on to consumers via higher retail prices – TIV would have predicted constant or even declining tax-inclusive retail prices.

Chapter V concludes.

CHAPTER II

CHILDCARE AND THE NEW PART-TIME: GENDER GAPS IN LONG-HOUR PROFESSIONS

2.1. Introduction

The median female worker in the United States makes approximately \$10,000 less than the median male worker in the United States [1]. Over the course of a lifetime, this difference in earnings leads to a significant wealth gap. Women hold 65% of U.S. student loan debt[1] [3] and are more likely to live below the poverty line at every age[2] [5]. A large body of work seeks to explain the persistent earnings gap (for a thorough review, see [6]), and finds that a difference in weekly work hours can explain much of the difference.

Historically, the difference in work hours was largely a difference in part-time versus full-time employment. As technology has made it possible to work remotely, workers have been faced with the both the possibility and, in some cases expectation, that they work much more than traditional full-time. This phenomenon has been documented in the press ([7]; [8]; [9]), and described in surveys of the workforce ([10]; [11]). This high-hour equilibrium is present in the top of the distribution and features strongly in professions that reward wage as a strongly nonlinear function of work hours [12].

Female MBAs and JDs self-report working more than 40 hours a week, but less than the hours self-reported by men in the same samples ([11], [13]).

[1]For the past 18 years, women have earned 57% of bachelor degrees in the U.S. [2].

[2]For individuals aged 65 to 74, 9.8% of women are in poverty and 7.7% of men. For individuals above the age of 75, 8% of men and 12.1% of women live in poverty, although this statistic is confounded by more women living later in life than men [4].

Self-reports from workers across all fields indicate this gender difference is a broad phenomenon [10]. In this paper, I use high-frequency records of tech worker programming activity to verify the presence of work activity outside of the traditional 9am to 5pm period in STEM professions. I then document precisely when during the week and day men are participating in programming labor and women are not, and find that women are less likely than men to work on nights and weekends.

Women may have a greater taste for leisure than men, and cultural factors may make leisure more rewarding during these hours. On the other hand, these are also times when formal childcare is less readily available. Female labor force participation is inextricably connected to childcare availability [14]. We know that the percent of mothers who work more than forty hours a week is significantly below the percent in other demographic groups, and the gap between the share of mothers and the share of non-mothers participating in these hours has been increasing since the 1990's ([15]; [10]; [11]). This is suggestive that childcare is the factor that limits mothers from working the hours that women without children and men are working.

In order to separate childcare availability from other potential mechanisms, I consider shocks to childcare availability. In my estimation, I exploit winter weather public school closures as a natural experiment in which childcare is unexpectedly unavailable. I consider female and male work activity on these days.

Male work activity does not respond to unexpected shocks in the public school calendar, but female work activity does. Female activity decrease by 34% on these days. These results are consistent with a story that households plan for childcare external to the household, but that women are more likely to be responsible for within-household childcare when external childcare is unavailable.

It is possible that the parental status of the women in my data is systematically different than the parental status of the men in my data. I could, for example, be comparing single men to women with children. In order to increase the likelihood that I am comparing men with children to women with children, I augment my primary data with estimated age for a subset of individuals. I use publicly available information from LinkedIn to estimate the individual's age using their education levels and corresponding graduation years[3]. I match individuals by limiting my LinkedIn search to those working in STEM fields, in the geographic location listed in my primary dataset, with a matching name and company affiliation. When age is included, the average effect for women disappears, and a much larger effect for women between the age of 30 and 40 emerges.

The wage gap is widest among college graduates and at top of the wage distribution [17]. Science, technology, engineering, and math (STEM) are often highlighted as fields in which increased gender parity is wanted, but increases in the number of female college students enrolled in these majors have not been reflected in the workforce [18]. Female STEM graduates are less likely to take STEM jobs, and are more likely to leave the STEM workforce. Women who stay in STEM advance more slowly, and are less likely to hold management positions. These gaps have been resistant

[3] "Scraping" data from LinkedIn in this manner has been ruled legal by federal courts [16]

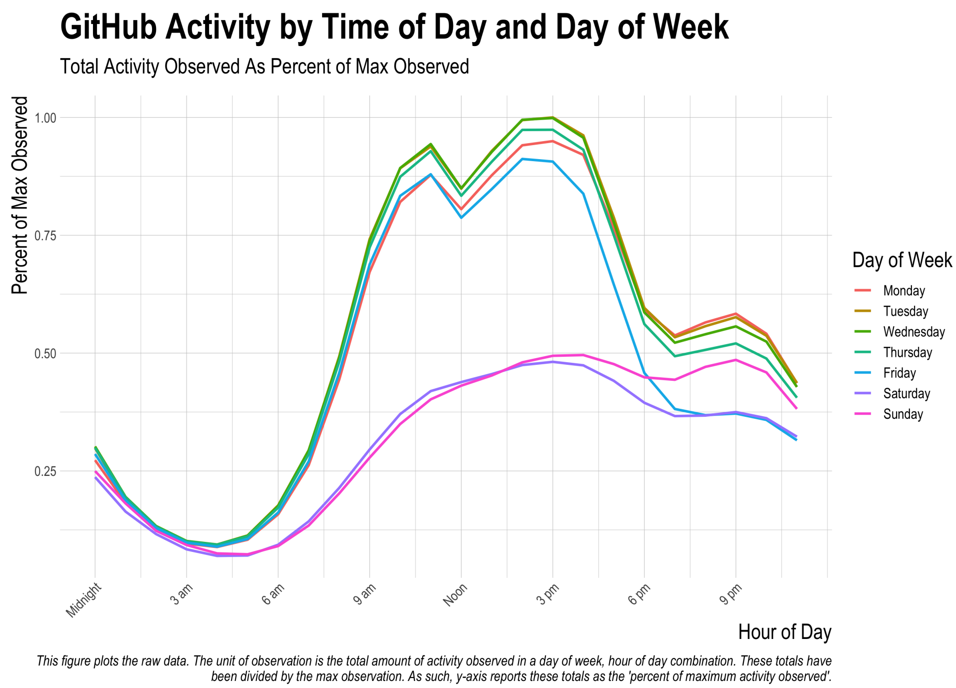
to intervention [19]. A lessened ability to work long hours may be the cause of the slow advancement and lower participation rate. My work suggests that in STEM careers, and perhaps other high-hour professions, the wage gap is not only an issue of childcare during traditional work hours, but all the other times as well.

2.2. Data

2.2.1. GitHub

In order to document worker activity at a precise timescale, I use a publicly available record of worker activity from GitHub, a version-control platform with more than 40 million users repo. Workers who use technology use version control to manage solo and collaborative projects. A basic example of version control involves saving multiple versions of a file to a local hard drive. Consider a common example of this practice, for *File*. A user might save this file as *File1*, *File2*, *File3*, etc. as both minor and major changes are made to the file. For complex projects, version control software provides a more user-friendly version of this practice. The worker first creates one file. After making changes to this file, the worker saves these changes using version control. These changes, and the lines of code or text within them, are tracked by the version control software in order to allow the worker to return to any previous version at any point. Crucially for this project, these changes as well as other activity on the version control platform are tracked by time and date. I am able to exploit the time-date system of tracking to create a dataset that describes the types of users interacting with code in a given time period.

FIGURE 2.1. **GitHub Activity by Time of Day and Day of Week**



In GitHub, activity is referred to as an event. For a detailed list and description of these events see Table A.2. Table A.1 summarizes how frequently each type of activity occurs. The second most common type of event is a “push” event, which is when a user takes file changes from their local computer to the remote copy of the file. Inside the push event details, I am able to observe each time the user has made a change to that file on their local machine. These local changes are “commits”, the most common event type, and I observe portions of their content and the timestamp of when the file was saved. Together, commits and push events represent 70% of the data. In my analysis, I limit my sample to these events. In the appendix, Figures A.1 and A.2, show activity across the day for each event type. Figure A.1 displays this for each event type, and Figure A.2

displays this for the ten most common events. All events outside the top ten represent less than 1% of the data. The data pattern for all types is qualitatively similar, with less noise as event type frequency increases.

I access the record of GitHub activity from two sources, GitHub Archive and GitHub Torrent. GitHub Archive is log of all public events that happen on GitHub. This data source is updated hourly and covers activity from 2011 forward. GitHub Torrent is meant to capture the relationships between the different aspects of GitHub – the relationships between users, files and users, and files with each other. This dataset records the stated location of users as well. GitHub torrent is updated monthly and covers 2013 to present.

I gather user name and event activity for all events except for commits from GitHub Archive. I use GitHub Torrent to collect the user location and commit activity. I merge these datasets together using the login name of the user. The raw data is processed minimally. The timestamp of activity is standardized to Universal Coordinated Time in both GitHub Archive and GitHub Torrent. In order to track when the user is working during the day, I adjust timestamps by the timezone provided in their geographic information. In the appendix, Figure A.3 shows the daily activity of users by timezone after this transformation. Each timezone follows a similar daily pattern.

I use two subsamples of the worker data for this analysis. I consider activity between 2017 and 2019 for users with a stated location in the United States and an identifiable first name. This sample is 145,333 users. I also analyze the sample of users with a stated location in Seattle or

Bellevue, Washington. There are 4,392 users in this smaller sample. Please see Table 2.1 for summary statistics on these samples, including the total number of observations and the gender compositions.

TABLE 2.1. **Data Description**

	United States		Seattle	
Total Observations	59,400,105		4,493,618	
Total Users	145,333		4,392	
Obs. Per User (Percentile)				
25	24		49	
50	97		199	
75	328		715	
90	858		2075	
95	1501		3813	
Gender Composition	*Women*	*Men*	*Women*	*Men*
No. Users	18339	126994	340	4,052
Percent of Users	12.6%	87.4%	7.7%	92.3%
Mean Users Observed Per Week	2931	22068	197	2259
Mean Users Observed Per Day	984	8584	70	794

This table describes the data for the United States sample and the sample for Seattle and includes the total number of observations and the total number of users. I include a description of the number of observations per user across different percentiles of frequency in the data. The final portion of the table provides various descriptors on the gender composition of the data.

2.2.2. Gender Imputation

I impute gender for each user. In the GitHub data, I am able to identify the first name of users who provide a first name in their profile information. In order to impute gender for each user, I use the the R package, '`gender`', which uses Census and Social Security Administration data to predict the gender of an individual when given their first name and geographic location [20]. I provide the first name of the user as collected from GitHub Archive and the geographic location as collected from GitHub Torrent.

2.2.3. School Snow Closure Records

In my analysis of winter weather school closures, I use a subsample of users in Seattle, Washington. I use the Twitter page of Seattle Public School District to identify snow and winter weather school closure dates. Most school districts do not maintain public, formal records of closures due to snow. Fortunately, Seattle Public School District maintains an active Twitter page that communicates news of school closures, delays, and changes to activity schedules. The tweets announcing these decisions remain on the Twitter page for the district. I locate relevant tweets on the Seattle Public School District page and verify these dates by searching for tweets by other users on the same day documenting the school closure and winter weather conditions.

2.2.4. LinkedIn

As part of my robustness analysis, I impute the age of users to identify likely parents. Through the GitHub data, I am able to identify user first name, last name, and company affiliations. With this information, I identify these users in LinkedIn using a web-based LinkedIn URL finder application, *PhantomBuster*. From the LinkedIn profiles, I scrape user education and dates of graduation using Selenium[4]. Using this information, I impute a likely age for the users.[5]

[4]For more information on the scraping process and code, see my co-authored work on COVID-19 with Ben Hansen, Grant McDermott, and Caroline Weber

[5]I assume individuals graduate with a Bachelor's degree at age 22, with a Master's degree at age 25, and with a PhD at age 30. The earliest listed degree and graduation date is used, as deviations from the average education path timing become more likely over time.

2.3. Overall Work Patterns

I first document the typical behavior of all users in my data. Given the high frequency of the data, it is possible to track the daily work schedule of the average user. In Figure 2.1, I plot the total amount of activity that occurs for all users on any given day of the week for every hour of that day. I scale this total amount of activity by dividing by the max total amount of activity observed for a day-of-week, hour-of-day combination.

On average, users are most active between 2pm and 3pm on Wednesday and Thursday. Activity is lowest between the hours of 1am and 7am across all days. On workdays, activity increases most sharply between 8am and 9am. There is a dip in activity between 11am and 1 o'clock that is consistent with the lunch hour. Activity decreases most sharply between 5pm and 6pm. This decrease is most pronounced on Friday.

Evenings follow a similar pattern across the week. Monday, Tuesday, Wednesday, Thursday, and Sunday nights are closely related as are Friday and Saturday nights. These grouping correspond to nights before work days and nights before weekend days. Monday and Tuesday night have the most activity, followed by Wednesday, then Thursday, and then Sunday. For the nights that occur on a work day and before a work day, there is a decrease in activity as the week progresses. After 5pm, the most work occurs at 9pm, regardless of day. Friday and Saturday are very similar from 8pm onward.

2.4. Description of Male and Female Work Habits

I begin the analysis of gender difference by examining the differences in male and female work timing across the day and week.

2.4.1. Empirical Method

For my analysis, I use the following general specification, which compares the activity for men and women over some time scale. For the following equation, I am using the example of comparing male and female activity across the work week.

$$Log(Activity_{tgd}) = \beta_0 + \theta Female_g + \sum_{d=1}^{7} \delta_d Day_t + \sum_{d=1}^{7} \gamma_d Female_g \times Day_t + \epsilon_{tgd}$$

Activity is a general term that refers to the combination of push and commit events as described in Section 2.2. This is the total amount of activity on date t for users of gender g on day of the week d. I include a fixed effect for activity by users identified as female, $Female$, in order to control for baseline differences in activity levels. In each specification, I consider various fixed effects, including month of year. In other specifications, these day of week effects are included as controls. I allow for errors with heteroskedasticity and auto-correlation by using Newey-West estimator and allowing for a two week lag. I am interested in the coefficients attached to the interaction between day of week and the female indicator variable, γ_d, which capture the differential day of week effects for women.

2.4.2. Results

Men and women display different daily and weekly work timing, on average. Figure 2.4 illustrates the difference in activity by hour of day during the work week, and Figure 2.2 illustrates the difference in activity by hour of day during the weekend. In Figure 2.3, I plot activity by day of week. Men tend to have a more diffuse work schedule, while female users are more active during times associated with a traditional work schedule.

FIGURE 2.2. **Work Activity by Gender on Weekends**

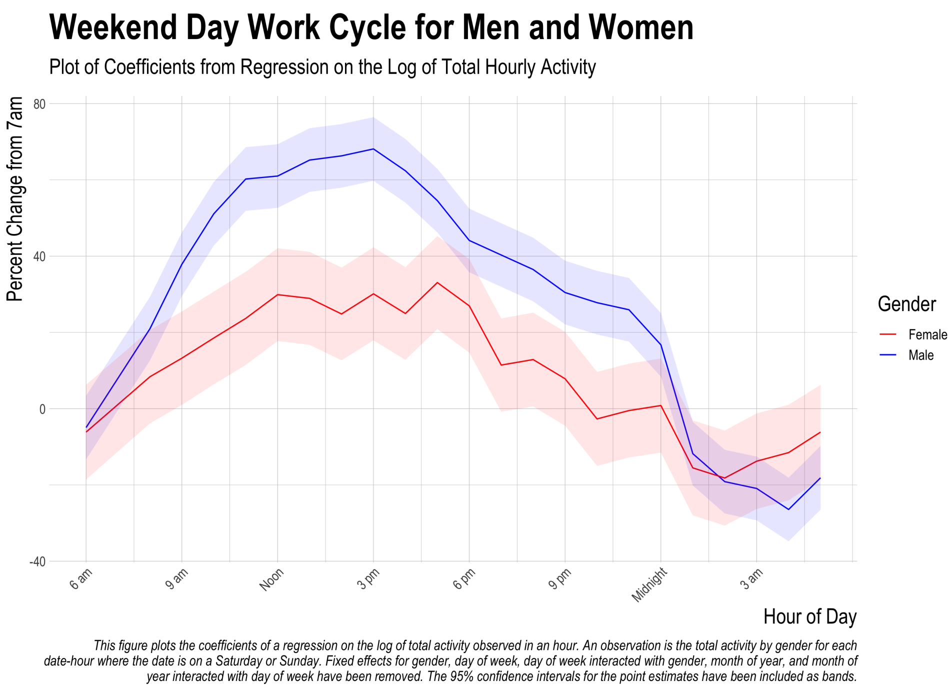
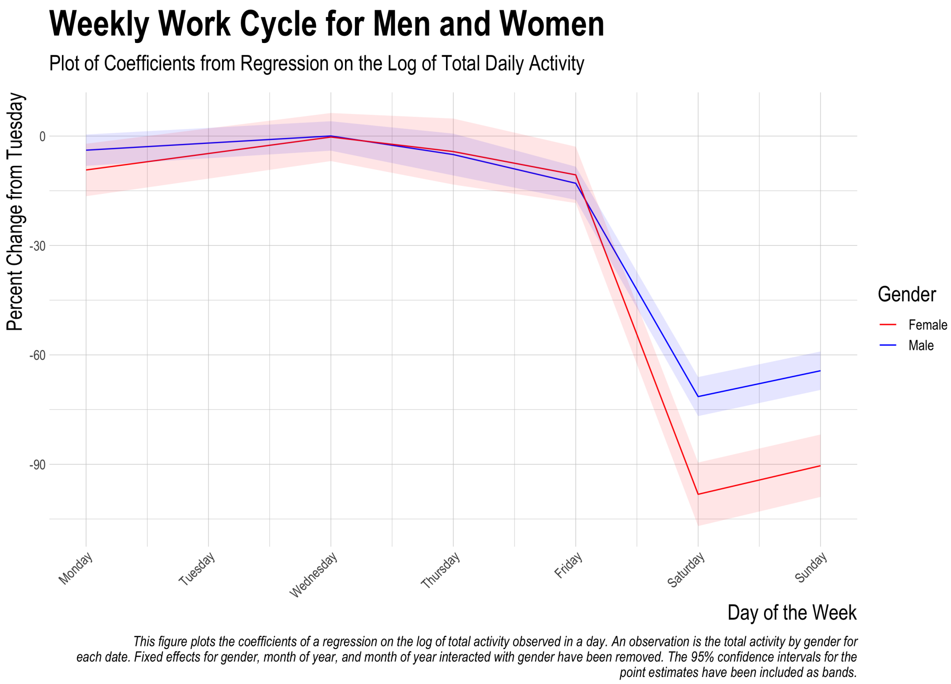


FIGURE 2.3. **Work Activity by Gender Across the Week**



2.4.2.1. Work Week

During the week, men and women both have the most activity during traditional 9am to 5pm work hours. However, the difference between work during this time and other hours is much more pronounced for women. The precise estimates for these differences are reported in Table 2.3. Between 6pm and 2am, female work activity declines more than male work activity for every hour, but the magnitude of the difference varies. The largest difference occurs between 6 and 7pm. The smallest difference occurs between 8 and 10pm. These details are shown in Figure 2.4.

TABLE 2.2. **Difference in Week Between Men and Women**

VARIABLES	(1)	(2)	(3)
Female × Monday	-0.054	-0.054	-0.055
	(0.078)	(0.074)	(0.074)
Wednesday	-0.003	-0.003	-0.003
	(0.074)	(0.071)	(0.071)
Thursday	0.008	0.008	0.007
	(0.079)	(0.076)	(0.076)
Friday	0.023	0.023	0.022
	(0.075)	(0.072)	(0.072)
Saturday	-0.268***	-0.268***	-0.270***
	(0.072)	(0.069)	(0.069)
Sunday	-0.260***	-0.260***	-0.262***
	(0.071)	(0.068)	(0.068)
Observations	2922	2922	2922
R-squared	0.841	0.853	0.853
Month	No	Yes	Yes
Female × Month	No	No	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table reports the coefficients from a regression of the log of activity. The reported coefficients above are the differential change for woman for the specified time period. E.g. Female activity decreases 27% more than male work activity on Saturday as compared to Tuesday. An observation is the total activity for a date for each gender. The reported errors are Newey-West with a two week lag. The coefficients are interpreted as percent changes in the amount of work from the baseline day of Tuesday. In all specifications, gender is controlled for.

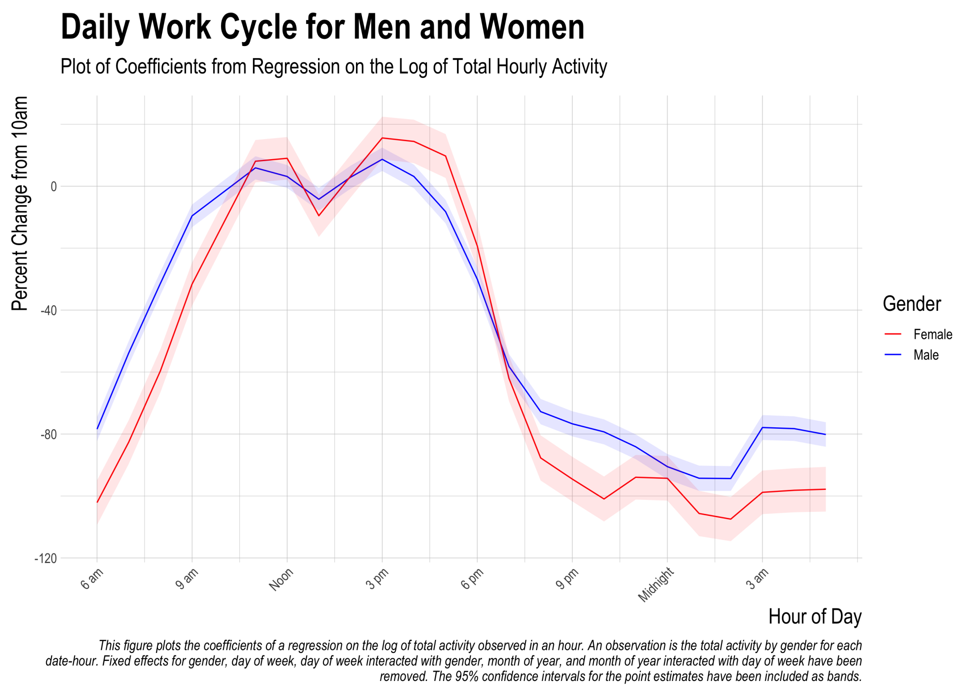
TABLE 2.3. **Difference in Workday Between Men and Women**

VARIABLES	(1)	(2)	(3)	(4)
Female × 2am to 8am	-0.272***	-0.272***	-0.272***	-0.272***
	(0.024)	(0.024)	(0.024)	(0.024)
9am to 3pm	0.019	0.019	0.019	0.019
	(0.021)	(0.021)	(0.021)	(0.021)
4pm to 5pm	0.109***	0.109***	0.109***	0.109***
	(0.027)	(0.027)	(0.027)	(0.027)
6pm to 7pm	-0.283***	-0.283***	-0.283***	-0.283***
	(0.032)	(0.032)	(0.032)	(0.032)
8pm to 10pm	-0.193***	-0.193***	-0.193***	-0.193***
	(0.033)	(0.033)	(0.033)	(0.033)
11pm to 2am	-0.226***	-0.226***	-0.226***	-0.226***
	(0.031)	(0.031)	(0.031)	(0.031)
Observations	50020	50020	50020	50020
R-squared	0.743	0.743	0.753	0.754
Day of Week	Yes	Yes	Yes	Yes
Female × Day of Week	No	Yes	Yes	Yes
Month	No	No	Yes	Yes
Day of Week × Month	No	No	No	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table reports the coefficients from a regression of the log of activity on the hour of the day. An observation is the total activity for a date-hour combination for each gender and includes observations from weekdays only. The reported errors are Newey-West with a two week lag. The coefficients are interpreted as percent changes from the baseline hour of 10am. In all specifications, gender is controlled for. The reported coefficients above are the differential change for women for the specified time period. E.g. Female activity decreases 19.3% more than male work activity during 6pm to 7pm as compared to 10am work activity.

FIGURE 2.4. **Work Activity by Gender on Weekdays**



During the traditional work day, women have less variable work activity. There is no statistically significant difference between men and women during 9am to 3pm. Women allocate their work to the 4pm to 5pm period of time 11% more than men do, but as illustrated in Figure 2.4, this is because male work decreases during these hours and female work does not.

In Section 2.5, I explore how child care responsibilities contribute to differences in male and female work activity. The large difference between 6pm and 7pm with the lessened difference between 8pm and 10pm is consistent with a story of women having non-work responsibilities in the household while children are awake and during traditional dinner preparation hours.

2.4.2.2. Weekends

On weekends, male work activity is allocated in a pattern that closely resembles the weekday work pattern. As shown in Figure 2.2, activity is highest at 3pm, which is the same as during the week. The sharpest increase in activity occurs between the hours of 8am and 9am, which is the same as during the work week as well. Female work activity on weekends is much flatter across the day, in contrast to the pattern during work days. As shown in Table 2.4, between 9am and 11pm, male work increases by 45% compared to the night hours. Female work increases by 19% less than this, or alternatively, 26%.

One explanation for this flat pattern for women on the weekends is that women are working less than men on these days. In Figure 2.3, I plot the coefficients of the regression of work activity by day of week. Activity by men and women decreases on Saturday and Sunday, but decreases more for users identified as female. As reported in Table 2.2, female work activity decreases by an additional 27% and 26% on Saturday and Sunday over the male decrease in activity. This is roughly equivalent to the magnitude of the difference between men and women at night.

TABLE 2.4. **Difference in Weekend Day Between Men and Women**

VARIABLES	(1)	(2)	(3)	(4)
Female × 9am to 11pm	-0.185***	-0.185***	-0.185***	-0.185***
	(0.046)	(0.046)	(0.046)	(0.046)
9am to 11pm	0.457***	0.457***	0.457***	0.457***
	(0.021)	(0.021)	(0.021)	(0.021)
Observations	20056	20056	20056	20056
R-squared	0.759	0.759	0.768	0.769
Day of Week	Yes	Yes	Yes	Yes
Female × Day of Week	No	Yes	Yes	Yes
Month	No	No	Yes	Yes
Day of Week × Month	No	No	No	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table reports the coefficients from a regression of the log of activity for each hour of the day. The reported coefficients above are the differential change for woman for the specified time period. E.g. Female activity decreases 18.5% more than male work activity from 9am to 11pm. The reported errors are Newey-West with a two week lag. An observation is the total activity for a date-hour combination for each gender. This table includes observations from Saturday and Sunday only. The coefficients are interpreted as percent changes from the baseline hour of 7am. In all specifications, gender is controlled for.

2.4.3. Conclusion on Work Patterns

Worker surveys show that women, and especially women with children, work fewer hours per week than non-mothers even when fully employed. Using observational data, the above sections show when in the work week men are working when women are not. Women work a more traditional Monday through Friday, 9am to 5pm week than men in this data.

2.5. Gendered Reactions to Childcare Shocks

There a variety of reasons why women may choose to schedule their work hours differently than men. Women may have a greater taste for leisure time on nights and weekends than men do. Women could also be engaged in informal labor that is time-sensitive. In this section, I consider childcare availability as a key explanation for this variation. Formal childcare is less available on nights and weekends, and so the comparative drop in activity on nights and weekends may be because women are taking care of children more than men are. In order to separate childcare from other mechanisms, I evaluate unexpected interruptions in the school calendar.

2.5.0.1. Identification Strategy

I utilize unexpected breaks in the school calendar in order to identify how men and women respond to childcare availability. I specifically look at school closures due to snow and other winter weather conditions. Snow days are helpful in two primary ways. First, snow closures are unexpected and so individuals are not able to plan for alternative childcare options in advance. Second, severe weather conditions leading to school closures typically cause a cancellation in daycare and can impact the ability of at-home childcare workers (nannies, babysitters, etc.) to commute to the individual's home or vice versa. This means that most childcare options will be unavailable to parents,

For my analysis, I use the following general specification, which compares the activity on a winter weather school closure day to activity on a normal day for both men and women.

$$Log(Activity_{tgd}) = \beta_0 + \theta Female_g + \sigma Snowday_{td} + \alpha Snowday_{td} \times Female_g + \sum_{d=1}^{7} \delta_d Day_t + \epsilon_{tgd}$$

Activity is a general term that refers to GitHub push and commit events as described in Section 2.2. This is the total amount of activity on date t for users of gender g on day of the week d. I include a fixed effect for users identified as female, $Female$, in order to control for baseline differences in activity levels. *Snowday* is a indicator variable for winter weather school closures. The attached coefficient σ is the impact of one of these school closures on the activity of the baseline group. The baseline group in the specification is the set of users identified as male. The coefficient α captures the differential effect of a school closure on female activity. I include fixed effects for the day of the week, δ_d.

Both α and σ are parameters of interest, as they capture the impact of a snowday on men and women. This paper is specifically interested in the differential impact of a childcare shock on women, α. I allow for errors with heteroskedasticity and auto-correlation by using Newey-West estimation and allowing for a two week lag.[6]

2.5.0.2. Results

Table 2.5 summarizes the results of this analysis. My preferred specification is shown in Column (4). The estimated impact of a school closure on female-identified users is consistent across the four columns, although statistical significance does vary. In all four specifications, the

[6]Estimations with robust standard errors produce similar results.

impact of a school closure on male users is not statistically significant. In my preferred specification, a snow day decreases the amount of activity by female-identified users by 34%. In my preferred specification, I include day of week, month of year, female interacted with day of week, and day of week interacted with month of year fixed effects.

TABLE 2.5. **Difference Between Men and Women in Public School Snowday Response**

VARIABLES	(1)	(2)	(3)	(4)
Female × Snowday	-0.310	-0.343*	-0.343***	-0.343***
	(0.192)	(0.192)	(0.115)	(0.112)
Snowday	0.115	0.132	0.008	0.010
	(0.136)	(0.136)	(0.084)	(0.084)
Observations	1032	1032	1032	1032
R-squared	0.917	0.917	0.971	0.974
Day of Week	Yes	Yes	Yes	Yes
Female × Day of Week	No	Yes	Yes	Yes
Month	No	No	Yes	Yes
Day of Week × Month	No	No	No	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table reports the coefficients from a regression of the log of activity. An observation is the total activity for a date for each gender. The reported errors are Newey-West with a two week lag. The coefficients are interpreted as percent changes from a non-snowday. In all specifications, gender is controlled for. The reported coefficients above are the differential change for woman on a snowday and the change overall.

2.5.1. Threats to Identification

In the following section, I more carefully explore two aspects of the childcare causality story I am considering.

It is possible that I am identifying a change in github activity, but that this activity is hobby- and not work-related. In order to evaluate this, I conduct my analysis again, but limit the analysis to the subset of activity that is explicitly associated with top tech companies. [7]

In the second part of this section, I proxy for parental likelihood by controlling for the estimated age of the users. In order to more carefully identify female users who are likely to have parental obligations, I consider the effect on women between the ages of thirty and forty. Women in this age group are more likely to have young children in their household ([21]). As described in Section 2.2, I connect LinkedIn and GitHub user information to construct a variable with the approximate age of these users.

2.5.1.1. Company-Owned Projects

In GitHub, I am able to identify if user activity is associated with a project that is owned by a major tech company. I conduct my main analysis again using the subsample of activity that is associated with these companies. This analysis is reported in Table 2.6. When we examine activity that is explicitly related to a top tech company, the magnitude of the decrease in activity by women is larger. In this subsample, the decrease in activity is 66%, whereas the larger sample shows a 34% decrease.

[7] These companies are as follows: Amazon, Comcast, Facebook, Google, Intel, IBM, Microsoft, Red Hat, and SAP.

TABLE 2.6. **Snowday Response, Explicitly Company-Related Activity**

VARIABLES	(1)	(2)	(3)	(4)
Female × Snowday	-0.541*	-0.666**	-0.666**	-0.666**
	(0.306)	(0.304)	(0.299)	(0.293)
Snowday	0.105	0.167	0.056	0.154
	(0.217)	(0.215)	(0.219)	(0.219)
Observations	914	914	914	914
R-squared	0.865	0.868	0.874	0.888
Day of Week	Yes	Yes	Yes	Yes
Female × Day of Week	No	Yes	Yes	Yes
Month	No	No	Yes	Yes
Day of Week × Month	No	No	No	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table reports the coefficients from a regression on the log of activity. An observation is the total activity for a date for each gender. The reported errors are Newey-West with a two week lag. This analysis only includes activity that is explicitly associated with a major Seattle technology company. The coefficients are interpreted as percent changes from a non-snowday. In all specifications, gender is controlled for. The reported coefficients above are the differential change for woman on a snowday and the change overall.

2.5.1.2. Estimated Age

In Table 2.7, I report the estimated effect of a school closure on female activity when considering women between the ages of 30 and 40. When an indicator variable for women between these ages is included, the overall effect for women loses statistical significance. Instead, the estimated effect on women between 30 and 40 is statistically significant and much larger at approximately -80%. This suggests that the decrease in work activity for women overall is being driven by an increase in childcare responsibilities, as women in this age group are more likely to have young children.

TABLE 2.7. **Snowday Response, Estimated Age Included**

VARIABLES	(1)	(2)	(3)	(4)
Female × Snowday	-0.377***	-0.126	-0.125	-0.125
	(0.146)	(0.145)	(0.145)	(0.140)
Female × Thirty to Forty × Snowday		-0.794***	-0.795***	-0.794***
		(0.226)	(0.226)	(0.213)
Observations	972	2834	2834	2834
R-squared	0.904	0.876	0.881	0.885
Day of Week	Yes	Yes	Yes	Yes
Female × Day of Week	Yes	Yes	Yes	Yes
Month	Yes	No	Yes	Yes
Day of Week × Month	No	No	No	Yes

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

This table reports the coefficients from a regression on the log of activity. An observation is the total activity for a date for each gender-age group. The reported errors are Newey-West with a two week lag. This analysis includes two age groups, individuals who are estimated to be between thirty and forty in one group and all others in the other. This age group is used to identify individuals who are likely to be parents of younger children. The coefficients are interpreted as percent changes from a non-snowday. In all specifications, gender is controlled for.

2.6. Robustness

2.6.0.1. Random Female Assignment to Male Coders and Placebo Snowdays

In the first robustness check, I remove all users identified as female from the sample. In this new sample, I randomly assign 11 % of users to a "female" group and evaluate the impact of the snow day school closures on work activity. I also construct a set of placebo school closures by moving the true school closures back one year in time. In this analyses, as reported in Table 2.9 and Table 2.8, I do not find equivalent statistical significance for our parameters of interest.

TABLE 2.8. **Snowday Response for "Placebo" Snowdays**

VARIABLES	(1)	(2)	(3)	(4)
Female × Placebo Snowday	0.063	0.066	0.066	0.066
	(0.191)	(0.192)	(0.115)	(0.112)
Placebo Snowday	0.196	0.195	0.114	0.125
	(0.135)	(0.136)	(0.084)	(0.084)
Observations	1032	1032	1032	1032
R-squared	0.917	0.917	0.971	0.974
Day of Week	Yes	Yes	Yes	Yes
Female × Day of Week	No	Yes	Yes	Yes
Month	No	No	Yes	Yes
Day of Week × Month	No	No	No	Yes

*** p<0.01, ** p<0.05, * p<0.1

I construct "placebo" snowdays by moving the snowday dates to the previous year. These are dates when Seattle Public Schools did not have public school closures. This table reports the coefficients from the regression which is run on the log of activity. An observation is the total activity for a date for each gender. The coefficients are interpreted as percent changes from a non-snowday. In all specifications, gender is controlled for. The reported coefficients above are the differential change for woman on a snowday and the change overall.

TABLE 2.9. **Snowday Response, "Placebo" Female**

VARIABLES	(1)	(2)	(3)	(4)
Placebo Female × Snowday	0.018	0.014	0.014	0.014
	(0.166)	(0.167)	(0.092)	(0.088)
Snowday	0.128	0.130	-0.029	-0.026
	(0.118)	(0.118)	(0.068)	(0.066)
Observations	1032	1032	1032	1032
R-squared	0.925	0.925	0.977	0.981
Day of Week	Yes	Yes	Yes	Yes
Female × Day of Week	No	Yes	Yes	Yes
Month	No	No	Yes	Yes
Day of Week × Month	No	No	No	Yes

*** p<0.01, ** p<0.05, * p<0.1

I construct "placebo" female observations by limiting the sample to only male users and randomly assigning 11% of these users to a new female group. An observation is the total activity for a date for each gender. This table reports the coefficients from the regression which is run on the log of activity. The coefficients are interpreted as percent changes from a non-snowday. In all specifications, gender is controlled for. The reported coefficients above are the differential change for woman on a snowday and the change overall.

2.6.0.2. Individual Analysis

In Table 2.10, I consider an alternate specification style. In previous analyses, the unit of observation is the aggregate activity for a given gender. In this table, the unit of observation is the individual user. By doing this, I am able to include individual fixed effects that account for differences in GitHub interaction style. By doing so, I can control for compositional changes in the users who are active on any given day. This is not my primary specification style as individual users are not typically observed at a high enough frequency to support this specification style.

TABLE 2.10. **Snowday Response, Individual Analysis**

VARIABLES	(1) *Aggregate*	(2) *Individual*	(3) *Individual 25% of Days*	(4) *Individual 50% of Days*
Female × Snowday	-0.343***	-0.072***	-0.173***	-0.214**
	(0.112)	(0.026)	(0.061)	(0.102)
Snowday	0.010	-0.028***	-0.027	-0.003
	(0.084)	(0.007)	(0.017)	(0.030)
Observations	1032	2.259e+06	913836	435504
R-squared	0.974	0.023	0.050	0.083
Day of Week	Yes	Yes	Yes	Yes
Female × Day of Week	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
Day of Week × Month	Yes	Yes	Yes	Yes
Number of Individuals		4,378	1,771	844

*** p<0.01, ** p<0.05, * p<0.1

This table reports the coefficients from a regression of the log of activity where the unit of observation is varied. In Column (1) the preferred specification in my analysis is reported where a unit of observation is the total activity on a date by gender. The reported errors are Newey-West with a two week lag. In Columns (2)-(4), a unit of observation is an individual's total activity on a date. In Column (2), all users in the Seattle Area with identifiable gender are included. In Column (3), I limit the analysis to users who are present in the data at least 25% of work days. In Column (3), I limit analysis to users who are present in the data at least 50% of days.

In Column (1), I repeat the preferred column from Table 2.5, the main specification of this paper. In Column (2), I consider the model with individual observations and fixed effects. In this column, all users are included. The decrease for women is statistically significant, but much smaller in magnitude. In this data, there are many users who interact with the platform very infrequently which causes the log specification to be less than ideal. In Column (3), I repeat this analysis, but only keep users who are present in the data at least 25% of work days. In Column (4), this is increased to 50% of work days. As I increase the frequency that users must be present in the data, we approach the magnitude of the decrease found in the aggregate analysis.

2.7. Conclusion

For women, there is an increase in happiness associated with having a family, and an increase in happiness associated with having a career [22]. Unfortunately, these two increases do not add together to lead to a happiness premium for women with families *and* careers [22]. In response to the demanding nature of raising children while working, we see women exit the labor force in their 30s and 40s, re-enter as children age, and postpone retirement [23]. For women who continue in the labor market, there is a motherhood penalty that has remained near constant in magnitude since 1986 [24].

In this work, I document that the female work week follows a more traditional pattern than the male work week for the tech workers in my sample. For women as compared to men, we see that work is concentrated between 9am to 5am and that work decreases more on the weekends. I use a natural experiment of unplanned public school closures due to winter weather to demonstrate that changes in childcare availability impact female work activity in my sample, but not male. This second analysis is consistent with the story that appears in the first half of the paper — women are not working when children are not in school, but men are.

In this work, I consider the impact of school closures on female work activity. The implications are potentially much broader. There are many situations in which households may be faced with an increase in childcare responsibilities. These circumstances may be idiosyncratic to the individual household, as in the case of a medical issue for example, and they may also appear during larger shocks. The results in this paper suggest that school closures during COVID19 are likely to impact female work activity much more than male activity. Statistics coming out of the United States Bureau of Labor and Statistics echo this. In September 2020, approximately 78,000 men and 617,000 women exited the labor force ([25]).

Broadly, life and parenthood are rife with shocks that demand increased household labor. The results from this paper suggest that these shocks will impact female workforce activity much more than male. Over a career, these differences may lead to broad differences in career trajectories and compensation.

CHAPTER III

FATALITIES AND GOVERNMENT TRANSFERS

This chapter is co-authored with Benjamin Hansen and Caroline Weber.

3.1. Introduction

A fundamental pretext of economics is that increased income should improve individual outcomes and welfare, holding all else constant. Despite this axiomatic assumption underlying all economic models (due to free disposal), only recently have experimental and quasi-experimental papers begun to explore how income affects labor supply, welfare, and health [26, 27, 28, 29].

While a variety of studies have concluded that generally increased income improves outcomes, other evidence has pointed to unseen nuances not captured in standard models of utility maximizing behavior. For instance, Ruhm [30] finds recessions, typically seen as avoidable and undesirable, actually led to fewer deaths in develop countries like the United States. Likewise, Evans and Moore [31, 32] find evidence mortality increases substantially on days when paychecks arise. Dobkin and Puller [33] find overdoses increase on the first of the month when disability payments arrive, and become more evenly distributed when income payments are distributed in a staggered manner. Thus despite the broadly accepted conclusion that income improves welfare, the receipt of income increase human activity which carries some level systemic risk. This particularly pronounced in populations with credit constraints where consumption smoothing is limited and the permanent income hypothesis appears to fail.

One population that is traditionally credit constrained are recipients of the Supplemental Nutrition Assistance Program (SNAP). Due to concerns about subsidizing drug use and abuse, some states test for drug use, while others ban SNAP receipt for those convicted of drug offenses. Moreover, while many studies find SNAP benefits largely increase spending on food, households have the potential to fiscally substitute between SNAP and other income. How does drug use respond to in-kind government transfers like SNAP?

This paper exploits the variation in SNAP receipt timing within the calendar month to identify the public health effects of government transfers. Although the federal government funds SNAP, each state is in charge of administering the program to its eligible residents. We use the variation in the calendar day(s) of benefit distribution between states, and variation in distribution regime across time within states, to identify the effect of SNAP transfers on drug and alcohol related fatal car accidents. Drug related fatal car accidents proxy for drug and alcohol use within the population.

A state may choose to distribute benefits to all recipients on one day of the month, or spread distribution out over multiple days. When distribution is spread over multiple days, each individual within the state receives all her benefits on one day, but that day is one of a subset of days that the state distributes benefits on. Considerable research documents that individuals near their budget constraint do not smooth income shocks ([34]; [35]; [36]; [37]). Instead, an individual rapidly expends income on immediate consumption ([35]; [38]; [37]). This spike in consumption allows us to identify the effect of the transfer. The variation in SNAP distribution date allows us to construct a control group.

The interaction between government transfers and drug use is of considerable interest to policymakers [39]. Fifteen states spent 1.3 million to screen Temporary Assistance for Needy Families (TANF) applicants for drug use in 2016 alone [40]. Screened applicants tested positive at rates between 0 and 2.14, depending on the state [40], which is lower than the national rate of 9.4 percent [41]. The positive rate of drug testing among TANF applications should not be misconstrued as the drug use rate, however, as applicants are aware that testing could be required, and have time to adjust their behavior accordingly. A benefit of our data is that individuals do not alter their behavior in response to the data collection. States are not currently allowed to test SNAP recipients for drug use, but the federal government is considering allowing them to [42].

Related literature. Our paper is related to a large literature in public, labor, and behavioral economics studying the effects of government transfer receipt and timing on household and individual behavior. For a thorough review, see [43] and [44]. Our work contributes to three strains: (1) evaluations of the fungibility of in-kind transfers, (2) exploration of intertemporal smoothing, and (3) documentation of SNAP externalities.

Nearly all empirical research shows the average SNAP household increases food expenditures more from SNAP receipt than if it received a cash transfer, but does not spend the entirety of the additional income on food. Various studies have estimated the marginal propensity to consume food (MPCF) out of SNAP income. Administrative records of SNAP transactions show a MPCF of .5 to .6 [45]. An evaluation of SNAP roll-out using survey data finds the SNAP MPCF is .16 to .32 with a confidence interval from .17 to .27 [46]. Additional analysis of survey data, using a SNAP expansion as variation, finds a MPCF of .53 to .64 [47]. Retail scanner data, augmented with method of payment data to identify SNAP consumers, finds a MPCF of .3 [48]. Notably, the SNAP MPCF are all strictly less than one – which suggests consumers are using the in-kind transfer to consume non-food goods. We explore a potential outlet for these remaining funds – drugs and alcohol.

We leverage the cyclicality of spending following income receipt for our identification strategy. Many studies document the decrease in food expenditures throughout the month ([49]; [50]; [34]; [51]), where the beginning of the month is associated with income receipt. For SNAP specifically, food spending peaks in the first three days after benefit receipt [37]. Of course, food expenditure is not the same as consumption, and it possible that consumers purchase storable goods at the beginning of the month while still smoothing consumption through the month. However, recent work has shown a decrease in food consumption over the course of the month as well ([35]; [38]; [37]).

The cyclicality of consumption is connected to a number of externalities. Test scores for children in SNAP households decrease near the end of the benefit cycle [52]. Crime increases at grocery stores at the end of the month [53] and over the course of the welfare benefit cycle [54]. There appear to be significant public health consequences as well. Hospital admissions for hypoglycemia increase at the end of the month for low-income individuals, but not high-income [55]. Hypoglycemia is associated with a lack of nutrition [56]. On the other side, the initial receipt of income increases intimate partner violence [57], which the author suggests may be due to an increase alcohol consumption.

Our work most closely relates to Dobkin and Puller [33], which documents the relationship between government transfer payments and hospitalizations for drug-related illness. Drug-related admissions increase by 23% in the first five days of the month, with a large component of this driven by Supplemental Security Income recipients. Our work differs in that we estimate a causal relationship by using exogenous changes in SNAP distribution dates over time. Additionally, our introduction of expenditure data allows us to identify the relationship between dollar amount and the resulting drug and alcohol related mortality.

Structure of paper. In Sections 3.2 and 3.3 we provide an overview of the SNAP program, its implementation in each state, and describe our SNAP and car crash data. Section 3.6 outlines our research design, including a detailed explanation of our identifying variation and econometric model. We provide our results in Section 3.7. We conclude in Section 3.8. An outline of the next steps in our research is presented in two parts: in the data section, specifically, Section 3.4 and Section 3.5 we describe the data components of our extensions. In Section 3.6.1, we describe the research designs.

3.2. Supplemental Nutrition Assistance Program

The Supplemental Nutrition Assistance Program (SNAP) provides low-income residents of the United States with food-purchasing assistance. Participants in the program receive a transfer of funds each month that can only be spent on food. Although the funds are not cash, they are redeemable for cash by supermarkets, convenience stores, and food retailers. The monthly amount varies for each household participating in the program, and is based on the number of individuals in the household, total income, and total expenses. The Federal Government pays for the program, and the Department of Agriculture Food and Nutrition Service Division oversees it. Each state is in charge of administering the distribution of SNAP benefits to its residents.

SNAP is the second largest government transfer program in the United States – only Medicaid is larger (Congressional Budget Office, 2013). 42 million individuals were enrolled in SNAP in 2017 [58], which was around 13% of the population at the time [59]. The average participant receives 125.51 per month in SNAP funds [58], and the total cost of the program was 68 billion in 2017 [58].

Federal food assistance has existed in some form in the United States since the Great Depression [60]. A pilot program, which eventually became the Food Stamp Program, was trialed between 1961 and 1964. The Food Stamp Act of 1964 made the pilot program permanent and extended it to every state [60]. The 2008 Farm Bill renamed the program the Supplemental Nutrition Assistance Program [61].

Distribution Schedules. Each state administers SNAP benefit distribution for its residents, although the benefit amount is set and paid for by the federal government. Seven states distribute benefits to every resident on the same day. The remaining states disburse benefits over multiple days. Although a state may distribute over multiple days, each individual within the state receives the entirety of her monthly benefits on one day. States that distribute over multiple days have a system of assigning participants to a distribution day. As such, states have chosen to assign the distribution date by SSN, birthday, last name, and case number. The last name, birthday, and SSN numbers allow for an additional layer of identification, which we discuss in Section 3.6.1. See Table 3.1 for a detailed listing of the date assignment scheme and distribution dates for each state in 2018.

Schedule Transitions. As mentioned above, each state selects the number of days to distribute SNAP benefits over. Within a state, this choice can vary over time. Sixteen states switch between distribution regimes, with three of those states switching more than once. When switching between distribution regimes, states have chosen to transition in one of three ways: a *simple transition*, a *50-50* transition, and a *smoothed transition*.
Consider some new distribution schedule, $Dist_{new}$, and an old distribution schedule, $Dist_{old}$.

TABLE 3.1. **SNAP Distribution Schedules and Schemes, 2017**

State	**Date Range**	**Scheme**
Alabama	4 - 23	case number
Alaska	1	-
Arizona	1 - 13	last name
Arkansas	4 - 13	SSN
California	1 - 10	case number
Colorado	1 - 10	SSN
Connecticut	1 - 3	last name
Delaware	2 - 23	last name
District of Columbia	1 - 10	last name
Florida	1 - 28	case number
Georgia	5 - 23	case number
Hawaii	3 - 5	last name
Idaho	1 - 10	birthday
Indiana	1 - 10	last name
Iowa	1 - 10	last name
Kansas	1 - 10	lastname
Kentucky	1 - 19	case number
Louisiana	5 - 14	SSN
Maine	10 - 14	birthday
Maryland	3 - 21	last name
Massachusetts	1 - 14	SSN
Michigan	3 - 21	case number
Minnesota	4 - 13	case number
Mississippi	4 - 21	case number
Missouri	1 - 22	last name - birthday
Montana	2 - 6	case number
Nebraska	1 - 5	SSN
Nevada	1	-
New Hampshire	5	-
New Jersey	1 - 5	case number
New Mexico	1 - 20	SSN
North Carolina	3 - 21	SSN
North Dakota	1	-
Ohio	2 - 20	case number
Oklahoma	1 - 10	case number
Oregon	2 - 20	SSN
Rhode Island	1	-
South Carolina	1 - 19	case number
South Dakota	10	-
Tennessee	1 - 20	SSN
Texas	1 - 15	case number
Utah	5 - 15	last name
Vermont	1	-
Virginia	1 - 9	case number
Washington	1 - 10	case number
West Virginia	1 - 9	last name
Wisconsin	2 - 15	SSN
Wyoming	1 - 4	last name

The above table is a list of the distribution schedules for each state in the United States. We have omitted Illinois, New York, Ohio, and Pennsylvania. These states await data confirmation. The *Date Range* refers to the first date of SNAP distribution and the last day. The *Scheme* is the method of assigning dates to SNAP recipients. "Case number" means that recipients are assigned a monthly SNAP receipt date based on their SNAP case number. "Last name" means the receipt date is based on the first letter (or first three letters) of the recipient's last name. "SSN" means the last digit of the recipient's Social Security Number is used and "birthday" means some aspect of the birthday (year, day, etc.) is used.

1. If a state uses a *simple* transition at some month t:

$$Dist_{t-1} = Dist_{old}$$
$$Dist_t = Dist_{new}$$

2. If a state uses a *50-50* transition at some month t:

$$Dist_{t-2} = Dist_{old}$$
$$Dist_{t-1} = .5Dist_{old} + .5Dist_{new}$$
$$Dist_t = Dist_{new}$$

3. If a state uses a *smoothed* transition at some month t:

   This state will smooth the transition over some period of months, k

$$\begin{aligned} Dist_{t-k} &= Dist_{old} \\ Dist_{t-k+1} &= Dist_{int}^{k+1} \\ &\vdots \\ Dist_{t-2} &= Dist_{int}^{2} \\ Dist_{t-1} &= Dist_{int}^{1} \\ Dist_t &= Dist_{new} \end{aligned}$$

   where $Dist_{int}$ is some distribution schedule with dates between $Dist_{old}$ and $Dist_{new}$, and each $Dist_{int}$ may or may not be the same. These switches provide additional identifying variation. Table 3.2 lists each switch we observe from 1998-2017.

TABLE 3.2. **Changes in Distribution Schedule, 1998-2017**

State	1998 Dates	2017 Dates	Transition
Alabama	4 - 18	4 - 23	50-50
Delaware*	5 - 11	2 - 23	simple
Florida	1 - 15	1 - 28	50-50
Georgia	5 - 14	5 - 23	50-50
Idaho*	1 - 5	1 - 10	simple
Indiana	1 - 10	1 - 23	50-50
Kentucky	1-10	1-19	smoothed
Maryland*	6 - 10	3 - 23	simple
Michigan	1 - 9	3 - 21	smoothed
Mississippi	5 - 19	4 - 21	simple
Montana	1	2 - 6	simple
North Carolina	3 - 12	3 - 21	simple
Oklahoma	1	1 - 10	smoothed
South Carolina	1 - 10	1 - 19	simple
Tennessee	1 - 10	1-20	simple
Virginia	1	1-9	smoothed

*State experienced more than one switch.

This table describes the transition between distribution schedules for each state that transitioned. It also lists the old date range and the new date range. The "range" lists the first date that SNAP benefits are distributed as well as the last date. A few states have multiple transitions. These states are denoted with a *. To date, each state with multiple transitions has chosen to transition in the same way, and so we choose to only list the state and its transition style once.

TABLE 3.3. **Specific SNAP Distribution Days, 2017**

State	Day of Month																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Alabama				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X					
Alaska	X																											
Arizona	X	X	X	X	X	X	X	X	X	X	X	X	X															
Arkansas				X	X			X	X	X	X	X	X															
California	X	X	X	X	X	X	X	X	X	X																		
Colorado	X	X	X	X	X	X	X	X	X	X																		
Connecticut	X	X	X																									
Delaware		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X					
D.C.	X	X	X	X	X	X	X	X	X	X																		
Florida	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Georgia					X		X		X		X		X		X		X		X		X		X					
Hawaii			X		X																							
Idaho	X	X	X	X	X	X	X	X	X	X																		
Indiana					X		X		X		X		X		X		X		X		X		X					
Iowa	X	X	X	X	X	X	X	X	X	X																		
Kansas	X	X	X	X	X	X	X	X	X	X																		
Kentucky	X		X		X		X		X		X		X		X		X		X									
Louisiana					X	X	X	X	X	X	X	X	X	X														
Maine										X	X	X	X	X														
Maryland				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X					
Massachusetts	X	X		X	X		X	X		X	X		X	X														
Michigan			X		X		X		X		X		X		X		X		X		X							
Minnesota				X	X	X	X	X	X	X	X	X	X															
Mississippi				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X							
Missouri	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X						
Montana		X	X	X	X	X																						
Nebraska	X	X	X	X	X																							
Nevada	X																											
N. Hampshire					X																							
N. Jersey	X	X	X	X	X																							
N. Mexico	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X								
N. Carolina			X		X		X		X		X		X		X		X		X		X							
N. Dakota	X																											
Oklahoma	X				X					X																		
Oregon	X	X	X	X	X	X	X	X	X																			
Rhode Island	X																											
S. Carolina		X		X		X		X		X		X	X		X		X		X		X							
S. Dakota	X																											
Tennessee	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X								
Texas	X		X		X	X	X		X		X	X	X		X													
Utah					X						X				X													
Vermont	X																											
Virginia	X			X			X		X																			
Washington	X	X	X	X	X	X	X	X	X	X																		
West Virginia	X	X	X	X	X	X	X	X	X																			
Wisconsin		X	X		X	X		X	X		X	X		X	X													
Wyoming	X	X	X	X																								

An "X" mark indicates a state distributes SNAP benefits on this day of the month. The vertical axis lists each state in the United States, with the exception of Illinois, New York, Ohio, and Pennsylvania, which we have omitted until data confirmation. The horizontal axis lists each day of the month that states distribute SNAP. No state distributes benefits later than the 28th.

Data Collection. We construct a novel dataset that details the distribution dates and date assignment scheme for each state and month for years 1998 - 2017. We construct our data using a time-series panel of distribution date ranges for the years 1998 - 2012 from the Economic Research Service of the United States Department of Agriculture (USDA). We augment this dataset with the specific days within in the range, see Table 3.3, the distribution scheme, and the transition method for distribution switches. We also carry the dataset forward to 2017.

These additions are added using the current schedules posted on the USDA Food and Nutrition Service "When Are Benefits Available?" webpage, https://www.fns.usda.gov/snap/snap-monthly-benefit-issuance-schedule, and historical versions of this webpage available at the Internet Archive, https://archive.org.

Program Cost and Scope. We use state-month SNAP expenditures and the count of individuals served for 1998-2017, as detailed in the SNAP National Data Bank Monthly State Participation and Benefit Surveys.[1]

3.3. Automobile Collisions

Our analysis focuses on the effect of SNAP distribution on the number of drug and alcohol related fatal car crashes. Ideally, we would measure drug and alcohol consumption for each individual who is treated and not treated with SNAP benefits. We could monitor alcohol and tobacco expenditures, but this ignores (1) the difference between expenditure and consumption, and (2) the vast black market for non-legal drugs. Drug and

[1] https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap

alcohol related car crashes are a consistent measure of the level of alcohol and drug consumption in a given community at a specific time.

We use Fatality Analysis Reporting System (FARS) data to measure the number of fatal crashes. FARS is a publicly available dataset maintained by the National Highway Traffic Safety Administration of the United States Department of Transportation. FARS records person, vehicle, and crash information for all fatal car crashes from 1975 to the present. This information includes if the driver of any vehicle involved in the car is under the influence of drugs or alcohol.

3.4. Medical Data

We have received approval from the Center for Disease Control's National Health Statistics Division (https://www.cdc.gov/nchs/nvss/index.htm) to extend our study to mortality at large using National Vital Statistics System (NVSS) data. NVSS micro-data and compressed vital statistics files contain the cause of death codes, exact date of death, and state of death. Using this data, in future work, we will extend our analysis to every death in the United States related to drug and alcohol use, and strengthen our identification by utilizing the variation in disbursement by last name and social security number (SSN).

3.5. SNAP Roll-Out

At the time of program introduction, SNAP was known as the Food Stamp Program (FSP). FSP began as a pilot program in 8 counties and eventually expanded to 43 counties during this pilot program period. The

Food Stamp Act of 1964 opened the program to all counties in the United States, but, crucially for our identification strategy, allowed counties to choose whether to participate [46]. In 1973, the Food Stamp Act was amended to mandate that all counties participate in FSP by 1975. We use a dataset of county-month FSP participation as provided in [44].

During this time-period, the NVSS micro-data is publicly available in pdf files. We are transcribing the pdf files into datasets that track the cause of death codes, exact date of death, and county of death for all fatalities in the United States during the FSP rollout period and the years preceding it. We will use this information in our future work.

3.6. Research Design

Identifying Variation. We use the variation in SNAP disbursement schedules between states and within states over time to identify the causal effect of SNAP income on fatal car crashes involving drugs or alcohol. We are able to use the exogenous variation in SNAP disbursement time, random across individuals, to approximate random assignment of SNAP treatment. For each state, we know the percent of SNAP disbursement that occurs on each day of the calendar month and the average SNAP income per person.

We use this information to estimate three effects: (1) the effect of distributing SNAP benefits on a day other than the 1st of the month, (2) the daily effect of SNAP disbursement on drug related fatal accidents, and (3) the effect of benefit generosity.

The majority of social services distribute benefit checks on the first of the month, and most paychecks are distributed near the first of the month

TABLE 3.4. **SNAP Distribution Date Based on SSN, 2017**

State	Day of Month						
	1	2	3	4	5	6	7
Arkansas				0,1	2,3		
Colorado	1	2	3	4	5	6	7
Louisiana					0	1	2
Massachusetts	0	1		2	3		4
Nebraska	1,2	3,4	5,6	7,8	9,0		
New Mexico	11,31,51,71,91	01,21,41,61,81	12,32,52,72,92	02,22,42,62,82	13,33,53,73,93	03,23,43,63,83	
North Carolina			1		2		3
Oregon	0,1	2	3	4	5	6	7
Tennessee	00-04	05-09	10-14	15-19	20-24	25-29	30-34
Wisconsin		0	1		2	3	
	8	9	10	11	12	13	14
Arkansas	4	5	6	7	8	9	
Colorado	8	9	0				
Louisiana	3	4	5	6	7	8	9
Massachusetts	5		6	7	8	9	
Nebraska							
New Mexico	14,34,54,74,94	04,24,64,84	15,35,55,75,95	05,25,45,65,85	16,36,56,76,96	06,26,46,66,86	17,37,57,77,97
North Carolina		4		5		6	
Oregon	8	9					
Tennessee	35-39	40-44	45-49	50-54	55-59	60-64	
Wisconsin	4	5		6	7	8	9
	15	16	17	18	19	20	21
Arkansas							
Colorado							
Louisiana							
Massachusetts							
Nebraska							
New Mexico	07,27,47,67,87	18,38,58,78,98	08,28,48,68,88	19,39,59,79,99	09,29,49,69,89	10,30,50,70,90	00,20,40,60,80
North Carolina	7		8		9		0
Oregon							
Tennessee	65-69	70-74	75-79	80-84	85-89	90-94	95-99
Wisconsin							

Some states choose to assign SNAP distribution day based on the recipient's Social Security Number. We have listed these states as well as the distribution scheme they use. The horizontal axis is each day of the month that states from this group distribute SNAP benefits. The numbers listed below the day are the SNAP recipients who receive their benefits on that day for that state. States with one number assign based on the last digit. States with two numbers assign based on the last two digits.

TABLE 3.5. **SNAP Distribution Date Based on Last Name First Letter, 2017**

State	**Day of Month**							
	1	2	3	4	5	6	7	8
Arizona	A, B	C,D	E,F	G,H	I,J	K,L	M,N	O,P
Connecticut	A-F	G-N	O-Z					
Delaware		A	B	C	D	E	F	G
D.C.	A,B	C	D,E,F	G,H	I,J,K	L,M	N,O,P,Q	R,S
Hawaii			A-I		J-Z			
Indiana					A,B		C,D	
Iowa	A,B	C,D	E,F,G	H,I	J,K,L	M,N,O	P,Q,R	S
Kansas	A,B	C,D	E,F,G	H,I,J	K,L	M	N,O,P,Q,R	S
Maryland				AAA-BAO	BAP-BQZ	BRA-CAQ	CAR-COQ	COR-DIZ
Missouri	Jan., A-K	Jan., L-Z	Feb., A-K	Feb., L-Z	Mar., A-K	Mar., L-Z	Apr., A-Z	May, A-K
Utah					A-G			
West Virginia	B,X,Y,Z	C,F	H,N,V	I,M,O,U	Q,S	A,W	J,K,P	D,E,R
Wyoming	A-D	E-K	L-R	S-Z				
	9	10	11	12	13	14	15	16
Arizona	Q,R	S,T	U,V	W,X	Y,Z			
Connecticut								
Delaware H	I	J	K	L	M	N	O	
D.C.	T,U,V	W,X,Y,Z						
Hawaii								
Indiana	E,F,G		H,I		J,K,L		M,N	
Iowa T,U,V	W,X,Y,Z							
Kansas	T,U,V	W,X,Y,Z						
Maryland	DJA-FIS	FIT-GON	GOO-HAX	HAY-JAB	JAC-KIM	KIN-LOX	LOY-MCO	MCP-NEF
Missouri	May, L-Z	Jun., A-K	Jun., L-Z	Jul., A-K	Jul., L-Z	Aug., A-K	Aug., L-Z	Sep., A-K
Utah			H-O			P-Z		
West Virginia	G,L,T							
Wyoming								
	17	18	19	20	21	22	23	
Arizona								
Connecticut								
Delaware	P	Q,R	S	T	U,V	W	X,Y,Z	
District of Columbia								
Hawaii								
Indiana	O,P,Q,R		S		T,U,V		W,X,Y,Z	
Iowa								
Kansas								
Maryland	NEG-PGZ	PHA-RIC	RID-SDZ	SEZ-STC	STD-TRA	TRB-WES	WET-ZZZ	
Missouri	Sep., L-Z	Oct., A-K	Oct., L-Z	Nov., A-K	Nov., L-Z	Dec., A-Z		
Utah								
West Virginia								
Wyoming								

Some states choose to assign SNAP distribution day based on the first letter of the recipient's last name. We have listed these states as well as the letter distribution scheme they use. The horizontal axis is each day of the month that states from this group distribute SNAP benefits. The letters listed below the day are the SNAP recipients who receive their benefits on that day for that state. Missouri uses a combination of the recipient's birth month and the last name first letter.

as well. Following in the work of [53], we examine the impact of adding an additional income shock to the monthly calendar. This occurs when a state chooses to move SNAP disbursement from the first of the month, where it would be grouped with other income, to its own day in the month.

We then consider the impact of SNAP disbursement on the daily amount of drug and alcohol related traffic fatalities. We construct construct a "weighted" treatment variable for each day-state combination over time, that describes the percent of SNAP distributed each day in every state. We compare the number of fatal accidents involving drugs and alcohol between states, for each day of the month, as a function of the SNAP distribution that day for the state. The identifying assumption is that in the absence of SNAP disbursement, the trend in the number of drug and alcohol related fatal car accidents would be parallel in all states. We relax this assumption by including fixed effects, which are detailed in the following section.

Our last analysis examines the impact of benefit generosity on the number of these fatalities.

Econometric Models.

Our first estimation considers the difference between total monthly traffic fatalities related to drug and alcohol use in states that distribute SNAP on the first of the month, and states that distribute SNAP away from the first of the month. This classification can vary within states across time. Consider the following estimating equation,

$$crash_{smy} = \beta_1 multiple_{sdmy} + \alpha_{sy} + \epsilon_{sdmy}$$

where $crash$ is the number of drug and alcohol related car crashes in a state s in month m during year y. $Multiple$ is an indicator variable for distribution away from the first of the month, in reference to the multiple income shocks faced by the individuals. We include state-year fixed effects.

Our following two estimations utilize a generalized differences-in-differences estimation strategy. The first estimation is described by the following,

$$crash_{sdmy} = \beta_1 percent_{sdmy} + \alpha_{smy} + \lambda_d + \epsilon_{sdmy}$$

where $crash_{sdmy}$ is the number of drug and alcohol related car crashes in a state s on day d in month m during year y. Our explanatory variable of interest is $percent_{sdmy}$, which is the percent of SNAP benefits distributed on that day. We include state-month-year fixed effects, α_{smy} and day of week fixed effects λ_d.

The second estimation is described by the following,

$$crash_{sdmy} = \beta_1 benefit_{sdmy} + \alpha_{smy} + \lambda_d + \epsilon_{sdmy}$$

where $benefit$ is the average dollar amount distributed to a SNAP participant.

Our dependent variable, the number of fatal accidents involving drugs and alcohol, is a non-negative count variable and therefore we use a Poisson regression model and assume $E(Y|X) = exp(X'\beta)$. We thus reform our estimating equations into the following log-likelihood function,

$$\ln L(\beta) = \sum_{sdmy=1}^{N} \{crash_{sdmy}\mathbf{x}_i'\beta - exp(\mathbf{x}_i'\beta) - \ln crash_{sdmy}!\}$$

For all estimations, we allow the standard errors to cluster at the state level, which has the additional effect of relaxing the Poisson model assumption of equality between the mean and variance.[2] The coefficients in the Poisson regression can be interpreted as semi-elasticities, or how a 1 one unit change in our independent variables predict a percentage change in the count of fatal car crashes involving drugs and alcohol.

3.6.1. Future Work

National Vital Statistics System. We use the same variation to identify the effect of SNAP disbursement on drug and alcohol related mortality at large. NVSS includes patient level data that allows us to add an additional layer of identification. A subset of states distribute SNAP benefits based on the last name, the birth date, or social security number (SSN) of the recipient. This allows us to identify a more narrow group of potentially treated individuals for our analysis.

Consider the following equation,

$$death_{sdmyi} = \beta_1 treat_{sdmyi} + \alpha_{smy} + \lambda_d + \epsilon_{sdmyi}$$

[2] We consider two-way clustering for state and year, but find similar results. Our panel extends from 1998 to 2017, which provides fewer year clusters than ideal and so we select year clustering in our preferred regression

where $death_{sdmyi}$ is the count of drug and alcohol fatalities in state s on day d in month m for year y , for individuals that have identifier i. The identifier can be the first letter of the last name, the last digit of the SSN, or some aspect of the birth date of the recipient - whichever method the state uses to select the day of disbursement for residents. Table 3.5 and 3.4 detail how states chose to distribute SNAP across last names and SSN numbers on 2017. The estimating equation is the same as the car crash analysis except our key explanatory variable is now $treat_{sdmyi}$ instead of $SNAP_{sdmy}$, where $treat$ is an indicator variable for individuals who would receive all of their benefits on that day if eligible for SNAP.

As before, we are working with count data, and use a Poisson approach.

$$\ln L(\beta) = \sum_{sdmyi=1}^{N} \{death_{sdmyi}\mathbf{x}_i'\beta - exp(\mathbf{x}_i'\beta) - \ln death_{sdmyi}!\}$$

Introduction of SNAP. Using the county level rollout of SNAP, we examine the total number of drug and alcohol related fatalities per month in SNAP-participating counties compared to non-SNAP-participating counties.

Consider the following equation,

$$death_{cmy} = \beta_1 treat_{cmy} + \alpha_c + \lambda_y + \epsilon_{cmy}$$

where $death_{cmy}$ is the number of deaths in county c in month m and year y. Our coefficient of interest is attached to $treat_{cmy}$ which indicates if a county is participating in SNAP or not. We control for county and year fixed effects.

3.7. Results

Distribution of SNAP on a day other than the first of the month leads to a 1.21 percent increase in the number of drug and alcohol related automobile fatalities per month. This point estimate is included in Table 3.6. Our preferred specification, Column(1), includes state-year fixed effects. The additional specifications include different fixed effects.

TABLE 3.6. **Drug and Alcohol Related Fatal Car Crashes and Multiple Day SNAP Disbursement**

	(1)	(2)	(3)
Multiple	1.26***	.155***	
	(.474)	(.059)	
Observations	10,902	10,902	
Fixed Effects			
State		X	
Year		X	
State-Year	X		

These are the results of our estimation of the effect of SNAP benefit disbursement on multiple days on drug and alcohol related fatal car accidents. *Multiple* is an indicator variable for distributions that occur on more than one days of the month. We estimate the effect using a Poisson distribution. The coefficients in the Poisson regression can be interpreted as semi-elasticities, or how distributing SNAP on more than one day predicts a percentage change in the count of fatal car crashes involving drugs and alcohol. We include combinations of fixed effects. An “X” indicates this set of fixed effects was included. Our preferred specification is Column (1).

We find a one percentage point increase in the share of SNAP benefits distributed in a state on a day leads to a .11 percent increase in the number of car crashes involving drugs and alcohol in a state on the distribution day. Table 3.7 shows this point estimate and its confidence interval and the point estimates and confidence intervals for three alternate specifications. These additional specifications include different fixed effects. In our preferred specification, Column (1), we include state-year-month fixed effects. These fixed effects relax our identifying assumption of parallel trends in drug and alcohol related fatal car accidents across states across time, to parallel trends within a a specific year-month combination. Column (2) includes day of week and state-year fixed effects. In Column (3), we consider year fixed effects and state fixed effects. In Column (4), we only include day of week effects. The point estimate is steady throughout the alternate specifications, with the exception of Column (4), which does not control for time-invariant differences in unobservables across states.

TABLE 3.7. **Drug and Alcohol Related Fatal Car Crashes and SNAP Disbursement**

	(1)	(2)	(3)	(4)
Percent	.112***	.112***	.113 ***	.070**
	(.037)	(.038)	(.038)	(.035)
Observations	331,798	331,798	331,798	331,798
Fixed Effects				
Day of Week	X	X	X	X
State			X	
Year			X	
State-Year		X		
State-Year-Month	X			

These are the results of our estimation of the effect of SNAP benefit disbursement on drug and alcohol related fatal car accidents. *Percent* is the percent of SNAP benefits distributed on a day. We estimate the effect using a Poisson distribution. The coefficients in the Poisson regression can be interpreted as semi-elasticities, or how a 1 percentage point change in the amount of SNAP distributed in a day predicts a percentage change in the count of fatal car crashes involving drugs and alcohol. We include combinations of fixed effects. An "X" indicates this set of fixed effects was included. Our preferred specification in Column (1).

Table 3.8 holds the same information for our estimation of the effect of benefit generosity. We find a one hundred dollar increase in benefit generosity leads to a .06 percent increase in drug and alcohol related automobile fatalities.

TABLE 3.8. **Drug and Alcohol Related Fatal Car Crashes and SNAP Disbursement Amount**

	(1)	(2)	(3)	(4)
Benefit	.00055*	.00058	.00089 ***	-.0041 ***
	(.000)	(.000)	(.000)	(.000)
Observations	331,767	331,767	331,767	331,767
Fixed Effects				
Day of Week	X	X	X	X
State			X	
Year			X	
State-Year		X		
State-Year-Month	X			

These are the results of our estimation of the effect of SNAP benefit amount on drug and alcohol related fatal car accidents. *Benefit* is the per person benefit amount distributed on a given day. We estimate the effect using a Poisson distribution. The coefficients in the Poisson regression can be interpreted as semi-elasticities, or how a 1 dollar increase in the amount of SNAP benefits distributed on a day predicts a percentage change in the count of fatal car crashes involving drugs and alcohol. We include combinations of fixed effects. An "X" indicates this set of fixed effects was included. Our preferred specification is Column (1).

3.8. Conclusions

Income is something that increases welfare theoretically in essentially every model of economics. Moreover, empirical evidence largely bears this out, as non-labor income allows individuals to consume more leisure, increases in income increase birth weights, reduces stress, and improves health [26, 27, 28, 62]. However, the activity increased income brings also increases systemic risk particularly for external causes of injury and death [30, 32**?**]. Despite an extensive literature documenting the benefits of SNAP, it's possible and consistent with other evidence on the timing of income receipt, that SNAP could influence some external risks. Moreover, based on the extent to which SNAP is fungible with income, and SNAP receipt allows credit constraint individuals to increase their consumption of non-SNAP goods, it's feasible some of the patterns of income timing and drug or alcohol use would also manifest themselves with SNAP.

This paper uses SNAP and the variation in SNAP benefit disbursement across states and time to identify the causal effect of government transfers on drug and alcohol related car crashes. Distributing SNAP benefits on days other than the first of the month, adding an additional income shock to the monthly calendar, increases the number of drug-related fatal automobile collisions by 1.21 percent. A one-percentage point increase in the share of SNAP benefits distributed on a day leads to a .2 percent increase in the number of drug-related fatal automobile collisions. A one hundred dollar increase in benefit generosity leads to a .06 percent increase in drug and alcohol related automobile fatalities. Future research should more fully investigate potential channels, impacts on drug overdoses more generally, and compare the differences in how income increases map into drug use versus in-kind transfers like SNAP.

CHAPTER IV

GETTING INTO THE WEEDS OF TAX INVARIANCE

This chapter is co-authored with Benjamin Hansen, Keaton Miller, and Caroline Weber.

4.1. Introduction

Tax invariance (TIV)—the principle that who remits taxes does not influence incidence—is a bedrock principle of tax design. TIV allows policymakers to focus on minimizing administrative and evasion costs without worrying about the welfare effects of alternative tax collection strategies. TIV is routinely taught in "Principles of Economics" courses [63, 64]. While recent empirical work suggests that TIV can fail under specific circumstances—when tax evasion opportunities vary along the supply chain [65, 66, 67], when there are price rigidities [68, 69, 70], or if tax salience is different for consumers and firms [71, 72]—it is unclear whether TIV simply does not hold, or just that it cannot be applied in particular settings.

We provide a more general test of TIV than has previously been possible by studying the cannabis market in Washington state.[1] The frequently-audited comprehensive regulatory reporting system makes tax evasion difficult. Prices both increase and decrease often, which means rigidities are unlikely. Tax salience is likely high for manufacturers, retailers, and consumers. Regulatory requirements ensure that owners are highly-

[1] We describe the market in Section 4.2.

skilled and well-capitalized. The posted retail prices include all taxes, so tax-inclusive prices are likely salient to consumers. Finally, tax leakage and competition are not relevant as the market is closed. Each gram of cannabis purchased in Washington was grown in Washington, and vice versa, and neighboring states did not have legal cannabis markets at the time.

We study an ideal reform for testing TIV. Prior to July 1, 2015, a 25% gross receipts tax applied to each transfer of cannabis. Cultivators remitted the tax when they sold to manufacturers, manufacturers remitted the tax when they sold to retailers, and retailers remitted the tax when they sold to consumers. The retail tax was required to be included in the posted price making it functionally equivalent to other excise and sales taxes. After the reform, the retail tax was increased to 37% and all other taxes were eliminated. Crucially, this change was unexpected by market participants; the reform was passed on June 27, 2015, and signed by the Governor on June 30 [73].

We measure the effects of this reform using an interrupted time series regression in first differences; that is, we ask how prices change in the week of the reform relative to weeks surrounding the reform. Identification rests on the assumption that, after controlling for product characteristics, prices would not have changed in the week of the reform (relative to a baseline trend) in the reform's absence. We conduct event study and placebo analyses which provide no evidence to reject this assumption. We employ this approach rather than a difference-in-differences design as the only potential comparison state is Colorado, which had a significantly different regulatory and industry structure—the assumption that prices in the two states co-move in the period of the reform is likely much stronger than the assumptions we impose.

Our setting features imperfect competition—retailers have substantial market power [74, 75] and manufacturers, while more competitive, retain market power too (see Table 4.1). Given the the emphasis on imperfect competition in tax incidence analysis [76], we examine how TIV predictions vary for a percent-based tax—the relevant tax in our setting—depending on the level of competition. We compare two extremes: perfect competition and a monopolist retailer and monopolist manufacturer. We show that manufacturers pass along their entire savings in response to the elimination of their tax in both situations. Under perfect competition, retailers leave their tax-inclusive prices unchanged. Under monopoly, retailers cut their prices to maximize profits under the new system. Our setting lies between these extremes.

We then examine how manufacturer prices change post-reform. Our framework predicts that manufacturers' prices should decrease 28.7% from pre-reform levels. Given that per-gram tax revenue would fall slightly in that scenario,[2] we also consider a second benchmark, the amount manufacturers needed to pass-through to leave retailers' per-gram profits and consumer-facing tax-inclusive prices constant post-reform (17.7%). We find that manufacturers reduce their prices by only 7.2%; we reject the null hypothesis of TIV based on either benchmark at the 0.1 percent level.

Finally, we examine retail behavior. Our framework predicts that retailers should either leave their tax-inclusive prices constant or decrease

[2] Revenue would have remained roughly constant if tax-exclusive prices remained constant—i.e. retailers had passed along their entire tax increase to consumers.

them. Instead, we find tax-inclusive retail prices *increased* by an average of 2.5%. Retailers pass through one-third of the tax increase to consumers. Another roughly one-third is borne by manufacturers, leaving retailers to bear about one-third of the increase. We find evidence that retailers maintained constant tax-exclusive markups, consistent with our model's pricing rule.

In summary, we find that TIV fails. A reform that should have left the welfare of manufacturers, retailers, and consumers unchanged or improved instead increased the profits of manufacturers at the expense of retailers and consumers. We conclude by discussing potential mechanisms for this result and implications for policymakers and future research.

4.2. Background

Our analysis focuses on the adult-use cannabis market in Washington state, which opened in July 2014 after a successful ballot initiative in 2012. We have written elsewhere about the history of this market [77, 78]—here we focus on features of the market and the reform that underlie our analysis.

Washington's cannabis market consists of three types of firms: cultivators, who grow cannabis plants, manufacturers, who transform plant material into marijuana products, and retailers, who sell products they obtain from manufacturers to consumers. Potential entrants have to pass background checks and undergo a lengthy regulatory process requiring substantial capital investment before entry. Cultivators face capacity constraints—the largest firms may cultivate 30,000 sq. ft. of plant canopy and may not merge to increase capacity. While retailers must be financially independent from other firms, a cultivator and a manufacturer may vertically integrate, though the capacity constraint remains. When the reform was implemented, approximately 94% (by weight) of usable marijuana—dried and cured cannabis flowers—was produced through a vertically-integrated process [79]. Thus, we focus our analyses on two types of firms, "manufacturers" and "retailers".[3]

The market features a closed supply: all cannabis sold by retailers is grown in the state, and every ounce grown legally within the state is sold at a Washington retailer. These rules are enforced through the state's "seed-to-sale" traceability system, which tracks each plant from cultivation through processing and retail. This system was implemented to respond to the informal federal regulations created by the "Cole Memo" [80]. The system provides information that can be used to check for tax evasion: retailers cannot sell cannabis without manufacturing records, which forces manufacturers to report accurately.[4] Reporting is enforced through frequent audits—firms typically face one or more visits per year—backed by significant penalties for non-compliance.

Washington's initial tax regime consisted of a 25% tax collected at every transfer of cannabis. Vertically-integrated manufacturers owed no tax on intra-firm transfers. The reform we analyze eliminated the 25% excise taxes within the supply chain and increased the retail excise tax rate to 37%. The excise tax applied to the sales-tax-inclusive price pre-reform and

[3] State law calls cultivators "producers" and manufacturers "processors"—we choose nomenclature to represent functional equivalents in other markets.

[4] Retailers can under-report their sales, but such behavior is detectable as retail sales can be compared to purchases from manufacturers. Our estimates are unaffected by dropping the few retailers that engage in significant under-reporting.

the sales-tax-exclusive price post-reform. Accounting for changes to the base and rate of the retail tax, the reform changed the retail tax rate by 6.93%.[5] This change was designed to be revenue neutral under the assumption that tax-*exclusive* prices remained constant (whereas TIV predicts constant tax-*inclusive* prices). We account for both the change in the rate and the base of the retail excise tax in our analyses. We provide calculations of revenue pre- and post-reform in Section 4.4. Other regulations concerning cannabis production, distribution, and sales were unaffected.

Our identification assumes that the policy change was unanticipated. The bill originated and was passed in the Washington House during the 2015 Regular Session, but stalled in the Senate. The bill was reintroduced in the First Special Session, but again stalled. Finally, on June 27, the last day of the Second Special Session, the bill passed both chambers. The Governor signed it on June 30 and the law went into effect the next day. Contemporaneous reporting portrayed the industry as unprepared. According to one retail manager, "[we] have a few hours to change an entire market's pricing structure. It is an exceptionally short window for such a tremendous change" [73].

4.3. A Framework for Tax Invariance

To motivate our empirical analyses, we introduce stylized models of manufacturers, retailers, and consumers. We assume a constant manufacturing marginal cost of mc. Given tax-inclusive retail price p, we

[5] The average sales tax rate during this period was 8.9%, thus $\log\left(\frac{1.25(1+.089)}{1.37+.089}\right) = -0.0693$

assume demand has constant price elasticity, that is $q(p) = kp^{\epsilon}$ with $\epsilon < -1$. We evaluate two extremes: (1) perfectly competitive manufacturers and perfectly competitive retailers and (2) a monopolist manufacturer and monopolist retailer. We expect our empirical setting lies between the two (see Section 4.4).

Let p_m^i be the price charged by manufacturers to retailers including all taxes. Let p_m^e be the price charged by manufacturers to retailers exclusive of taxes. Given a manufacturing tax rate of τ_m, $p_m^e = p_m^i(1 - \tau_m)$. Similarly, let p_r^i be the tax-inclusive retail price and let p_r^e be the tax-exclusive retail price. Given a retail tax rate of τ_r, $p_r^i = p_r^e(1 + \tau_r)$. While these definitions are not parallel, they match our empirical setting.

4.3.1.

Perfect Competition-Perfect Competition In perfect competition, the tax-exclusive price earned by manufacturers is equal to their marginal cost, and so the tax-inclusive price is $p_m^i = \frac{mc}{1-\tau_m}$. Perfectly competitive retailers face this price as their marginal cost, and so the tax-inclusive retail price is $p_r^i = mc\frac{1+\tau_r}{1-\tau_m}$. The total tax revenue collected is $TR = k\left(\frac{mc(1+\tau_r)}{1-\tau_m}\right)^{\epsilon}\frac{mc}{1-\tau_m}(\tau_r + \tau_m)$. To see TIV holds, define $\tau = \frac{1+\tau_r}{1-\tau_m}$. Then $p_r^i = mc \cdot \tau$ and $TR = k(mc)^{\epsilon+1}\tau^{\epsilon}(\tau - 1)$. Given some τ, a policy maker can freely move one of τ_r or τ_m, solve for the other, and hold p_r^i and TR constant.

4.3.2. Monopolist-Monopolist

Under monopoly, the retailer's profit maximization problem is $\max_{p^i_m}\left(\frac{p^i_r}{1+\tau_r}-p^i_m\right)k(p^i_r)^\epsilon$ which implies $p^i_r = \frac{\epsilon}{1+\epsilon}p^i_m(1+\tau_r)$. Note that the retailer's tax-exclusive price is a constant markup over their marginal cost p^i_m. The quantity is $q = b(p^i_m)^\epsilon$ where $b \equiv k\left[\frac{\epsilon}{1+\epsilon}(1+\tau_r)\right]^\epsilon$. The wholesaler's problem is $\max_{p^i_m}[p^i_m(1-\tau_m)-mc]b(p^i_m)^\epsilon$ which implies $p^i_m = \frac{\epsilon}{1+\epsilon}\frac{mc}{1-\tau_m}$. Thus, the tax-inclusive price charged by the manufacturer is independent of the retail tax and $p^r_i = \left(\frac{\epsilon}{\epsilon+1}\right)^2\frac{1+\tau_r}{1-\tau_m}mc$ and $TR = k\left(\frac{\epsilon}{\epsilon+1}\right)^{2\epsilon+2}\left(mc\frac{1+\tau_r}{1-\tau_m}\right)^{\epsilon+1}\left[\frac{\tau_r}{1+\tau_r}+\frac{\tau_m}{1+\tau_r}\frac{\epsilon+1}{\epsilon}\right]$. Mechanically, if τ is defined as above, the term in brackets cannot be simplified to be a function of τ alone. Given some p_r, if a policy maker shifts τ_r and τ_m to hold p_r constant, TR must change. Thus, TIV fails. Intuitively, the percentage taxes act as demand shifters as in Weyl and Fabinger [76], but the wholesaler does not internalize the retailer's response to retail percentage taxes because its effective demand elasticity is unchanged.

Given TIV generally does not hold in this monopolist-monopolist case, we want to understand the effect of a movement from a manufacturing tax to a retail tax. First, suppose that the policy $\omega_1 = \{\tau_r = 0, \tau_m = \tau\}$ is replaced with $\omega_2 = \{\frac{\tau}{1-\tau}, 0\}$. From the equations above, it is clear that p^i_r (and thus quantities) remains constant. The manufacturer passes through all of its tax savings, and earns identical profits. However, the retailer's profits decrease because the τ savings on the manufacturer's price is more than offset by the $\frac{\tau}{1-\tau}$ tax on their price. By the same logic, TR increases as $\frac{\varepsilon+1}{\varepsilon} < 1$. Now consider the policy $\omega_3 = \{\tau', 0\}$ where $\tau' = \frac{p^i_m(\omega_1)\tau}{p^i_r(\omega_1)-p^i_m(\omega_1)\tau}$ is "naive-revenue-neutral": it would raise the same total revenue *if* the tax-inclusive retail price was the same after the reform. In this case, since $\tau' < \frac{\tau}{1-\tau}$, $p_r^i(\omega_3) < p_r^i(\omega_1)$ and $q(\omega_3) > q(\omega_1)$. Since $\varepsilon < -1$, profits for both firms and total tax revenues increase. Finally, suppose ω_1 is replaced with $\omega_4 = \{\tau'', 0\}$ where τ'' is chosen to be "true-revenue-neutral": $TR(\omega_1) = TR(\omega_4)$. Since $TR(\omega_3) > TR(\omega_1)$, $\tau'' < \tau'$ and thus ω_4 increases profits for both the retailer and the manufacturer beyond ω_3. This is consistent with the notion that, under monopoly, ad valorem taxes improve welfare over unit taxes [81, 82, 83].

In summary, the combination of market power and percent taxes leads traditional TIV to fail. However, revenue-neutral policies (whether "naive" or "true") that shift taxes from manufacture to retail lead to full pass-through from the manufacturer to the retailer and a decrease in tax-inclusive retail prices. Firms and consumers benefit from the change.

4.4. Data and Methods

Our data consist of administrative records from the "traceability" (or seed-to-sale) system maintained by the Washington State Liquor and Cannabis Board (WSLCB). We obtain data on all plants, products, and sales. We restrict our analysis to "usable marijuana" products—74.5% of the total transactions observed in our data. Within this category, products are differentiated by "strain" (analogous to fruit cultivars), potency, and whether the marijuana is loose or pre-rolled into a joint. These characteristics are captured by our fixed effects.

Harvested flowers and other plant material are converted into an "inventory lot" that is assigned a unique identifier (ID). Products or material within a single inventory lot are assumed to be homogeneous. Large inventory lots of finished product are split into smaller "retail" lots for sale to retailers. Each retail lot consists of multiple sealed packages of a specific weight of cannabis (e.g. 1 gram, 3.5 grams, etc) which are considered identical. When lots are sold to retailers, the system records the date, quantity, and price, and assigns a new lot ID. Thus, retail lot IDs uniquely identify the retailer, manufacturer, cultivator, product, and package size.[6] We observe each retail sale and link the price, quantity, and transaction time to the relevant inventory lots.

We aggregate by inventory-lot-week. We exclude firms with less than two months of pre- and post-reform data. The reform also changed technical reporting rules which affect the price data. We clean the price data for each firm to reflect the prices faced by consumers using an algorithm based on rounding behavior verified by spot checks of historical menus.[7] See Appendix B for details.

Table 4.1 reports summary statistics for retail inventory lots for the six weeks pre-reform (the basis for our analyses in Section 4.5). The average tax-inclusive retail price was \$13.03 per gram and the tax-exclusive price was \$9.57 per gram. Retailer tax-exclusive prices are more than double manufacturer tax-inclusive prices. Both manufacturer and retail prices change week-over-week by more than one percent almost 40 percent of

[6] A small number of lots have multiple package sizes, which we identify and correct for.

[7] Cannabis retailers have limited access to financial services and so choose to set tax-inclusive prices that are round numbers (e.g. \$8 or \$10.25) to lower cash-handling costs. While this represents a potential friction, the effective minimum price change is smaller than the effects we estimate.

the time, split fairly evenly between price increases and price decreases, suggesting prices are not rigid.

The average market share of retailers in the 10-mile radius around their location was 31%, suggesting that there is substantial market power at retail, consistent with Hollenbeck and Uetake [74] and Mace et al. [75]. The manufacturer market is effectively state-wide and the average market share is 1.4%. No manufacturer has more than 7% of the total market share. However, manufacturers may exert market power on individual retailers—on average, about seven manufacturers supply 75% of a retailer's inventory.

Across competitive environments, our framework predicts that manufacturers should pass through their savings to retailers via a $\log(1 - 0.25) = -28.7\%$ decrease in manufacturer tax-inclusive prices. If we estimate a different price response to the reform, we can reject the framework. Since the reform was not quite revenue-neutral, it may be possible to construct alternative models which both rationalize any price responses we observe and which feature a TIV result. To rule out this concern, we construct a second benchmark for manufacturer price changes: given pre-reform prices, to maintain both a constant tax-inclusive retail price and constant per-gram retail profits (and therefore to satisfy TIV), manufacturers would have to decrease their prices by an average of 64 cents, or 17.7%.[8]

Under a revenue-neutral reform, TIV predicts that retailers would reduce their tax-*exclusive* prices by 6.93% (the amount of the change in the retail tax rate) and maintain constant tax-*inclusive* prices. Under monopoly, we predict tax-inclusive prices will decline. As we calculate the reform is

[8] 13.03/(1.37+0.089)-13.03/(1.25*(1+0.089)) = 64 cents.

TABLE 4.1. **Pre-Reform Retail Summary Statistics**

Variable	**Obs.**	**Mean**	**Std. Dev.**
Prices and Taxes			
Tax-Inclusive Retail Price ($/g)	63,668	13.033	3.798
Tax-Exclusive Retail Price ($/g)	63,668	9.570	2.783
Probability of > 1% Retail Price Increase	63,668	0.17	0.375
Probability of > 1% Retail Price Decrease	63,668	0.204	0.403
Manufacturer Price ($/g)	63,668	4.103	1.309
Probability of > 1% Manufacturer Price Increase†	7,954	0.177	0.382
Probability of > 1% Manufacturer Price Decrease†	7,954	0.196	0.397
Retail State + Local Sales Tax Rate	63,668	1.089	0.006
Tax Revenue Pre-Reform ($/g)	63,668	4.489	1.246
Competition			
Market Share of Retailer in 10 Mile Radius	63,668	0.313	0.282
Market-level Manufacturer Market Share	63,668	0.014	0.016
Retail-Level Manufacturer Concentration Index	63,668	6.997	2.691
Benchmarks Assuming TIV			
Expected Tax Revenue Post-Reform ($/g)	63,668	4.104	1.200
Manufacturer Pass-Through Cents	63,668	-0.640	0.185
Manufacturer Pass-Through Percent Change	63,668	-0.177	0.058

An observation is an inventory-lot-week pre-reform. The data come from our retail analysis set and cover the six weeks prior to the tax reform. Tax revenue is calculated using both excise and state and local sales taxes. The retail-level manufacturer concentration index is calculated as follows: for a given retailer, we sort their suppliers by the weight of inventory sold, and count the number needed to comprise at least 75% of total sales. The "benchmarks assuming TIV" account for changes in the base and rate of the retail excise tax. The "manufacturer pass-through" statistics assume constant tax-inclusive retail prices and indicate the post-reform changes to manufacturer prices that would have left retailer variable-profit-per-gram constant.

† These probabilities are calculated for the subset of retail-processor-strain-weight group-weeks when the inventory lot changes (and thus a new purchase from a manufacturer has occurred).

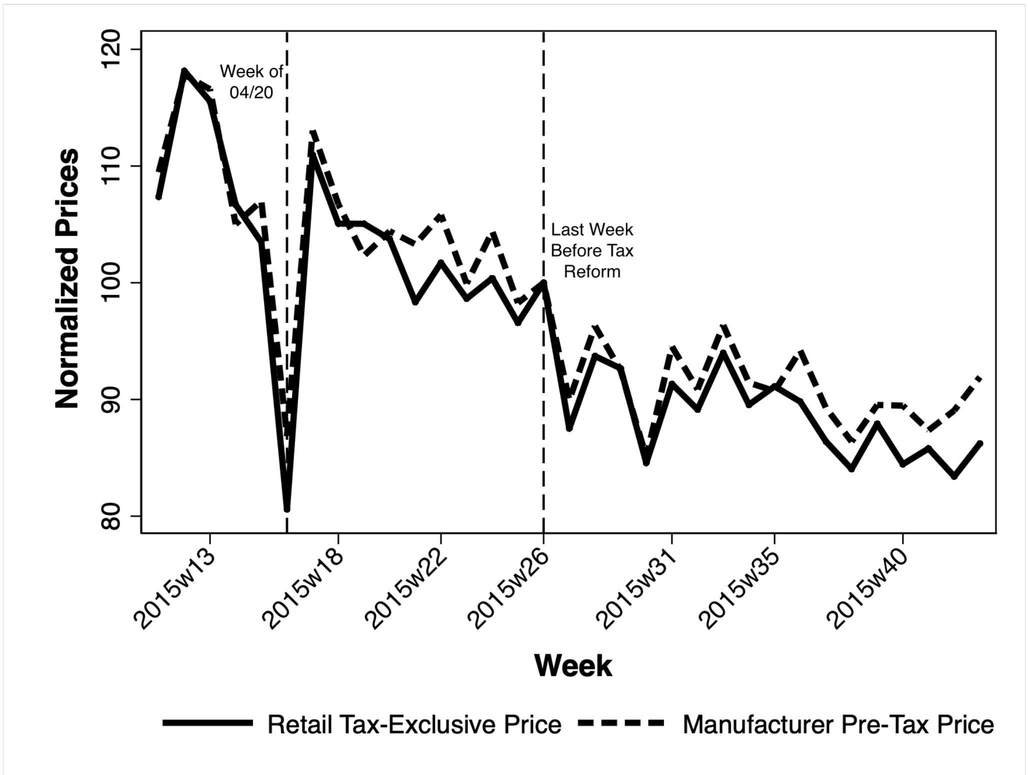
slightly revenue-decreasing,[9] our framework suggests retailers should reduce tax-inclusive prices further.

To summarize, if we estimate a decrease in average manufacturer tax-inclusive prices of less than 28.7% (in a statistically significant sense), we reject our model and reject TIV indirectly. If we estimate a decrease in average manufacturer tax-inclusive prices of less than 17.7%, we reject TIV directly. If we estimate any *increase* in retailer tax-inclusive prices, we reject TIV directly.

Figure 4.1 plots the panel of retail tax-exclusive prices normalized to the week before reform. For each week, we take inventory lots in their first week of sale and match them with the price paid to the manufacturer, restricting observations to those where the first retail sale and manufacturer sale both happened pre- or post-reform; thus, this illustrates the relation between retailer per-gram revenue and variable costs. The two series move in a highly correlated way through the entire pre- and post-reform period (including the period around April 20, an industry promotional event). This implies a constant markup of the retail tax-exclusive price over variable costs (the manufacturer price) that appears to be preserved in response to the tax reform. This figure depicts a set of products that is changing over time. To disentangle the effects of the reform from long-run trends and control for potential compositional changes, we employ regression and (in Section 4.5.3) event study analyses.

[9] If prices had remained constant, the reform would have decreased the average total tax revenue per gram from \$4.49 to \$4.10.

FIGURE 4.1. Retail and Manufacturer Prices



This figure plots average prices in Washington's cannabis industry for four months before and after the tax reform, normalized to 100 in the week before the reform. For each week, we take inventory lots in their first week of sale and match them with the price paid to the manufacturer, restricting observations to those for which the first retail sale and manufacturer sale both happened pre- or post-reform (before any applicable taxes are paid from the manufacturer to the government). This therefore illustrates the relation between retailer per-gram revenue and variable costs. The left dashed line in the figure marks 4/20 (an industry promotional period) and the right dashed line marks the week before the tax reform.

We model responses to the tax reform as an interrupted time-series in first differences:

$$\Delta log(p_{it}) = \alpha_0 + \alpha_1 \Delta TaxReform_t + \alpha_2 FE_i + u_{it}, \tag{4.1}$$

where i is described below, and t indicates the week. p is the wholesale or retail price per gram, $TaxReform$ is an indicator variable that is one after July 1, 2015 and zero before, and FE are fixed effects. α_1 is the parameter of interest.[10] Our analysis window spans six weeks before and after the reform—we examine the robustness of our estimates to this bandwidth. We two-way cluster standard errors on manufacturer and retail location [84].[11] Our identifying assumption is that within a given product, there are no shocks in the week of the reform that would have a significant and systematic impact on prices besides the direct effect of the tax reform. Given the short interval between observations (i.e a week, not a year), this assumption is plausible and we will provide placebo regression evidence that this assumption is reasonable.

For the manufacturer analysis, we aggregate to the manufacturer-retailer-strain-week level, so that i is a manufacturer-retailer-strain tuple, and then take first differences.[12] Each manufacturer-retailer-strain tuple does not sell every week. We thus calculate the minimal-length difference

[10] Without fixed effects, this regression is equivalent to an interrupted time series regression in levels with fixed effects at the level of our first differences and a control for time to the reform.

[11] Firm clusters or two-way clusters on firm and week yield similar standard errors.

[12] Aggregation beyond the inventory lot is required because each lot is sold only once. Other possible aggregations produce similar estimates with lower power (though statistical significance remains).

and include difference-length fixed effects.[13] The maximum difference-length allowed is 4 weeks. We are thus estimating the magnitude of price changes in response to the reform within a specific firm-product pairing. When we add retailer-manufacturer-strain fixed effects, we allow each retailer-manufacturer-strain to have a separate time trend.

For our main retail analysis, we aggregate to the inventory-lot-week level so that i is an inventory lot.[14] Retail sales from an inventory lot are frequent, so we construct one-week differences. We are thus estimating the change in the retail price of an inventory lot in response to the tax reform holding all possible product and firm variation constant. Sales of retail inventory lots typically last multiple weeks, so we include fixed effects for the week since the first week a particular inventory lot sold. When we add inventory lot fixed effects, we allow prices in each inventory lot to have a separate time trend.

We separately examine the first week of retail sales for each inventory lot and include only those that were purchased from manufacturers in the same week. Similar to our manufacturer analysis, we aggregate by retailer-manufacturer-strain and take varied differences. We include difference-length fixed effects. In these regressions, we ask how prices for *new inventory lots purchased post-reform* change relative to *pre-reform lots of the same strain from the same manufacturer*. This allows us to examine whether prices

[13] These fixed effects are not significant. Our estimates are similar when restricted to one-week differences, but with less power.

[14] We are able to work at this level because retailers repeatedly sell out of a single inventory lot.

change more or less if the inventory was purchased post-reform relative to inventory that had already been purchased and was selling pre-reform.

4.5. Results

4.5.1. Manufacturer Price Response

Table 4.2 reports estimates of Equation (4.1) for manufacturers. The estimate in Column (1), which includes no fixed effects, implies that prices changed by -6.5% in response to the tax reform (statistically significant at the 0.1% level). When we include manufacturer-retailer-strain fixed effects in Column (2)—our baseline specification—the point estimate becomes -7.2% (significant at the 0.1% level). This is roughly one-third of the 17.7% price decrease needed to preserve retailer per-gram profits (and therefore to minimially satisfy TIV), and one-quarter of the 28.7% decrease predicted by our framework. We can reject the null hypothesis that our estimate is consistent with TIV at the 0.1% level. Column (3) repeats Column (2) for the price in levels instead of logs—we find that the reform decreased manufacturer prices by 23 cents, about one-third of the 64 cent bound. Across specifications, the observed price adjustment was greater than 1% for more than 75% of our observations—suggesting firms were aware of this reform and prices are not rigid. Even if we rescaled our estimate assuming that any observation with minimal adjustment was caused by rigidities or lack of awareness, the data would still reject the null hypothesis of TIV.

The bottom panel of Table 4.2 repeats the specification of each column for a placebo reform dated one year later. The estimates are near zero across all four columns, providing support that our regression specifications are

TABLE 4.2. Manufacturer Price Response

	(1) Δlog(Price)	(2) Δlog(Price)	(3) ΔPrice
		Tax Reform	
ΔTax Reform	-0.065^{***}	-0.072^{***}	-0.228^{***}
	(0.015)	(0.018)	(0.068)
Observations	12,087	12,087	12,087
Manufacturer Firms	199	199	199
P-Value for Test of TIV-Predicted Pass-Through	0.000	0.000	0.000
		Placebo	
ΔPlacebo	0.001	0.000	0.014
	(0.012)	(0.014)	(0.040)
Observations	21,288	21,288	21,288
Manufacturer Firms	180	180	180
Bandwidth	6 weeks	6 weeks	6 weeks
MRS FE?	No	Yes	Yes

This table reports estimates of Equation (4.1) – other variables in that equation are included, but not reported. An observation is a manufacturer-retailer-strain-week. The outcome is the change in the log of the price per gram charged by the manufacturer to the retailer (except for in column (3) which is the same outcome, but not logged). MRS stands for manufacturer-retailer-strain fixed effects. The estimates are weighted by the total grams sold across the two weeks of the difference. The P-value tests the null hypothesis that the estimated pass-through is equal to that predicted by TIV. For the placebo regressions, we repeat the analysis one year later. These regressions are estimated with `reghdfe` in Stata. Standard errors twoway-clustered by manufacturer and retailer are in parentheses [84]. *5% significance level. **1% significance level. ***0.1% significance level.

valid. The top panel of Figure B.1 considers bandwidths from 2 to 8 weeks and confirms that our estimates are not sensitive to the bandwidth chosen. We provide an event study of these results in Section 4.5.3.

4.5.2. Retail Price Response

Table 4.3 reports estimates of Equation (4.1) for retailers. The estimate in Column (1), which includes no fixed effects, implies that the reform decreased the tax-exclusive price by 4.5% (significant at the 0.1% level). We include inventory lot fixed effects in Column (2)—our baseline specification. The estimates are very similar; the coefficient in Column (2) implies that the reform reduced tax-exclusive retail prices by 4.4% (significant at the 0.1% level). Combined with the rate change, this implies that tax-*inclusive* prices increased by 2.5%; retailers passed through roughly one-third of the tax to consumers. We find that we can reject the null hypothesis of TIV-consistent pricing behavior at the 0.1 percent level.

As firms might have taken time to adjust (and the Independence Day holiday may have generated temporary price adjustments), Column (3) repeats Column (2) for two week differences and drops the first week after the reform, so that the effect of the reform is identified from the difference between the week before and the week after the reform. The estimates are approximately the same, indicating that neither of these concerns play a large role. We will return to a broader discussion of timing in Section 4.5.3.

Table 4.3 Column (4) repeats Column (2) with the dependent variable in levels—we estimate that average retail tax-exclusive prices fell by 41 cents per gram. This implies that retailers are an average of 41 cents per gram

TABLE 4.3. Retail Tax-Exclusive Price Response

	(1) Δlog(Price)	(2) Δlog(Price)	(3) Δlog(Price)	(4) ΔPrice	(5) Δlog(Price)	6) Δlog(Price)
			Tax Reform			
ΔTax Reform	-0.045^{***}	-0.044^{***}	-0.046^{***}	-0.413^{***}	-0.049^{**}	0.011
	(0.006)	(0.007)	(0.006)	(0.065)	(0.018)	(0.017)
Δlog(Manufacturer Price)						0.887^{***}
						(0.084)
Observations	145,357	145,357	145,357	145,357	11,265	11,265
Retail Firms	110	110	110	110	110	110
Implied Tax-Inclusive Price Change	0.024	0.025	0.023	0.230	0.020	0.080
P-Value for Test of Constant Tax-Inclusive Price	0.000	0.000	0.000	0.000	0.270	0.000
			Placebo			
ΔPlacebo	-0.006^{*}	-0.004	0.001	-0.029	-0.016	-0.004
	(0.003)	(0.003)	(0.002)	(0.017)	(0.012)	(0.009)
Δlog(Manufacturer Price)						0.642^{***}
						(0.053)
Observations	253,123	253,123	253,123	253,123	11,534	11,534
Retail Firms	106	106	106	106	105	105
Bandwidth	6 weeks	6 weeks	6 weeks	6 weeks	6 weeks	6 weeks
MRS FE?	No	No	No	No	Yes	Yes
Inventory Lot FE?	No	Yes	Yes	Yes	No	No
Difference Length	1 week	1 week	2 weeks	1 week	1-4 weeks	1-4 weeks
Restricted to First Week Sales?	No	No	No	No	Yes	Yes

This table reports estimates of Equation (4.1) – other variables in that equation are included but not reported. An observation is an inventory-lot-week. The outcome is the log of the tax-exclusive price per gram charged by the retailer to consumers (except for in column (4) which is the same outcome, but not logged). MRS stands for manufacturer-retailer-strain fixed effects. The estimates are weighted by the total grams sold in the first week of the difference. The P-value tests the null hypothesis that the tax-inclusive price remained constant as predicted by TIV. For the placebo regressions, we repeat the analysis one year later. These regressions are estimated with reghdfe in Stata. In the last two columns we only include observations in their first week of being sold at retail and only if the cannabis was also purchased from the processor in that same week. Standard errors twoway-clustered by manufacturer and retailer are in parentheses [84]. *5% significance level. **1% significance level. ***0.1% significance level.

worse off on existing inventory as a result of the reform. On fresh inventory, firms were roughly 18 cents per gram worse off (41 less the estimated 23 cent decrease in manufacturer prices estimated in Table 4.2). In other words, under TIV this reform should have caused manufacturer and retail tax-exclusive prices to fall by 64 cents leaving profit and consumers unaffected. Instead, it caused smaller manufacturer price cuts leaving both retailers and consumers worse off.

Table 4.3 Columns (5) and (6) take an alternative approach to identification examining inventory lots only in their first week and only if retailers purchased the inventory lot from the manufacturer in that week. For this, we create a panel of retail-processor-strain-weight group-weeks. The estimates are quite similar—a 4.9% decrease in Column (5) versus a 4.4% decrease in Column (2), suggesting that retailers' price responses are largely unaffected by whether they are still selling inventory lots purchased pre-reform or selling new inventory lots purchased post-reform. Column (6) adds the first-differenced log manufacturer price. When included, the coefficient on the wholesale price is not statistically different from one and the coefficient on $\Delta TaxReform$ is now approximately zero. This suggests that retailers largely maintained a constant tax-exclusive markup. This is consistent with the pricing rule derived in Section 4.3. In other words, while retail behavior *as a whole* is inconsistent with TIV, after conditioning on the pass-through from manufacturers, retailers behaved, on average, in a way consistent with marginal-cost pricing (and therefore potentially consistent with TIV).

The bottom panel of Table 4.3 repeats each column in the top panel for a one-year-later placebo reform. If our estimates are valid, the coefficient on $\Delta Placebo$ should be roughly zero—exactly what we find. Even in Column (1) where the estimate is marginally significant, the coefficient is very close to zero. The bottom panel of Figure B.1 considers bandwidths from 2 to 8 weeks and confirms that our estimates are not sensitive to the bandwidth chosen.

4.5.3. Event Studies

The analyses above indicate that prices changed at the time of the reform—yet it is possible that these changes were part of the ongoing evolution of the market, something that the placebo tests one year later cannot rule out. Moreover, the estimates above do not indicate whether there is additional adjustment towards TIV beyond the first week. To address these issues, we conduct event studies for both the manufacturer and retailer responses using our baseline specifications from Tables 4.2 and 4.3. For manufacturers, we do not drop the $t-1$ tax reform coefficient due to our varied difference lengths.[15] Figure 4.2 plots the relevant coefficients and confidence intervals.

In both event studies in Figure 4.2, there is no clear trend in prices pre-reform. The entire response happens in period t, the reform date. Given the varied difference lengths for manufacturers, this implies that manufacturers adjust their prices the first time they sell a particular retail-strain pair post-reform. This is compelling evidence that our estimates are

[15]E.g., for a two week difference that spans t-1 to t+1, both the t and t+1 coefficients are relevant.

unlikely to be driven by the ongoing market evolution and are instead a true response to the reform. The immediate nature of the response suggests that prices in this market follow a unit root, further supporting our first-difference specification. Moreover, this suggests that our results are not driven by learning in the short run and there is no evidence in Figure 4.1 to suggest substantial adjustments based on long-run learning either [85, 86].

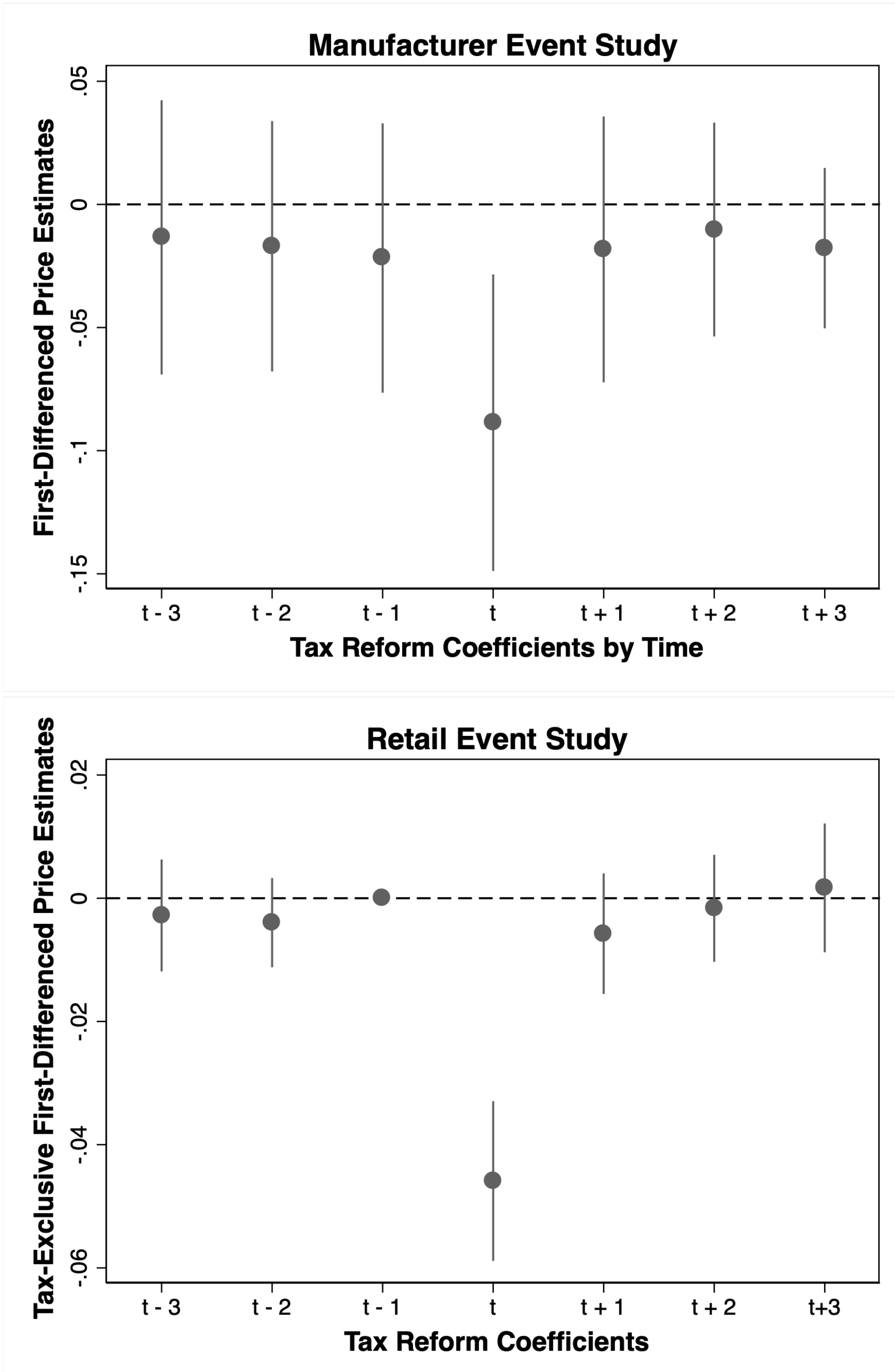
Appendix Figure B.2 replicates the event study plots one year later, further emphasizing the placebo findings in previous sections—our identification strategy is effective in this setting when tested in other periods with similar cyclicality and holiday patterns. If one wanted a difference-in-differences design, one could subtract the placebo estimates from the main estimates in Tables 4.2 and 4.3 and take the same approach here for the event studies. The estimates would be very similar.

4.6. Discussion and Conclusion

TIV is a key component of tax policy design and analysis—it states that taxes may be collected at any point in the supply chain without concern as to the ultimate incidence. While the literature has documented cases in which TIV fails, these results have come with caveats driven by specific frictions or asymmetries present in the markets studied. We study a reform in a market with none of the these issues and show that TIV fails. A reform intended to be welfare-neutral or even welfare-enhancing had negative consequences for both retailers and consumers.

This result is driven by manufacturers, who on average reduced prices significantly less than TIV would predict. Conditional on manufacturer

FIGURE 4.2. Manufacturer and Retail Price Event Study



This figure plots estimates of Table 4.2 Column (2) (top panel) and Table 4.3 Column (2) (bottom panel) with additional leads and lags of $\Delta TaxReform$. The plotted coefficients are leads and lags of $\Delta TaxReform$. We include in the regression (but do not plot) leads and lags are for periods $t-4$ and before and $t+4$ and after as is standard in event study designs. The dots indicate the point estimates and the lines indicate 95% confidence intervals. See the notes for Tables 4.2 and 4.3 for regression details.

prices, we find evidence that retailers applied a constant markup over marginal costs, consistent with our model.

These results are not likely driven by market power; the wholesale market is more competitive than the retail market and thus we would expect violations of TIV to be more likely for retailers than for manufacturers. Similarly, if manufacturers employed average-cost pricing mechanisms [87, 88], we would expect the reform to cause similar or larger price drops than under marginal-cost pricing. While the reform eliminated incentives for inefficient vertical integration and, in the long run, production increased [79], increased production efficiency should similarly drive down prices. The frequency of price changes—and the prevalence of at least some drop in manufacturer prices in response to the reform—suggest that managerial inattention is not relevant [89]. Our event studies suggest the response is immediate, which decreases the likelihood that learning can explain our findings.

Others have found asymmetric firm behaviors in related settings. Benzarti et al. [90] is particularly relevant—they find increases in value-added taxes are passed-through to consumers at twice the rate of decreases. In our setting, retailers, which experienced a tax increase, passed-through taxes in a way that is consistent with standard models of profit maximization, while manufacturers, which experienced a tax decrease, failed to pass-through their savings as predicted. Unlike the VAT context, however, our setting features a simultaneous change and a marketplace where firms and consumers are highly aware of relevant prices; furthermore both retailers and manufacturers engage in repeated transactions with each other over a long period of time. More broadly, the industrial organization literature has identified potential asymmetries in firm responses to changes in demand and costs [91, 92].

We view our results as consistent with models that generate asymmetric responses to changes in market conditions due to behavioral phenomena, as opposed to information, transaction, or competitive frictions. In particular, anchoring and loss aversion may explain the outcomes we observe [93, 94, 95]. While the modal response by manufacturers in the week of the reform was to adjust their prices, the default option of "doing nothing" by maintaining constant tax-inclusive manufacturer prices (and thus realizing a significant increase in variable per-unit profits) may have anchored their negotiations with retail firms. The relatively common and small changes in manufacturer prices we do observe may be a result of competition—manufacturers may "do something" if they incorporate quantity or reputation effects into their analysis of post-tax outcomes [96] and competitors may be compelled to act as a consequence. In contrast, in aggregate, retailers may have overcame their default "do nothing" option (constant tax-inclusive prices) because this option represented a loss in variable per-unit profit. Once the default was overcome, they made decisions consistent with profit maximization.

Our findings have wide-ranging implications for tax policy. First, designers of new taxes may face welfare tradeoffs when choosing where in a supply chain to locate a tax. Both efficiency and equity considerations arise. When considering efficiency, variation in elasticities or competitive structures across the market may affect optimal tax placement. In terms of equity, if a policy goal is to ensure all market participants bear portions of the tax, it may be necessary to impose taxes on these different groups directly.

Second, policymakers considering changes to existing tax policy face greater consequences for doing so. While it may be possible to implement revenue-neutral reforms, restructuring will create clear winners and losers. In this case, manufacturers benefited—despite being in an arguably more competitive market—while retailers and consumers were harmed.

Taken together, these concerns point to broader political economy issues surrounding tax policy [97, 98]. Political systems may be designed to limit the ability of policymakers to enact tax reforms and thus rational actors may unknowingly design systems which have additional inefficiencies as described here. Indeed, in Washington state, the legislature may not reform measures passed by ballot initiative for two years after passage. Though local government officials knew from the moment of passage that the gross receipts tax was likely to have negative consequences on the market, their hands were tied. Flexibility in political and policy systems may help avoid these concerns—though at the cost of volatility and asymmetric responses.

Finally, these results demonstrate a need for further experimental and modelling work. Modern models of competition, growth, trade, inflation, and the business cycle generally make assumptions about taxes which are appealing from a tractability standpoint. These assumptions generally imply TIV [e.g. 99, 100, 101]. Instead of failures of TIV being the exception, our work provides evidence that TIV simply may not hold in practice because of the ubiquitous nature of default options. In the absence of TIV, it may be necessary to conduct experiments which examine the way in which firms and consumers respond to tax policy and construct models which more accurately capture these responses.

CHAPTER V

CONCLUSION

In this dissertation, I consider questions related to public and labor economics in the United States.

In Chapter II, I find both men and women work on evening and weekends, but men work more than women during these times. I examine how work activity responds to unexpected winter weather public school closures and find that Women reduce work activity by 34%. These results are consistent with the emerging theory that asymmetric childcare responsibilities could be a reason why men and women in high-wage professions are working different amounts on average.

In Chapter III, my co-authors and I find distributing SNAP benefits on days other than the first of the month increases the number of drug-related fatal automobile collisions by 1.21 percent and a one-percentage point increase in the share of SNAP benefits distributed on a day leads to a .2 percent increase in the number of drug-related fatal automobile collisions. The public policy implications of these findings are nuanced and should be taken in context with the marginal propensity to consume drugs and alcohol from income of all forms.

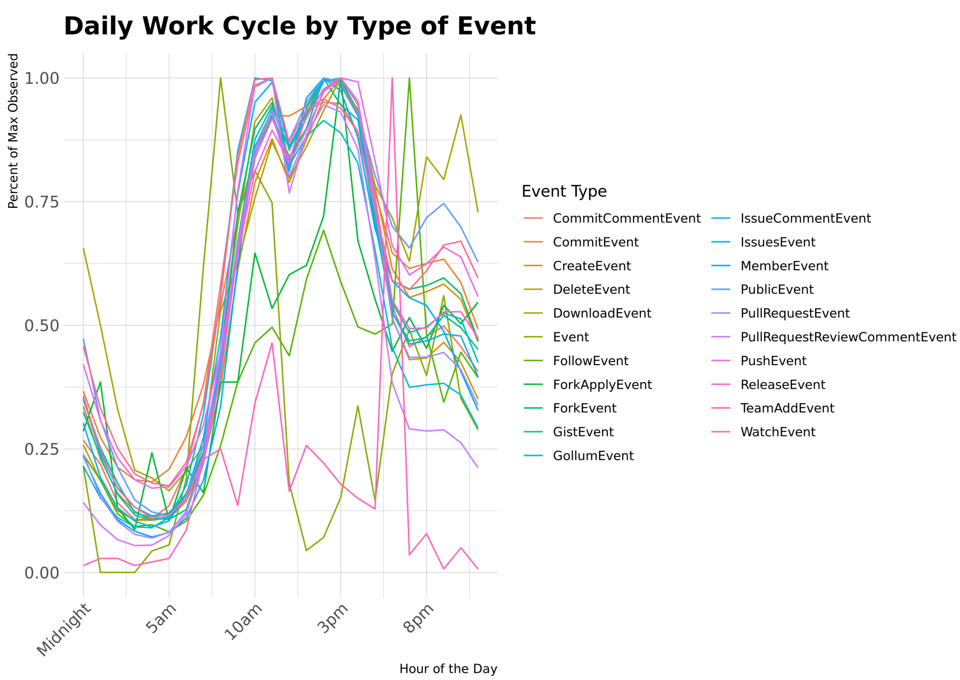
In Chapter IV, I find, in joint work with co-authors, that tax invariance does not hold. When a 25 percent tax remitted by manufacturers was eliminated in Washington state and the retail excise tax was simultaneously increased from 25 to 37 percent, manufacturers kept two-thirds of their tax savings instead of passing all their savings through to retail firms via lower prices as predicted by TIV. One-third of the retail tax increase was passed on to consumers via higher retail prices – TIV would have predicted constant or even declining tax-inclusive retail prices. These findings suggest that tax policy should be carefully designed from the beginning, as tax restructuring could have welfare implications.

These insights contribute to our understanding of human behavior and optimal public policy.

APPENDIX A

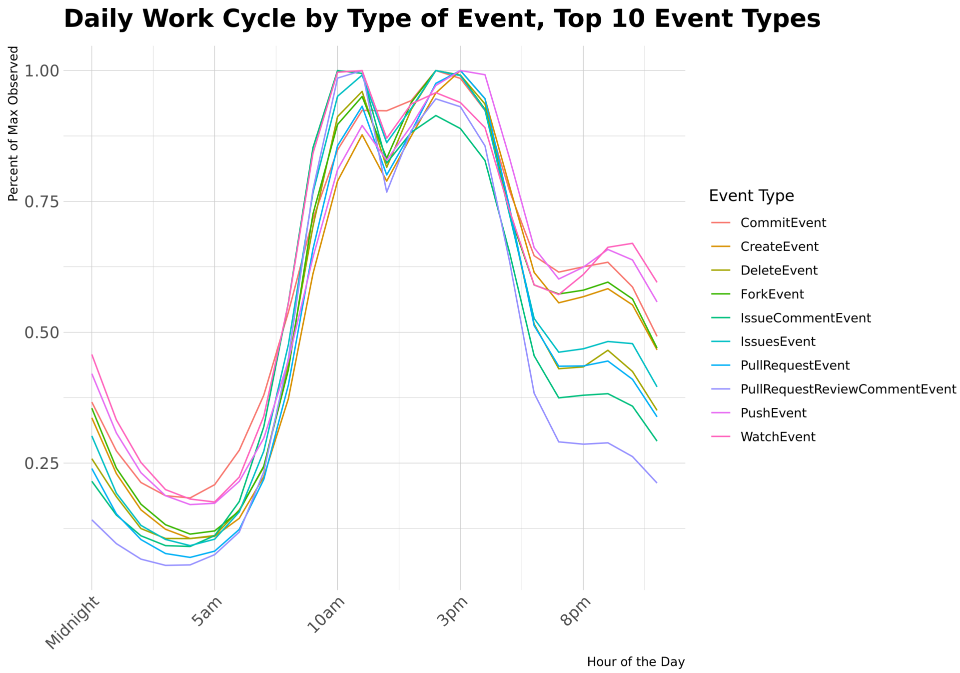
CHAPTER II SUPPLEMENTARY FIGURES

FIGURE A.1. **Work Day for All Event Types**



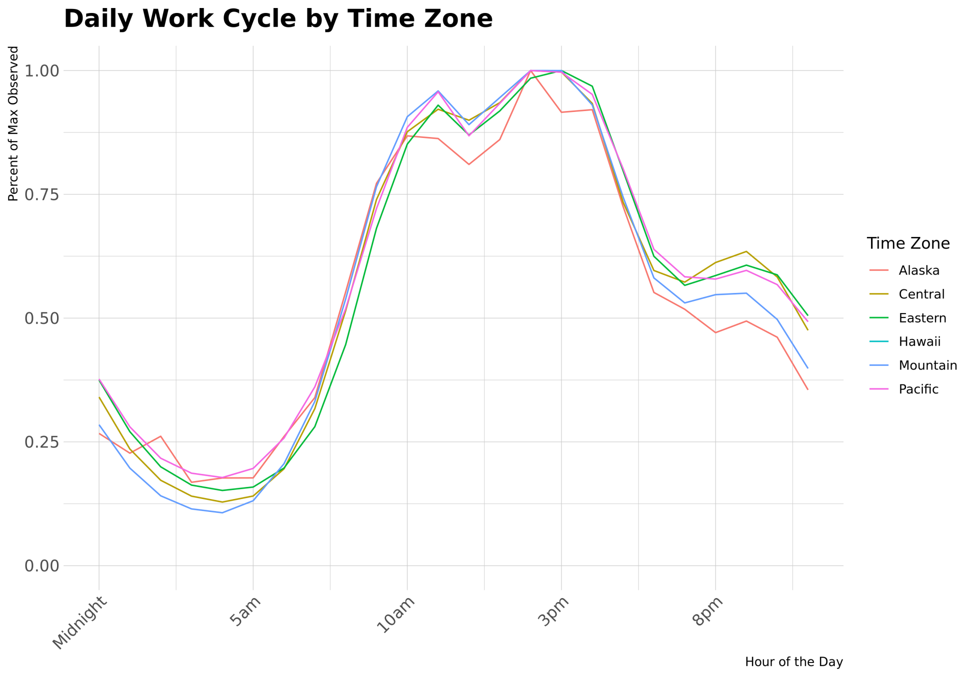
We observe all public activity in GitHub. This public activity comes in the form of different "events". In the figure above, each line represents one type of event. The majority of event types follow the same daily cycle, but a few event types are significantly noisier and reflect a different temporal pattern. Each of these event types with a different temporal pattern represents less than 1% of the data.

FIGURE A.2. **Work Day for Common Event Types**



We observe all public activity in GitHub. This public activity comes in the form of different "events". In the figure above, each line represents one type of event. For this plot, we limit the data to types of events that represent more than 1% of observations. For these event types, the daily pattern is qualitatively similar. The greatest heterogeneity occurs in the evening. Events that involve interactions between users are less common in the evening. For the majority of my analysis, I use "commit" and "push" events. These event types follow each other closely.

FIGURE A.3. **Work Day Across United States' Time Zones**



All data in GitHub Archive and GitHub Torrent is recorded in Coordinated Universal Time. I restore local time using the geographic state of the user and the time zone associated with that state. For states with multiple time zones, I use the time zone that covers the majority of the state. In the figure above, the data from each time zone is shown as an individual line. These lines reflect that data after the event times have been converted from Coordinated Universal Time to local time. The daily work pattern is similar across time zones.

TABLE A.1. **Observations by Event Type**

Event Type	Percent of Observations
Commit	39.43
Push	30.49
Issue Comment	6.71
Create	6.26
Pull Request	4.17
Watch	3.97
Issues	3.14
Pull Request Review Comment	1.73
Delete	1.54
Fork	1.12
Gollum	.47
Commit Comment	.26
Release	.22
Member	.20
Follow	.12
Gist	.08
Public	.06
Event	.02
Download	.01
Fork Apply	.00
Team Add	.00

I observe public activity by all users in GitHub. This activity is described as "events", and there are many types of events. This table documents the relative occurrence of each type of event in the data. Commit Events and Push Events are by far the most common events. Together, these events are 69.9% of the data. A commit event changes the local copy of a file. A push event changes the copy on the remote server. The remote server is shared across users who have access to a project. All event types listed above are described in detail in the Data Appendix, see Table A.2.

TABLE A.2. **Description of Event Types**

Event Type	Description
Commit	Save changes to the local copy of a file.
Push	Save changes to the remote copy of a file.
Issue Comment	An issue comment has either been created, edited, or deleted. This comment is attached to an issue that has been filed with existing code.
Create	A new branch or tag has been created. A branch is a copy of the main work that can be edited without impacting the main work. A user may choose to make the branch the main work at a later point.
Pull Request	When a user would like to contribute code to a project, this new code is issued as a pull request.
Watch	A watch event occurs when someone stars a repository. When a user stars a repository, they are choosing to follow this project.
Issues	The user identifies an issue with existing code.
Pull Request Review Comment	A pull request is being reviewed by another member of the project.
Delete	A branch or tag is deleted.
Fork	A user copies an existing project. This copy is not linked to the original project.
Gollum	Create a Wiki page.
Commit Comment	Comment on a commit that has already occurred.
Release	Release a new version of a software package.
Member	Add or remove a member of a project.
Follow	Follow the activity of another user.
Gist	Create or update a gist. A gist is a snippet of code.
Public	Switch a repository from public to private.
Download	Download a package.This event is no longer supported.
Fork Apply	Apply a patch to a fork. A patch covers the parts of someone's fork that you would like to apply to your code. This event is no longer supported.
Team Add	Add a repository to a team. A team is a group of members. This team is a subgroup of an organization.

We observe all public activity in the GitHub user platform. This activity can be one of the types listed above. All activity is described as an "event". The left column is the event and the right column describes the event. Files are edited locally. These changes are first saved locally. These changes can then be saved to the remote server. The events listed above include actions that a user takes on the local file, actions on the remote version of the file, and actions on other users' files.

APPENDIX B

CHAPTER IV SUPPLEMENTARY INFORMATION

In this appendix, I detail our data cleaning procedure for Chapter IV and provide supplementary figures.

I begin by detailing our methods for cleaning prices in the face of changing reporting requirements and tax rates. I then discuss other cleaning steps to transform the raw data into the set used in our analyses.

The retail sales prices reported by firms in the "seed-to-sale" traceability system were supposed to be all-tax-inclusive pre-reform and tax-exclusive post-reform. However, compliance varied from firm to firm and changed over time. For example, some firms reported prices with the sales tax included and some reported prices without the sales tax.

This reporting confusion means that we must infer, for each firm, how they reported their prices and therefore the true tax-inclusive and tax-exclusive prices they charged. For each firm-week, we assign a "multiplier" that reveals the relationship between the reported price and the price faced by consumers. This chosen multiplier is selected from a set of multipliers based on possible tax rates for the firm. We merge in the state and local sales tax rates for each firm in order to construct this choice set.[1] To understand the relationship between the multiplier, reported prices, and

[1]For five firms, the state and local tax rates do not match the rates they are using, so we adjust these. And a few firms do not ever change their local tax rate for reporting purposes—we make that adjustment as well. This transforms these firms from very unround to very round, but otherwise has approximately no effect on the data as the difference between the statutory and reported local tax rates is very small.

faced prices, consider the following equation:

$$Price_{Consumer} = Price_{Reported} \times Multiplier$$

We algorithmically determine which tax-based multiplier makes the prices faces by consumer's ($Price_{Consumer}$) most round for each week, where roundness is the closeness of the price to a 25 cent increment of a dollar. For each product type, $Price_{Reported}$ is the modal observed price for the week, where idiosyncratic discounts have been removed.[2]

We consider two orthogonal methods of determine the proper set of multipliers. Our results are robust to the method used. Ultimately, we find the modal firm never included the sales tax, included the excise tax pre-reform, and excluded the excise tax post-reform.

B.1. Cash Market Identification

In order to determine how each firm reports their prices in the traceability system, we take advantage of two characteristics of retail prices. First, publicly advertised prices (or 'list' prices) are nearly universally all tax-inclusive. Second, retailers nearly always choose to set prices in whole-dollar or (rarely) quarter-dollar increments.[3] We use these two facts to determine the difference between the list prices faced by consumers and the prices reported in the traceability system.

[2]We determine that a price is a one-off discount if the price for that transaction is 5% to 95% (in increments of 5 percentage points) or 33%/66.67% less than the previously reported price.

[3]We verified this through conversations with retailers as well as using historical menus available through The Internet Archive and a full set of menus for almost all firms we took screen shots of on 7/18/2017.

We assign each firm a multiplier before and after the tax change. We begin by assigning the modal firm's multiplier choices to all firms—all firms' prices were adjusted by the state and local sales tax pre-reform and all firm's prices were adjusted by the excise tax plus state and local sales tax post-reform. We then make the following adjustments based on the results from our algorithm:

1. We leave prices unadjusted (i.e. a multiplier of 1) where are algorithm finds that this choice maximizes roundedness and at least 85 percent of weekly sales are round with this multiplier choice.[4] This applies to 16% of firms.

2. We adjust the multiplier post reform to account for only the excise tax when the algorithm finds that this choice maximizes roundness and at least 85 percent of weekly sales are round with this multiplier choice. This applies to one firm (out of 110).

   There are three additional firms for whom an only excise tax adjustment makes them most round, but their roundness in the immediate post period is less than 85 percent. We leave two of the firms alone because they were also left alone in the pre-reform period because of unroundness and we could either adjust them both before and after the reform or leave them both alone with similar effects to the log price change. The third firm becomes more round a few weeks

[4]For the 4.5% of firms that suggest the multiplier could be 1 but are quite unround, there is too much uncertainty to confidently make an adjustment. Leaving these firms' multipliers unchanged, if wrong, will bias our estimates towards our main null hypothesis in the retail section of the paper—that firms did not adjust their prices in response to the reform.

after the reform and keeps this multiplier through the end of our data (and we have confirmed the multiplier in the menu screen shots), so we make this multiplier adjustment.

3. There are two firms for whom the multiplier that makes them round post-reform is the excise tax + state and local sales taxes divided by the state and local sales tax rate. In both cases, we have clear evidence that this is because they adjusted their prices post-reform by making their prices sales-tax exclusive post-reform. One firm keeps this choice permanently and we see this in the menu screen shots at the end of our data. The other firm eventually adjusts to the modal firms' multiplier. Our assumption keeps prices roughly constant through this reporting change.

B.2. Product Batch Price Stability

To provide additional evidence that our multiplier decisions are not systematically biasing our estimates, we consider a completely different mechanism for determining multipliers—we pick the set of multipliers that makes the tax-inclusive prices for the most number of inventory lots for a given firm the same pre- and post-reform.

There are a couple of reasons why this is a reasonable alternative to consider. A number of inventory lots did leave prices constant in response to the tax reform and the main null hypothesis in our retail analysis is that firms did not change their tax-inclusive prices—this is what we would expect if the tax reform was indeed tax invariant.

We consider two variants of this. One is to begin with the modal firms' multipliers and adjust it to another multiplier if it decreases the number of price changes by any margin. The second variant is to begin with our estimates based on roundedness and then make adjustments for firms that under the best set of multipliers leaves at least 25% of their inventory lots constant in response to the reform. The latter changes the multipliers for only four firms and three of those four leave the percent price changes quite similar. The former method decreases our baseline estimate by 0.4 percentage points and the latter decreases our baseline estimate by 0.2 percentage points. This evidence strongly supports our price cleaning methods and suggests that any remaining bias is extremely small.

B.3. Additional Cleaning

In addition to adjusting retail prices, we also drop some extreme outliers in the data. In particular, we drop all wholesale transactions with a usable weight above 2,500 grams[5] and all retail transactions if the usable weight was above 28.5 grams.[6] We also drop all wholesale or retail price per grams above $42.[7] We censor the THC content data if it is zero or above 40 in both the manufacturer and retailer data.[8] We also drop wholesale prices less than $1. This effectively drops samples from our data, which are sold

[5]This is about 0.025% of wholesale transactions.

[6]This is because the maximum legal sale was one ounce. This step drops 0.15% of retail transactions.

[7]This is less than 0.03% of wholesale transactions and less than 0.04% of retail transactions.

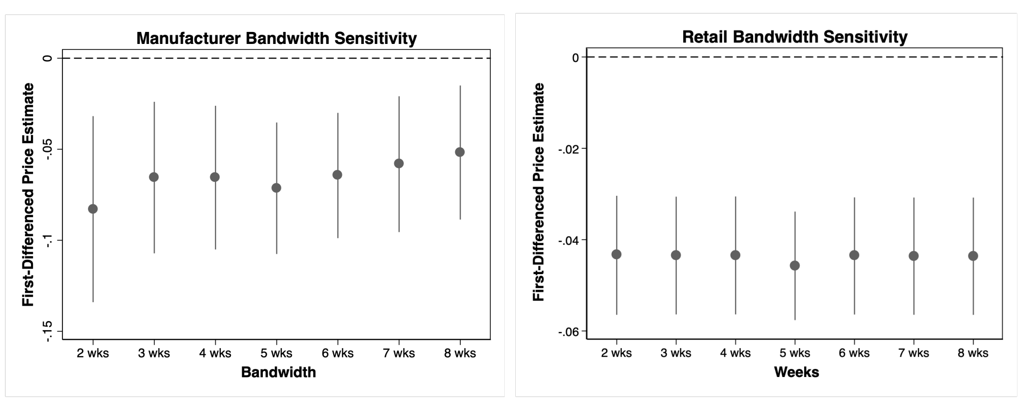
[8]This affects 0.2% of wholesale transactions and 5% of retail transactions.

well below market value. We typically see these as the first recorded sale from a parentlot.

Lastly, we drop some firms or firm-days in our data set. In particular, we require for each firm that the first sales transaction occurs two months before the tax reform and continues to have transactions through the two months after the reform (either because they had not yet opened, had closed, or because they took a long hiatus from selling any cannabis). A few retailers conduct a "soft opening" by opening briefly, closing for more than a month, and then re-open permanently. In these cases, we drop the first brief selling period and consider their first activity date the first date upon re-opening in our data. We also drop 20 retail firms for whom at some point in the 8 weeks before or after the reform report their data only once per day—this is a clear indicator of poor overall data quality and, because of this, determining the appropriate multipliers for these firms is difficult.

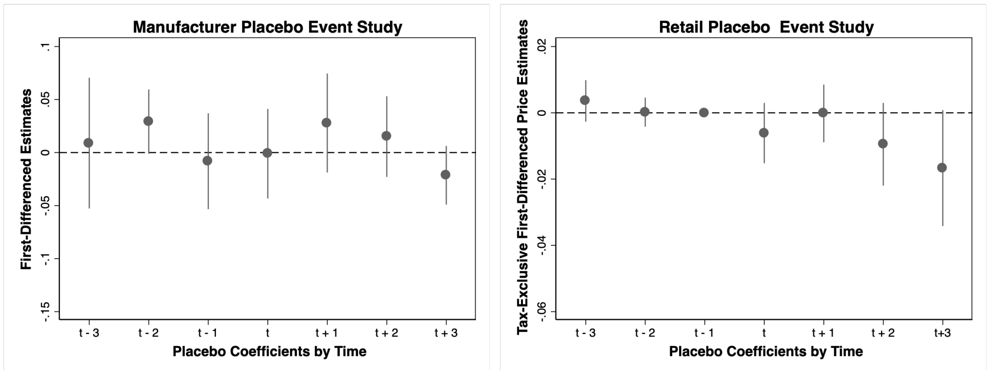
B.4. Supplementary Figures

FIGURE B.1. Manufacturer and Retail Price Bandwidth Choices



This figure plots estimates of Table 4.2 Column (2) in the top panel and Table 4.3 Column (2) in the bottom panel, varying the bandwidth. The bandwidth in our baseline specifications is 6 weeks. The estimates plotted are for the coefficient on $TaxReform$. The dots indicate the point estimates and the lines indicate 95% confidence intervals. See the notes for Tables 4.2 and 4.3 for regression details.

FIGURE B.2. Manufacturer and Retail Price Event Study



This figure plots placebo estimates of Table 4.2 Column (2) (top panel) and Table 4.3 Column (2) (bottom panel) with additional leads and lags of $\Delta Placebo$. The plotted coefficients are leads and lags of $\Delta Placebo$. We include in the regression (but do not plot) leads and lags are for periods $t - 4$ and before and $t + 4$ and after as is standard in event study designs. The dots indicate the point estimates and the lines indicate 95% confidence intervals. See the notes for Tables 4.2 and 4.3 for regression details.

REFERENCES CITED

[1] U.S. Census Bureau. Occupation by sex and median earnings in the past 12 months. 2017. URL `https://factfinder.census.gov/`.

[2] National Center for Education Statistics. Bachelor's degrees conferred by postseconday institutions, by race/ethnicity and sex of student: Selected years, 1976-77 through 2015-2016. 2017. URL `https://nces.ed.gov/programs/digest/d17/tables/dt17`$_3$22.20.*asp*.

[3] American Association of University Women. Women's Student Debt Crisis in the United States. 2019.

[4] U.S. Census Bureau. Poverty Status in the Past 12 Months by Sex and Age. 2017. URL `https://data.census.gov/cedsci/`.

[5] Joint Economic Committee. Gender Pay Inequality: Consequences for Women, Families and the Economy. 2016. URL `https://www.jec.senate.gov`.

[6] Francince D. Blau and Lawrence F. Katz. The Gender Wage Gap: Extent, Trends, and Explanations. 55(3):789–865, 2017.

[7] Claire Cain Miller. Women Did Everything Right. Then Work Got 'Greedy.'. April 2019. URL `https://www.nytimes.com`.

[8] Claire Cain Miller. Millenial Men Aren't the Dads They Thought They'd Be. July 2015. URL `https://www.nytimes.com`.

[9] Claire Cain Miller. The 24/7 Work Culture's Toll on Families and Gender Equality. May 2015. URL `https://www.nytimes.com`.

[10] Youngjoo Cha and Kim A. Weeden. Overwork and the Slow Convergence in the Gender Gap in Wages. 79(3):457–484, 2014.

[11] Mary C. Noonan, Mary E. Corcoran, and Paul N. Courant. Pay Differences Among the Highly Trained: Cohort Differences in the Sex Gap in Lawyers' Earnings. 84(2):851–870, 2005.

[12] Claudia Goldin and Lawrence F. Katz. A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation. 34(5): 705–746, 2016.

[13] Marianne Bertrand, Claudia Goldin, and Lawrence F. Katz. Dynamic of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. 2(3):228–255, 2010.

[14] Francince D. Blau and Lawrence M. Kahn. Female Labor Supply: Why Is the United States Falling Behind. 103(3):251–256, 2013.

[15] Kim A. Weeden, Youngjoo Cha, and Mauricio Bucca. Long Work Hours, Part-Time Work, and Trends in the Gender Gap in Pay, the Motherhood Wage Penalty, and the Fatherhood Wage Premium. 2(4): 71–102, 2016.

[16] United States Court of Appeals for the Ninth Circuit. hiQ Labs, Inc. v. LinkedIn Corporation, 2019. No. 17-16783. September 9, 2019.

[17] Claudia Goldin, Sari Pekkala Kerr, Claudia Olivetti, and Erling Barth. The Expanding Gender Earnings Gap: Evidence from the LEHD-2000 Census. 107(5):110–114, 2017.

[18] Jamin D. Speer. Bye Bye Ms. American Sci: Women and the Leaky STEM Pipeline. 2019. URL https://sites.google.com/site/jaminspeer/research.

[19] Anniek van den Hurk, Martina Meelissen, and Annemarie van Langen. Interventions in education to prevent STEM pipeline leakage. 41(2): 150–164, 2019.

[20] Cameron Blevins and Lincoln Mullen. Jane, John ... Leslie? A Historical Method of Algorithmic Gender Prediction. 9(3), 2015.

[21] Pew Research Center. For most highly educated women, motherhood doesn't start until the 30s. Technical report, 2015.

[22] Marianne Bertrand. Career, Family, and the Well-Being of College-Educated Women. 103(3):244–250, 2013.

[23] Claudia Goldin and Joshua Mitchell. The New Life Cycle of Women's Employment: Disappearing Humps, Sagging Middles, Expanding Tops. 31(1):161–182, 2017.

[24] Eunjung Jee, Joya Misra, and Mart Murray-Close. Motherhood penalties in the U.S., 2986-2014. 2018. URL https://equitablegrowth.org/working-papers/motherhood-penalties/.

[25] U.S. Bureau of Labor Statistics. Employment Situation Summary – September 2020, 2020.

[26] Guido W Imbens, Donald B Rubin, and Bruce I Sacerdote. Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. *American economic review*, 91(4):778–794, 2001.

[27] Jonathan Gardner and Andrew J Oswald. Money and mental wellbeing: A longitudinal study of medium-sized lottery wins. *Journal of health economics*, 26(1):49–60, 2007.

[28] Jason M Lindo. Parental job loss and infant health. *Journal of health economics*, 30(5):869–879, 2011.

[29] Hilary Hoynes, Doug Miller, and David Simon. Income, the earned income tax credit, and infant health. *American Economic Journal: Economic Policy*, 7(1):172–211, 2015.

[30] Christopher J Ruhm. Are recessions good for your health? *The Quarterly journal of economics*, 115(2):617–650, 2000.

[31] William N Evans and Timothy J Moore. The short-term mortality consequences of income receipt. *Journal of Public Economics*, 95 (11-12):1410–1424, 2011.

[32] William N Evans and Timothy J Moore. Liquidity, economic activity, a. hoynnd mortality. *Review of Economics and Statistics*, 94(2):400–418, 2012.

[33] Carlos Dobkin and Steven L. Puller. The effects of government transfers on monthly cyclyes in drug abuse, hospitalization and mortality. *Journal of Public Economics*, 91:2137–2157, 2007.

[34] Justine S. Hastings and Ebonya Washington. The First of the Month Effect: Consumer Behavior and Store Responses. *American Economic Journal: Economic Policy*, 2(2):142–162, 2010.

[35] Jesse M. Shapiro. Is there a daily discount rate? Evidence from the food stamp nutrition cycle. *Journal of Public Economics*, 89:303–325, 2005.

[36] Jessica E. Todd. Revisiting the Supplemental Nutrition Assistance Program cycle of food intake: Investigating heterogeneity, diet quality, and a large boost in benefit amounts. *Applies Economic Perspectives and Policy*, 37(3):437–458, 2015.

[37] Parke E. Wilde and Christine Ranney. The Monthly Food Stamp Cycle: Shopping Frequency and Food Intake Decisions in an Endogenous Switching Regression Framework. *American Journal of Agricultural Economics*, 82(2):200–213, 2000.

[38] Travis A. Smith, Joshua P. Berning, Xiaosi Yang, Gregory Colson, and Jeffrey H. Dorfman. The Effects of Benefit Timing and Income Fungibility on Food Purchasing Decisions among Supplemental Nutrition Assistance Program Households. *American Journal of Agricultural Economics*, 98(2):564–580, 2016.

[39] US HHS. Drug Testing Welfare Recipients: Recent proposals and continuing controversies. *Assistant Secretary for Planning and Evaluation*, 2011.

[40] Bryce Covert and Josh Israel. States spend millions to drug test the poor, turn up few postivie results. *Thinkprogress*, 2017.

[41] National Institute on Drug Abuse. National Survey of Druge Use and Health. 2016.

[42] CBS. Drug testing for food stamps may be coming soon. 2018.

[43] Marianne Bitler. The Health and Nutrition Effects of SNAP: Selection Into the Program and a Review of the Literature on Its Effects, 2015.

[44] Hilary W. Hoynes, Leslie McGranahan, and Diane W. Schanzenbach. SNAP and Food Consumption. 2016.

[45] Justine S. Hastings and Jesse M. Shapiro. How Are SNAP Benefits Spent? Evidence from a retail panel. *American Economic Review*, Forthcoming, 2018.

[46] Hilary W. Hoynes and Diane W. Schanzenbach. Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program. *American Economic Journal: Applied Economics*, 1 (4):109–139, 2009.

[47] Timothy K.M. Beatty and Charlotte J. Tuttle. Expenditure Response to Increases in In-Kind Transfers: Evidence from the Supplemental Nutrition Assistance Program. *American Journal of Aggircultural Economics*, 97(2):390–404, 2014.

[48] Gregory A. Bruich. The effect of SNAP benefitd on expenditures: New evidence from scanner data and the November 2013 beneit cuts, 2014.

[49] Jacob Goldin, Tatiana Homonoff, and Katherine Meckel. Is there an Nth of the Month Effect? The Timing of SNAP Issuance, Food Expenditures, and Grocery Prices. *Working paper*, 2016.

[50] Karen S. Hamrick and Margaret Andrews. SNAP Participants' Eating Patterns over the Benefit Month: A Time Use Perspective. *PLOS One*, 11(7), 2016.

[51] Michael A. Kuhn. Cyclical Food Insecurity and Electronic Benefit Transfer. *Working paper*, 2018.

[52] Chad Cotti, John Gordanier, and Orgul Ozturk. When Does it Count? The Timing of Food Stamp Receipt and Educational Performance. *Working paper*, 2017.

[53] Jillian B. Carr and Analisa Packam. SNAP Benefits and Crime: Evidence from Changing Disbursement Schedules. *Review of Economics and Statistics, forthcoming*, 2018.

[54] C. Fritz Foley. Welfare Payments and Crime. *The Review of Economics and Statistics*, 93(1):97–112, 2011.

[55] Hilary K. Seligman, Ann F. Bolger, David Guzman, Andrea Lopez, and Kirsten Bibbins-Domingo. Exhaustion of Food Budgets at Month's End and Hospital Admissions for Hypoglycemia. *National Institute of Health*, 33(1):116–123, 2014.

[56] Mayo Clinic. Hypoglycemia. 2018.

[57] Lin-Chi Hsu. The Timing of Welfare Payments and Intimate Partner Violence. *Economic Inquiry*, 55(2), 2017.

[58] USDA FNS. Supplemental Nutrition Program Participation and Costs. 2018.

[59] U.S. Census Bureau. Annual Estimates of the Resident Population. 2016.

[60] USDA FNS. A Short History of SNAP. 2018.

[61] USDA FNS. 2008 Farm Bill. 2008.

[62] Mikael Lindahl. Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. *Journal of Human resources*, 40(1):144–168, 2005.

[63] Campbell R McConnell, Stanley L Brue, and Sean Flynn. *Economics: Principles, Problems, and Policies*. McGraw-Hill, 2018.

[64] N Gregory Mankiw. *Principles of Economics*. Cengage Learning, 2020.

[65] Joel Slemrod. Does it matter who writes the check to the government? The economics of tax remittance. *National Tax Journal*, 61(2): 251–275, 2008. ISSN 00280283, 19447477. URL http://www.jstor.org/stable/41790444.

[66] Wojciech Kopczuk, Justin Marion, Erich Muehlegger, and Joel Slemrod. Does tax-collection invariance hold? Evasion and the pass-through of state diesel taxes. *American Economic Journal: Economic Policy*, 8 (2):251–86, May 2016. doi: 10.1257/pol.20140271. URL http://www.aeaweb.org/articles?id=10.1257/pol.20140271.

[67] Anne Brockmeyer and Marco Hernandez. Taxation, information, and withholding: Evidence from Costa Rica. Working paper, 2016.

[68] Joah Muysken, Tom Van Veen, and Erik De Regt. Does a shift in the tax burden create employment? *Applied Economics*, 31(10):1195–1205, 1999. doi: 10.1080/000368499323418. URL https://doi.org/10.1080/000368499323418.

[69] Emmanuel Saez, Manos Matsaganis, and Panos Tsakloglou. Earnings determination and taxes: Evidence from a cohort-based payroll tax reform in Greece. *The Quarterly Journal of Economics*, 127(1): 493–533, 2012. ISSN 00335533, 15314650. URL http://www.jstor.org/stable/41337215.

[70] Etienne Lehmann, François Marical, and Laurence Rioux. Labor income responds differently to income-tax and payroll-tax reforms. *Journal of Public Economics*, 99:66 – 84, 2013. ISSN 0047-2727. doi: https://doi.org/10.1016/j.jpubeco.2013.01.004. URL http://www.sciencedirect.com/science/article/pii/S0047272713000170.

[71] Raj Chetty, Adam Looney, and Kory Kroft. Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77, September 2009. doi: 10.1257/aer.99.4.1145. URL http://www.aeaweb.org/articles?id=10.1257/aer.99.4.1145.

[72] Amy Finkelstein. E-Z tax: Tax salience and tax rates. *The Quarterly Journal of Economics*, 124(3):969–1010, 08 2009. ISSN 0033-5533. doi: 10.1162/qjec.2009.124.3.969. URL https://doi.org/10.1162/qjec.2009.124.3.969.

[73] R. La Corte. Gov. Inslee signs recreational marijuana reform bill. *Associated Press*, June 30 2015. URL ap.org.

[74] Brett Hollenbeck and Kosuke Uetake. Taxation and market power in the legal marijuana industry. *Available at SSRN 3237729*, 2019.

[75] Chris Mace, Elena Patel, and Nathan Seegert. Marijuana taxation and imperfect competition. *National Tax Journal*, 73(2):545–592, 2020.

[76] E Glen Weyl and Michal Fabinger. Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy*, 121(3):528–583, 2013.

[77] Keaton Miller and Boyoung Seo. Tax revenues when substances substitute: Marijuana, alcohol, and tobacco. *Kelley School of Business Research Paper*, (18-27), 2018.

[78] Ben Hansen, Keaton Miller, Boyoung Seo, and Caroline Weber. Taxing the potency of sin goods: Evidence from recreational cannabis and liquor markets. *National Tax Journal*, 2020.

[79] Benjamin Hansen, Keaton Miller, and Caroline Weber. Vertical integration and production inefficiency in the presence of a gross receipts tax. Working paper, 2020.

[80] James M. Cole. Memorandum for all United States attorneys. `https://www.justice.gov/iso/opa/resources/3052013829132756857467.pdf`, 2013. Accessed: 2017-07-20.

[81] D. B. Suits and R. A. Musgrave. Ad valorem and unit taxes compared. *The Quarterly Journal of Economics*, 67(4):598–604, 1953. ISSN 0033-5533. doi: 10.2307/1883604. URL `https://doi.org/10.2307/1883604`.

[82] Susan E. Skeath and Gregory A. Trandel. A pareto comparison of ad valorem and unit taxes in noncompetitive environments. *Journal of Public Economics*, 53(1):53 – 71, 1994. ISSN 0047-2727. doi: https://doi.org/10.1016/0047-2727(94)90013-2. URL `http://www.sciencedirect.com/science/article/pii/0047272794900132`.

[83] Michael Keen. The balance between specific and ad valorem taxation. *Fiscal Studies*, 19(1):1–37, 1998. doi: https://doi.org/10.1111/j.1475-5890.1998.tb00274.x.

[84] A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249, 2011. doi: 10.1198/jbes.2010.07136.

[85] Ulrich Doraszelski, Gregory Lewis, and Ariel Pakes. Just starting out: Learning and equilibrium in a new market. *American Economic Review*, 108(3):565–615, 2018.

[86] Yufeng Huang, Paul B Ellickson, and Mitchell J Lovett. Learning to set prices in the Washington state liquor market. *Manuscript. University of Rochester. Simon Business School*, 2018.

[87] Robert L Hall and Charles J Hitch. Price theory and business behaviour. *Oxford Economic Papers*, (2):12–45, 1939.

[88] Carlo Altomonte, Alessandro Barattieri, and Susanto Basu. Average-cost pricing: Some evidence and implications. *European Economic Review*, 79:281–296, 2015.

[89] Xavier Gabaix. Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2, pages 261–343. Elsevier, 2019.

[90] Youssef Benzarti, Dorian Carloni, Jarkko Harju, and Tuomas Kosonen. What goes up may not come down: Asymmetric incidence of value-added taxes. *Journal of Political Economy*, page forthcoming, 2020. doi: 10.1086/710558. URL `https://doi.org/10.1086/710558`.

[91] Andrew Butters, Daniel Sacks, and Boyoung Seo. How do national firms respond to local shocks? Evidence from excise taxes. *Kelley School of Business Research Paper*, 2019.

[92] Stefano DellaVigna and Matthew Gentzkow. Uniform pricing in us retail chains. *The Quarterly Journal of Economics*, 134(4):2011–2084, 2019.

[93] Daniel Kahneman, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press, 1982.

[94] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1):193–206, 1991.

[95] B Douglas Bernheim and Antonio Rangel. Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics*, 124(1):51–104, 2009.

[96] Julio J Rotemberg. Fair pricing. *Journal of the European Economic Association*, 9(5):952–981, 2011.

[97] Stanley L Winer and Walter Hettich. Structure and coherence in the political economy of public finance. *The Oxford handbook of political economy*, 7:441, 2006.

[98] Jürgen von Hagen. Political economy of fiscal institutions. In Donald A. Wittman and Barry R. Weingast, editors, *The Oxford Handbook of Political Economy*. 2008.

[99] Kenneth L Judd. Capital-income taxation with imperfect competition. *American Economic Review*, 92(2):417–421, 2002.

[100] Marc J Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725, 2003.

[101] Jordi Galí. *Monetary policy, inflation, and the business cycle: An introduction to the new Keynesian framework and its applications*. Princeton University Press, 2015.