Title
Worker and Firm Responses to Environmental Policies

Author
Gibson, Matthew

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Worker and Firm Responses to Environmental Policies

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Matthew Gibson

Committee in charge:

Professor Julie Cullen, Co-Chair
Professor Mark Jacobsen, Co-Chair
Professor Prashant Bharadwaj
Professor Joshua Graff Zivin
Professor Craig McIntosh
Professor Junjie Zhang

2015

The dissertation of Matthew Gibson is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Co-Chair

University of California, San Diego

2015

EPIGRAPH

Truth lives, in fact, for the most part on a credit system. Our thoughts and beliefs 'pass,' so long as nothing challenges them, just as bank-notes pass so long as nobody refuses them.

—William James

TABLE OF CONTENTS

Signature Page . . . iii

Epigraph . . . iv

Table of Contents . . . v

List of Figures . . . viii

List of Tables . . . ix

Acknowledgements . . . xi

Vita . . . xii

Abstract of the Dissertation . . . xiii

Chapter 1 The effects of road pricing on driver behavior and air pollution . . . 1
- 1.1 Introduction . . . 1
- 1.2 Background . . . 4
- 1.3 Data . . . 7
- 1.4 Estimation . . . 8
- 1.5 Results . . . 11
  - 1.5.1 Traffic . . . 11
  - 1.5.2 Interaction with public transit . . . 16
  - 1.5.3 Price response . . . 17
  - 1.5.4 Pollution . . . 19
  - 1.5.5 Robustness checks . . . 22
- 1.6 Conclusion . . . 24
- 1.7 Figures & tables . . . 32
  - 1.7.1 Figures . . . 32
  - 1.7.2 Tables . . . 36
- 1.8 Supplementary material . . . 42
  - 1.8.1 Public transit results . . . 42
  - 1.8.2 Supplemental figures . . . 45
  - 1.8.3 Supplemental tables . . . 47

Chapter 2 Regulation-induced pollution substitution . . . 52
- 2.1 Introduction . . . 52
- 2.2 Background . . . 57
  - 2.2.1 The Clean Air Act . . . 57
  - 2.2.2 Regulation of water and land emissions . . . 58
  - 2.2.3 Abatement strategies and variable costs . . . 59

2.3 Theory . . . 61
2.4 Data . . . 64
2.5 Estimation . . . 65
2.5.1 Estimating equations . . . 65
2.5.2 Defining treatment . . . 67
2.5.3 Treatment exogeneity . . . 69
2.6 Results . . . 73
2.6.1 Air emissions . . . 73
2.6.2 Cross-media substitution, all industries . . . 74
2.6.3 Cross-media substitution, by industry . . . 75
2.6.4 Leakage . . . 76
2.7 Additional results, robustness & placebos . . . 78
2.7.1 Air emissions . . . 78
2.7.2 Cross-media substitution . . . 80
2.7.3 Placebos . . . 81
2.8 Conclusion . . . 82
2.9 Figures and tables . . . 88
2.9.1 Figures . . . 88
2.9.2 Tables . . . 91
2.10 Supplementary material . . . 98
2.10.1 Alternative theoretical models . . . 98
2.10.2 Additional tables . . . 101

Chapter 3 Time Use and Productivity: The Wage Returns to Sleep . . . 114
3.1 Introduction . . . 115
3.2 Identifying the effect of sleep on productivity and wages . . . 118
3.2.1 Previous research . . . 118
3.2.2 A productive sleep model . . . 120
3.3 Empirical strategy . . . 122
3.3.1 Estimating equations . . . 122
3.3.2 Local sunset time instruments . . . 124
3.3.3 Wage setting and measurement error . . . 130
3.4 Data . . . 132
3.5 Short-run results: the effect of sleep on wages . . . 136
3.5.1 Primary short-run results . . . 136
3.5.2 Short-run robustness checks . . . 139
3.6 Long-run results: the effect of sleep on wages . . . 141
3.6.1 Primary results . . . 141
3.6.2 Long-run robustness checks . . . 143
3.6.3 Instrument validity . . . 147
3.6.4 County average wages . . . 150
3.7 Conclusion . . . 152
3.8 Supplementary material . . . 159

3.8.1 Background and summary statistics 159
3.8.2 Solar mechanics 162
3.8.3 Measurement error in seasonal estimate 163
3.8.4 Auxiliary results and robustness checks 168
3.8.5 ATUS robustness checks 168

LIST OF FIGURES

Figure 1.1: Timeline of road pricing in Milan 32
Figure 1.2: Air pollution monitoring stations in Milan 32
Figure 1.3: Effect of Area C charge suspension on vehicle entries 33
Figure 1.4: Effect of Area C charge suspension on vehicle entries, by 15-minute interval 34
Figure 1.5: Effect of placebo suspensions on Area C vehicle entries 35
Figure 1.6: Marginal drivers 43
Figure 1.7: Milan's Area C 45
Figure 1.8: Seasonal pattern of Area C vehicle entries 46

Figure 2.1: Pollution changes, holding output fixed 88
Figure 2.2: Residual air emissions by distance from nearest non-attainment monitor 88
Figure 2.3: Event study estimates 89
Figure 2.4: PDF of NAICS6 coefficients 90

Figure 3.1: Seasonal sunset time 126
Figure 3.2: Vernal equinox, Mar 20 127
Figure 3.3: ATUS county-level geocoding 159
Figure 3.4: ATUS raw correlation: sleep and sunset time 160
Figure 3.5: Sleep seasonality 160
Figure 3.6: ATUS occupations do not exhibit seasonality 161
Figure 3.7: Seasonal estimate versus true estimate 167
Figure 3.8: Historical time zone sorting 173

LIST OF TABLES

Table 1.1: Descriptive statistics, daily level 36
Table 1.2: Weekday effect of Area C charge suspension on vehicle entries . . . 36
Table 1.3: Weekday effect of Area C charge suspension on sensor-level traffic volume, by distance outside Area C boundary 37
Table 1.4: Weekday effect of Area C charge suspension on portal-level vehicle entries, by public transit availability 38
Table 1.5: Price elasticity of Area C vehicle entries 39
Table 1.6: Weekday pollution effect of Area C charge suspension, by location . 40
Table 1.7: Weekday pollution effects of placebo suspensions 41
Table 1.8: Weekday effect of Area C charge suspension on portal-level vehicle entries, by public transit availability and congestion 44
Table 1.9: Comparison of traffic volume effects and elasticities to other empirical road pricing studies 47
Table 1.10: Comparison of Area C charge suspension to unpriced January 2012 interim period 48
Table 1.11: Weekday effect of unpriced January 2012 interim period on sensor-level traffic volume, by distance outside Area C boundary 49
Table 1.12: Weekday effect of unpriced January 2012 interim period on portal-level vehicle entries, by public transit availability 50
Table 1.13: Effect of 2011 placebo suspension on Area C vehicle entries 51

Table 2.1: Particulate abatement strategies 91
Table 2.2: Top ten industries, by TRI-reportable emissions 91
Table 2.3: Aggregated TRI emissions categories 92
Table 2.4: Monitor distance and emissions growth rates 92
Table 2.6: Effect on air emissions 93
Table 2.7: Effect on emissions ratios 94
Table 2.9: Effect on emissions ratios, by 2-digit NAICS code 94
Table 2.11: Leakage effect, within firm & 6-digit NAICS code 95
Table 2.12: Effect on emissions levels 96
Table 2.13: Placebo effect on emissions levels 96
Table 2.14: Placebo leakage effect, within firm & 6-digit NAICS code 97
Table 2.15: TRI *PM* descriptive statistics, by attainment status 101
Table 2.16: TRI *PM* descriptive statistics, by leakage dummy 102
Table 2.17: Historical CAA particulate standards 102
Table 2.18: Monitor distance and emissions levels 103
Table 2.20: General-equilibrium spillover test 104
Table 2.21: Effect on air emissions, intrafirm spillover controls 105
Table 2.22: Effect on toxicity-weighted air emissions 106
Table 2.23: Effect on air emissions, plants open 1993-2010 107
Table 2.24: Effect on air emissions, difference specification 108

Table 2.25: Effect on emissions ratios, by 3-digit NAICS code 109
Table 2.26: Intra-firm leakage effect on emissions ratios, within firm & 6-digit NAICS code 109
Table 2.27: Effect on toxicity-weighted emissions ratios 110
Table 2.28: Aggregate effect on emissions ratios 111
Table 2.29: Leakage effect, within firm & 2-digit NAICS code 112
Table 2.30: Leakage effect, continuous firm size controls 113

Table 3.1: ATUS Summary Statistics 133
Table 3.2: Short-run effects 136
Table 3.3: Robustness of short-run estimates: Controls 140
Table 3.4: Robustness of short-run estimates: Time controls and sample 141
Table 3.5: Long-run effects 142
Table 3.6: Robustness of long-run estimates: Controls 144
Table 3.7: Robustness of long-run estimates: Sample 146
Table 3.8: Effects on log median home value 149
Table 3.9: Effects on average wage 151
Table 3.10: Waking time use as a function of sunset time 152
Table 3.11: Waking non-work hours as a function of sunset time, selected groups 153
Table 3.12: Causal medical studies of sleep and performance 159
Table 3.13: QCEW summary statistics 161
Table 3.14: Short-run effects: Hourly workers 169
Table 3.15: Long-run effects: Hourly workers 170
Table 3.16: Modified Breusch-Pagan heteroskedasticity test 170
Table 3.17: Long-run effects, weighted and unweighted 171
Table 3.18: Robustness: County characteristics 172
Table 3.19: Hedonic robustness 174
Table 3.20: QCEW robustness 174

ACKNOWLEDGEMENTS

Thank you to Prashant Bharadwaj, Julie Cullen, Josh Graff Zivin, Mark Jacobsen, Craig McIntosh, Junjie Zhang, Andy Brownback, Gil Hersch, Sarojini Hirshleifer, Jamie Mullins, Arman Rezaee, Jeff Shrader, and participants in the UCSD environmental economics and applied economics seminars. Your questions and comments improved this work considerably. Thank you Kay Gibson, Terry Gibson, Julian Gibson, Alex Gibson, and Claudia Gibson Navarro for your support. Thank you Anne Lounsbery and James Kloppenberg, for teaching me to think carefully. Chapter 1, in full, has been submitted for publication of the material as it may appear in the Journal of Urban Economics, 2015, Gibson, Matthew; Carnovale, Maria, Elsevier, 2015. The dissertation author was the primary investigator and author of this paper. Chapter 2, in full, is currently being prepared for submission for publication of the material. Chapter 3, in full, is currently being prepared for submission for publication of the material. Gibson, Matthew; Shrader, Jeffrey. The dissertation author was the primary investigator and author of this material.

VITA

2001	Jacob Wendell Prize, Harvard University
2005	A. B. in History and Literature *magna cum laude*, Harvard University
2014	Clive Granger Research Fellowship
2015	Ph. D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Worker and Firm Responses to Environmental Policies

by

Matthew Gibson

Doctor of Philosophy in Economics

University of California, San Diego, 2015

Professor Julie Cullen, Co-Chair
Professor Mark Jacobsen, Co-Chair

This work examines responses to environmental policies. Treating workers and firms as optimizing agents, it derives theoretical predictions and evaluates them using data.

In Chapter 1, exploiting the natural experiment created by an unanticipated court injunction, we evaluate driver responses to road pricing. We find evidence of intertemporal substitution toward unpriced times and spatial substitution toward unpriced roads. The effect on traffic varies with public transit availability. Net of these responses, Milan's pricing policy reduces air pollution substantially, generating large welfare gains.

In addition, we use long-run policy changes to estimate price elasticities.

Chapter 2 examines the unintended consequences of pollution regulations. By regulating air emissions in particular counties, the Clean Air Act (CAA) gives firms incentives to substitute: 1) toward polluting other media, like landfills and waterways; and 2) toward pollution from plants in other counties. Using EPA Toxic Release Inventory data, I examine the effect of CAA regulation on these types of substitution.

Chapter 3 takes advantage of time zones, which influence worker sleep, to study the relationship between sleep and wages. Because sleep influences performance on memory and focus intensive tasks, it plausibly affects economic outcomes. We identify the effect of sleep on wages by exploiting the relationship between sunset time and sleep duration. Using a large, nationally representative set of time use diaries from the United States, we provide the first causal estimates of the impact of sleep on wages. A one-hour increase in seasonal weekly sleep increases a worker's wage by 1%. At the location level, a one-hour increase in long-run weekly mean sleep increases mean wage by 4.5%. Our results highlight the economic importance of sleep and pose potentially fruitful questions about the effects of time use on labor market outcomes.

These findings illustrate the richness of human responses to the new incentives and constraints imposed by environmental policies. They suggest ways in which efficient policies might account for such responses.

Chapter 1

The effects of road pricing on driver behavior and air pollution

Abstract

Exploiting the natural experiment created by an unanticipated court injunction, we evaluate driver responses to road pricing. We find evidence of intertemporal substitution toward unpriced times and spatial substitution toward unpriced roads. The effect on traffic varies with public transit availability. Net of these responses, Milan's pricing policy reduces air pollution substantially, generating large welfare gains. In addition, we use long-run policy changes to estimate price elasticities.

1.1 Introduction

Growing air pollution, congestion, and accident externalities from vehicle traffic have produced increasing interest in policy remedies. Beijing and Mexico City bar vehicles from their roads on some days based on their license plate numbers (Davis, 2008; Viard and Fu, 2014; Wang et al., 2014). Many German cities have created Low

Emissions Zones (Wolff, 2014), which prohibit dirtier vehicles within their borders. Stockholm, London, and Milan charge fees to enter congested downtown areas. In the US, the Department of Transportation is currently sponsoring a large number of road pricing experiments, including San Francisco's Golden Gate Bridge, Interstate 95 near Miami, SR520 near Seattle, and Interstate 35W near Minneapolis (DeCorla-Souza, 2004; Xie, 2013). Economists have raised concerns over non-price policies because behavioral responses can be so large that net policy benefits may be zero, or even negative (Davis, 2008; Gallego et al., 2013). Theory suggests that road pricing might be more efficient (Vickrey, 1963; Arnott et al., 1993), but this prediction depends on driver responses. On which margins do drivers respond to road pricing, and how large are such responses?

Confounding factors typically make traffic policies difficult to evaluate. Drivers know the policy start date well in advance and may begin to adjust their behavior beforehand, which will attenuate estimated effects. Municipalities typically increase public transit service at the same time they implement road pricing or a driving restriction. This makes it impossible to estimate the effect of the policy in isolation. For example, Eliasson et al. (2009) point out that Stockholm expanded bus service at the same time it implemented a congestion charge. Because the buses used for the expansion were older and dirtier, the reduction in emissions within the charge area was muted. Milan first implemented a congestion charge concurrent with, "traffic calming measures, new bus lanes, increased bus frequency, increases in parking restrictions and fees, and medium-term policies such as park-and-ride facilities and underground network extensions" (Rotaris et al., 2010).

To address these identification challenges, we exploit a natural experiment: in late July 2012, an Italian court unexpectedly suspended Milan's road pricing policy, called "Area C." The city reinstated pricing eight weeks later. Using unique traffic data at 15-minute resolution, our study examines behavioral responses to Milan's policy, which requires drivers entering the city center to pay d'5 on weekdays 7:30AM - 7:30PM. Drivers respond to pricing in two ways: 1) shifting trips to the unpriced period, just before 7:30AM or after 7:30PM; and 2) driving around the boundary of the priced area.

Net of these behavioral responses, we find the pricing policy decreases traffic by 14.5 percent and air pollution by 6 to 17 percent. The latter effect is large, particularly given that the priced area is small and Milan has an unusually clean vehicle fleet. We calculate that this pollution reduction increases welfare by approximately $3 billion annually. Routes without public transit experience large traffic changes from pricing, while those with public transit experience much smaller changes. We provide evidence that this surprising result may arise from residential sorting: residents who live near public transit may strongly prefer public transit. In addition, we use long-run changes in Milan's pricing policy to estimate elasticities: city-center entries by charged vehicles decrease .3 percent in response to a one percent price increase.

This study contributes to the empirical literature on second-best road pricing policies (Small et al., 2005; Small and Verhoef, 2007; Xie, 2013). Closely related to our analysis are Olszewski and Xie (2005), which analyzes the cordon charge and expressway pricing in Singapore, Santos and Fraser (2006) and Santos (2008) on the London cordon charge, and Eliasson et al. (2009) on the Stockholm cordon charge. These studies find cordon charges do reduce traffic within the priced area. Also related are Foreman (2013) and Small and Gomez-Ibanez (1998), which find evidence of intertemporal substitution in response to time-varying tolls. Our work complements the theoretical literature on second-best road pricing (Lévy-Lambert, 1968; Marchand, 1968; Verhoef et al., 1996), particularly the literature on cordon charges (Mun et al., 2003; Verhoef, 2005). Finally, we contribute to the literature on environmental effects of traffic policies. Many such studies have found no evidence of air quality improvements (Transport for London, 2005; Transport for London, 2008; Invernizzi et al., 2011). Authors commonly attribute this to driver substitution behaviors or exploitation of policy loopholes (Davis, 2008; Gallego et al., 2013). In important work, Wolff (2014) finds that German Low Emissions Zones reduce the concentration of particles with a diameter of 10 microns or less (PM10) by approximately 9 percent; this study is particularly significant given efforts by European cities to meet stringent air quality standards.

Our study is unique in obtaining unconfounded causal estimates of behavioral responses to road pricing and net road pricing effectiveness. This is the first analysis to examine removal, rather than imposition, of a traffic policy. Other studies have used indirect measures of traffic (such as gasoline sales or vehicle registrations) or hourly vehicle counts, but to the best of our knowledge ours is the first to combine direct, high-resolution measures of traffic volume with air pollution data. Finally, our finding that the net effect of pricing varies with public transit availability is novel. It contributes to the literature on public transit and air quality (Friedman et al., 2001) and adds a new dimension to the literature on traffic policies.

The remainder of the paper proceeds as follows. Section 2.2 provides policy background and describes the natural experiment. Section 2.4 covers data, Section 2.5 describes our estimating equations, and Section 2.6 discusses results. Section 2.8 concludes.

1.2 Background

Located in the center of Milan, Area C includes approximately 8.2 square kilometers (4.5 percent of city land area) and 77,000 residents (6 percent of population). The boundary follows the *Cerchia dei Bastioni,* the route of the walls built under Spanish control in 1549. Many of the portals still stand today, though the walls are largely gone. Figure 1.7 illustrates the area.

Milan provides high levels of public transit, including four subway lines, 19 tram lines, 120 bus lines, and 4 trolley lines. Together these lines transport 700 million passengers across 155 million kilometers per year. The 80-kilometer subway network is larger than all other Italian subways combined (Azienda Transporti Milanesi, 2013). Public transit has a 41 percent mode share in the city, followed by cars at 30 percent, walking at 17 percent, bicycles at 6 percent, and motorbikes at 6 percent (Martino, 2012). The average round-trip commute in Milan takes 53 minutes, comparable to US cities like Dallas (52 minutes), Seattle (55 minutes), and Los Angeles (56 minutes; Toronto Board of Trade, 2011).

Milan is one of the most polluted large cities in Europe. From 2002 through 2010 the city exceeded the EU standard for PM10 on an average of 133 days per year (Danielis et al., 2011). Since the mid 1990s the city has experimented with traffic policies intended to curb its air pollution problem. Milan's first major road pricing program, called Ecopass, ran from January 1, 2008 to December 31, 2011. Drivers paid a fee to enter Area C that varied with the emissions from their vehicles. Vehicles meeting the Euro 3 standard paid nothing, while the dirtiest diesel vehicles paid d'10.[1] The charge applied weekdays 7:30AM-7:30PM. Drivers could pay by internet, phone, or at the bank. The city enforced the charge using license plate-reading cameras located at the 43 entrances to Area C (Danielis et al., 2011). Drivers who entered without paying faced fines of d'70-d'275 (la Repubblica, 2008). Approximately 2 percent of entering vehicles each day incurred fines (Martino, 2012).

In June 2011 the voters of Milan overwhelmingly approved continued road pricing, with 79 percent in favor (Danielis et al., 2011).[2] As of January 16, 2012, the

[1] Vehicles built prior to imposition of EU emissions standards were prohibited from October 15 through April 15. Drivers received a 50% discount on the first 50 entries and a 40% discount on the next 50 entries. Residents of Area C were also eligible for discounts (Rotaris et al., 2010).

[2] 49 percent of voters participated. The referendum did not specify the exact form the continued program would take.

city implemented a d'5 congestion charge for most vehicles entering Area C weekdays 7:30AM-7:30PM. This policy was named Area C.[3] Motorcycles and public vehicles (e.g. ambulances) were exempted.[4] Administrative details were largely the same as those for Ecopass. Drivers gained the option to pay by direct debit, using a radio reflector placed in the vehicle (similar to FasTrak or E-ZPass in the US). Violators were fined d'87 (Carra, 2012).

On July 25, 2012, a court unexpectedly suspended the Area C congestion charge in response to a lawsuit by Mediolanum Parking (Povoledo, 2012). More than ten previous lawsuits against Ecopass and Area C had failed, so the suspension provoked surprise from the press (Carra and Gallione, 2012). Charge enforcement halted the next day, July 26. There was no press coverage prior to the court injunction, suggesting the decision was completely unanticipated. The duration of the suspension was unknown and some observers believed it would be permanent (Carra, 2012). Political forces marshaled on both sides. The mayor declared, "We will save Area C." Meanwhile the opposition called suspension the "death" of Area C, "the defeat of ideological fervor and the victory of Milan's productivity and good sense" (Carra, 2012). The city altered neither public transit service nor parking fees in response to the injunction. On September 6, the city announced the charge would be reinstated as of September 17, 2012.[5] For a timeline of these events, see Figure 1.1.

[3]Vehicles classified diesel Euro 3 or below, or gasoline Euro 0 or below, were prohibited. Private vehicles over 7m long were also prohibited. Scooters, motorcycles, and alternative-fuel vehicles, including hybrids, were exempted. Residents paid d'2 per entry (City of Milan, 2012; Milan Tourism, 2012).

[4]This category includes mopeds and powered scooters.

[5]The reinstated charge now ends at 6PM on Thursdays, rather than at 7:30PM as before (Corriere della Sera, 2012a). Other features are unchanged.

1.3 Data

Our traffic data come from AMAT and the Settore Pianificazione e Programmazione Mobilità e Trasporto Pubblico Comune di Milano. For Area C, we have entries by vehicle type and entry portal at 15-minute resolution, 2008-2012. There are 43 entry portals. These data are recorded by the license plate cameras used to enforce the Area C charge. In addition, we have counts of passing vehicles at 15-minute resolution, 2008-2012. These data are measured by 748 buried sensors, mostly outside Area C.[6] Table 1.1 reports descriptive statistics for both data sets at the daily level (aggregating over sensors/cameras and 15-minute intervals).

Our pollution and weather data come from ARPA Lombardia, the provincial air quality agency. We have pollution and weather data at the monitor level, from 2003 through February 2013. Measured pollutants include carbon monoxide (CO), particles 10 microns or less in diameter (PM10), and particles 2.5 microns or less in diameter (PM2.5). CO is measured hourly, while particulates are measured daily. There are eight pollution measurement stations in the city of Milan proper (see Figure 1.2), of which two are inside Area C. The number of monitors varies by pollutant and over time, as not all stations monitor all pollutants.

Table 1.1 provides descriptive statistics at the monitoring station-day level. The rightmost column includes EU pollution standards for comparison. The European Commission (EC) has the power to levy large fines against non-attainment cities. For example, the EC fined Leipzig d’700,000 per non-attainment day for failing to meet the PM10 standard (Wolff and Perry, 2010).

[6] According to AMAT, the buried sensors are less accurate than the cameras. Neither buried sensor data nor camera data are available prior to 2008 (the cameras had not yet been installed and activated).

1.4 Estimation

To explore the effect of policy suspension on traffic we estimate a series of equations within the following framework:

$$traffic_t = \beta * suspension_t + \lambda * suspension_t * wkend_t + \bar{\gamma} * \overline{\mathbf{time_t}} + \bar{\theta} * \overline{\mathbf{trend_t}} + \bar{\eta} * \overline{\mathbf{weather_t}} + \varepsilon_t \quad (1.1)$$

The $traffic$ variable measures either Area C entries or passing cars, over a day or a 15-minute period, with t indexing days. The $\overline{\mathbf{time_t}}$ vector includes dummies for year, month, week, weekend, day of week, and holidays, plus interactions of weekend with year. In addition, it includes dummies for the two-week interim period between Ecopass and Area C and the interaction of the interim period with weekend.[7] (While the interim period is non-random, we briefly analyze it in Section 1.5.5.) In our primary results below we report estimates using a 7th-degree trend in date, following Davis (2008). Weather controls comprise ten-piece linear splines in temperature and positive precipitation. We control for weather because it plausibly influences the choice of public versus private transportation, or car versus motorcycle. The *suspension* variable is a dummy equal to one for the period when the charge was suspended. The error term ε includes shocks to traffic not captured by our controls, for example, an unusually bad auto accident or the Pope's visit on June 2, 2012. In this and all subsequent equations, the coefficient of interest is β, the weekday effect of charge suspension. The weekend effect $(\beta + \lambda)$ is generally not statistically different from zero, suggesting limited scope for weekday-weekend substitution, so we do not report it in the estimation results. Weekend observations are still used in all of our estimation, however, as long-run trends may

[7] We do not explicitly control for Ecopass because of the year dummies 2008-2011.

influence both weekday and weekend traffic.

The key identifying assumption underlying both equation 1.1 and subsequent models is the exogeneity of the *suspension* variable. That is, we assume that conditional on our rich seasonal and weather controls, the timing of charge suspension is unrelated to other determinants of traffic volume and pollution. This is reasonable because the charge was suspended unexpectedly by a court, as discussed in Section 2.2.

For the analysis of spatial substitution, we estimate two panel models at the sensor-day level, with sensor fixed effects (FE). The first specification is as follows (*s* indexes sensor):

$$traffic_{st} = \bar{\beta} * suspension_t * \overline{\mathbf{distance_s}} + \bar{\lambda} * suspension_t * wkend_t * \overline{\mathbf{distance_s}} \\ + \overline{\alpha_s} + \bar{\gamma} * \overline{\mathbf{time_t}} + \bar{\theta} * \overline{\mathbf{trend_t}} + \bar{\eta} * \overline{\mathbf{weather_t}} + \varepsilon_{st} \quad (1.2)$$

In equation (3) $\overline{\mathbf{distance_p}}$ is a vector of dummies for sensors in several distance bins, where distance is measured from the outside of the Area C boundary. The second specification is similar, but instead of grouping sensors by distance, we group them into ring and non-ring roads (described in more detail in Section 2.6).

To analyze heterogeneity by public transport availability, we estimate a panel version of equation 1.1 with portal fixed effects (*p* indexes portal):

$$traffic_{pt} = \bar{\beta} * suspension_t * \overline{\mathbf{pubtrans_p}} + \bar{\lambda} * suspension_t * \overline{\mathbf{pubtrans_p}} * wkend_t \\ + \overline{\alpha_p} + \bar{\gamma} * \overline{\mathbf{time_t}} + \bar{\theta} * \overline{\mathbf{trend_t}} + \bar{\eta} * \overline{\mathbf{weather_t}} + \varepsilon_{pt} \quad (1.3)$$

In the equation above, $\overline{\mathbf{pubtrans_p}}$ is a vector containing a dummy for the presence of public transit, and another for the absence of public transit. We also estimate versions of the model comparing portals with and without bus, tram, and metro service. [8]

To investigate the effect of suspension on daily average pollution we estimate the following equation:

$$\begin{aligned}\ln(avg\_pollution)_t &= \beta * suspension_t + \lambda * suspension_t * wkend_t \\ &+ \bar{\gamma} * \overline{\mathbf{time_t}} + \bar{\theta} * \overline{\mathbf{trend_t}} + \eta * \ln\left(avg\_pollution_{t-1}\right) + \bar{\delta} * \overline{\mathbf{atmosphere_t}} + \varepsilon_t \qquad (1.4)\end{aligned}$$

The dependent variable is the log average level of a pollutant measured over a day, with t indexing days. We conduct the analysis in logs to make the estimates for different pollutants more easily comparable. To avoid the endogeneity problems that arise in a dynamic panel specification, we average over monitors and estimate the model separately for each pollutant and area of Milan.[9] In order to control for the persistence of pollutants emitted on the previous day, we include one lag of the dependent variable. The lagged pollution variable also controls for the previous day's atmospheric conditions, avoiding the need for functional form assumptions on lagged atmospheric variables. ARPA normalizes the pollution measurements for temperature and pressure. The vector $\overline{\mathbf{atmosphere_t}}$ includes 4-knot cubic splines in humidity, wind speed, solar radiation, and precipitation, plus a dummy for positive precipitation. As in equation 1.1, our specification also includes a 7th-degree trend in date and time dummies.

[8] For example, a given portal will have *bus* equal to 1 if a bus line crosses the boundary of Area C through that portal. This is a simplification that ignores the effect of being *near* (but not *on*) a bus line. If the two effects have the same sign, as is plausible, this specification will bias us against finding a difference between portals with and without a bus line. Similarly, the $tram$ and $metro$ variables equal 1 only if the mode in question passes directly through or beneath the portal in question.

[9] Estimation results from a dynamic panel specification are available upon request. They are extremely similar, as the asymptotic bias is of order $^1/_T$ and our data contain thousands of days.

1.5 Results

1.5.1 Traffic

We first provide some semi-parametric evidence on the effect of charge suspension for vehicle types subject to the charge (buses and motorcycles are excluded). Figure 1.3 plots the residuals from a daily model that omits the *suspension* variable. We fit separate degree-zero local polynomials for the period June-July 2012 (charge), August-September (no charge), and October-November (charge). The graph shows a sharp increase in weekday entries into Area C upon charge suspension, consistent with a surprise announcement.[10] All three fitted lines are flat; there is no evidence of a seasonal trend in the residuals before, during, or after charge suspension. This indicates that our time fixed effects, together with a polynomial trend in date, are effectively controlling for seasonal patterns that might otherwise bias our estimates. There are several large positive residuals between the Sept. 17 reinstatement of pricing and Oct. 1. This may reflect commuters delaying a mode switch before purchasing an October public transit pass.

Table 1.2 records results from our linear model for all vehicles, charged vehicles (buses and motorcycles excluded), motorcycles (including mopeds and scooters), and other vehicles (primarily police cars and ambulances, which are exempt from the charge). Charge suspension results in approximately 27,000 additional entries per day and the estimate is statistically significant at the one percent level. This represents an increase of approximately 14.5 percent. The composition of entries also changes. Entries by charged vehicles increase by roughly 29,000 while entries by motorcycles, which are exempt from the Area C charge, fall by roughly 2000. The latter result is not statistically significant. The estimate for other vehicles, predominantly public vehicles like police cars, is small in magnitude and not statistically distinguishable from zero at any conventional significance

[10] The large positive residuals in the October-November period correspond to weekends. Interacting the weekend dummy with month did not appreciably reduce the magnitude of these residuals.

level. This provides a placebo test, as drivers of public vehicles are exempt and so do not face a price change from charge suspension. We employ Newey-West standard errors to account for autocorrelation in ε_t out to seven lags. Because of occasional missing data, these standard errors will be biased slightly downward. We have also estimated our models with standard errors clustered at the week level and the results (available upon request) are not meaningfully different. Note that the assumption of independence across clusters fails for days near the boundary of a week, so this is not our preferred method of estimating standard errors.

Table 1.9 compares our estimated change in entries by charged vehicles (expressed as a percentage) to results for 10 other pricing policies. Eight find effects broadly comparable in magnitude, ranging from -3 percent to -22 percent. These include the London (-18 percent) and Stockholm (-22 percent) cordon charges studied by Santos (2008) and Eliasson et al. (2009), respectively. Singapore's central Restricted Zone yields two appreciably larger estimates, -44 percent and -52 percent. Small and Verhoef (2007) suggest this policy produced such a dramatic response because the charge was initially set extremely high.

To examine intertemporal substitution, Figure 1.4 plots the coefficients from a series of 96 regressions, with each 15-minute interval of the day modeled separately.[11] The estimates show intertemporal substitution in both the morning and the evening. Charge suspension results in approximately 500 fewer entries in the 15 minutes just before the charge begins at 7:30AM and just after it ends at 7:30PM. (This indicates that under the charge, drivers were shifting trips into these periods.) Indeed in the morning the negative estimates are statistically distinguishable from zero (at the 5 percent level) for the entire hour 6:30-7:30AM. Charge suspension increases entries during the

[11] We use Newey-West standard errors to account for serial correlation. For most 15-minute intervals, serial correlation falls to near zero after 7 lags. For the period 11:30PM-5:15AM, however, there are spikes in serial correlation at 14, 21, and 35 days. We hypothesize that this results from the preponderance of public and commercial vehicles during this window.

7:30AM-7:30PM period, consistent with the daily average estimates reported in Table 1.2. The increases achieve local maxima just after 7:30AM and before 7:30PM, suggesting intertemporal substitution by commuters. The hours 9AM-3PM, however, see roughly uniform increases in traffic under charge suspension. This indicates that non-commuters comprise a large share of marginal drivers.

Such a pattern of responses is the inverse of what is often called "peak spreading." In theory peak spreading affects driver welfare through two channels: 1) by reducing trip duration; and 2) by rescheduling trips (Arnott et al., 1993; Lindsey and Verhoef, 2000). The former welfare effect is positive, but the sign of the latter is theoretically ambiguous. Spreading the peak traffic load increases aggregate schedule delays[12], but pricing better aligns trip times with drivers' values of schedule delay (Arnott and Kraus, 1998; Lindsey and Verhoef, 2000). Because most air pollution emitted by vehicles is persistent within a day (Seinfeld and Pandis, 2012), peak spreading may not change welfare along this dimension.

Finally we investigate spatial substitution toward roads outside Area C. Table 1.3 presents results from a panel model at the sensor-day level, estimated from the buried sensor data. (Note these data measure passing cars per unit time and the resulting estimates are not directly comparable to those from camera data.) Traffic at the average sensor increases approximately 8 percent (469 vehicles per day) and the estimate is statistically significant at the one percent level. This overall result conceals an interesting spatial pattern. Consistent with the models based on camera data, suspension of the charge increases traffic inside Area C. Traffic on the roads within 1km outside the Area C boundary, however, decreases by approximately 18 percent. This estimate is significant at the ten percent level. Both point estimates for roads more than 2km outside the boundary are positive, with one statistically significant at five percent and the other not significant.

[12] In keeping with the theoretical literature, by "delay" we mean a deviation from the desired arrival time, either earlier or later.

This is consistent with an increase in radial trips (e.g. commutes from a residential neighborhood into the center) from charge suspension. Overall this pattern of results suggests that some drivers respond to the charge by driving around Area C. For drivers seeking to avoid the priced area, the natural route typically involves the Circonvallazione Esterna, a ring of larger roads located .6km-2km outside the Area C boundary. Table 1.3 shows the estimated effect of charge suspension on these roads is large, negative, and significant at the five percent level. Some of this decrease may reflect reduced circumferential commuting to public transit stations. Evaluating this type of spatial substitution has proved difficult in other settings due to confounding factors. In London, for example, the city substantially improved ring-road infrastructure because Transport for London anticipated spatial substitution (Santos, 2004). To the best of our knowledge, ours is the first study to recover an unconfounded driver response on this dimension.

In interpreting these results, it is reasonable to ask whether they capture the short-run response to a pricing holiday or a long-run response. The initial six-month trial of the Stockholm cordon charge provides some evidence on this point. Eliasson et al. (2009) observe, "...there was some doubt as to whether any traffic reduction would actually take place during a brief and transient trial. Could it be that people would decide to 'sit out' the trial period without changing their travel habits? We now know that the trial indeed had an immediate effect." Effects from Stockholm's initial trial proved very similar to long-run effects (Börjesson et al., 2012). In Milan the suspension of the Area C charge was widely publicized, so the vast majority of residents knew about the change. Evidence on residents' expectations is qualitative and limited. Press accounts suggested Area C pricing might not return (Corriere della Sera, 2012b), but there was likely a range of beliefs about this. The key question is not, however, whether residents expected the suspension to be permanent, but whether they behaved as though it were. We have some suggestive evidence on this point from Figure 1.3. If residents exhibited habit persistence or slowly updated their beliefs about the suspension, we would expect an upward trend in the residuals during the suspension period. No such trend is apparent; instead the full magnitude of driver responses emerges immediately. This pattern also implies that switching costs across modes, routes, and travel times are not first-order considerations for marginal drivers. Therefore we believe our estimates largely reflect long-run behavior.

There is one respect in which the observed behavior during suspension likely does not correspond to long-run behavior: vehicle portfolios. Intuition suggests they adjust slowly. Given the possibility of renewed pricing, a risk-averse resident of Milan might well have been reluctant to purchase a new vehicle in response to charge suspension. If vehicle portfolios constrained a reasonably large fraction of the population, our estimates represent lower bounds on the magnitude of long-run effects. While our use of a natural experiment does incur this cost on the external validity dimension, it brings benefits on the internal validity dimension (e.g. avoidance of policy endogeneity concerns).

It is impossible to conduct a full welfare analysis of the Area C policy using our data, but previous work casts light on some of its efficiency properties. Verhoef (2005) studies a cordon charge in a general-equilibrium model of a monocentric city with endogenous population density. He finds the optimal cordon location is at 22 percent of the distance from the city center to the city limits. We can compare the location of Milan's cordon to this benchmark. Like many older European cities, Milan is monocentric. From the land area of Area C, we can calculate an idealized radius of $\sqrt{\frac{8.2km^2}{\pi}} = 1.62$ kilometers. Proceeding in like fashion for the city limits, we obtain an idealized radius of $\sqrt{\frac{181km^2}{\pi}} = 7.59$ kilometers. Dividing yields a ratio of $\frac{1.62km}{7.59km} = .21$. This suggests that at minimum Milan's cordon is not badly located. We can also evaluate the level of Milan's charge. Mun et al. (2003) simulate a cordon charge using data from Osaka and find the optimal charge is equivalent to roughly 30 minutes' worth of labor income. In Milan this would be about d'9. While Milan differs from Osaka on many dimensions, it is possible that a charge modestly above the current d’5 level would increase welfare. Verhoef (2005) finds that an optimal cordon charge achieves 88-90 percent of the gains from a first-best pricing policy. De Borger and Proost (2001) find an optimal combination of a cordon charge and parking fees can achieve 70 percent of first-best gains. Taken together, this body of research implies that the Area C pricing policy may be reasonably efficient.

1.5.2 Interaction with public transit

In addition, we investigate the interaction of charge suspension and public transit availability. To that end we estimate a panel model with a portal-day as the unit of observation. The results in Table 1.4 indicate commuters on routes with public transit available respond much less to the suspension of the charge. Portals on a metro line, for example, show a response that is not statistically distinguishable from zero.[13]

There are at least two plausible explanations for these results. The first relies on cost differences. Assume an identical distribution of preferences for driving on two routes, one with public transit ("Route A") and one without ("Route B"). If a sorting equilibrium holds, commuters on the two routes must achieve equal utility. This implies that if Route A has cheap public transit, it must have expensive car travel. This could be a direct result of public transit, as when road lanes are devoted to tram lines, or a product of transit planning, as when metro lines are placed beneath more congested roads. If a city applies the same charge to cars on both routes, the percentage price change for Route A is much smaller and theory predicts a smaller traffic response.

Alternatively, the results in Table 1.4 could spring from residential sorting (preference heterogeneity). Suppose people with strong preferences for public transit live near

[13]Portals with public transit may still be generating welfare changes if the composition of traffic is changing, but we cannot evaluate this with our data.

Route A. They might not own cars. They might, for example, dislike the claustrophobic conditions that prevail on buses and trains at rush hour. Such individuals might be relatively unresponsive to changes in the price of driving. Assume the initial cost of driving is the same for both routes. Then for a given road price change, there will be more infra-marginal drivers on Route B than Route A.

It is difficult to choose between these explanations using the available data, but Table 1.8 provides suggestive evidence. When we interact charge suspension with a time-invariant measure of rush-hour congestion,[14] the response is larger for congested portals without public transit than for uncongested portals without public transit. (Congestion does not matter for portals with public transit.) This result is inconsistent with an explanation based *solely* on cost differences, which would predict smaller responses on more congested routes. It provides some evidence of preference heterogeneity, but does not exclude the possibility that cost differences drive some of the responses to charge suspension. For further discussion, see Appendix Section 1.8.1.

1.5.3 Price response

Milan's sequence of traffic policies, including both Ecopass and Area C, presents an opportunity to recover another important feature of driver behavior: price responsiveness. Under Ecopass, the weighted average weekday price for passenger vehicles was approximately d'0.72 (author's calculation, based on Rotaris et al. 2010). The Area C policy increased the weekday price to d'5. This provides potentially exogenous price variation, although the usual concerns about policy endogeneity obtain. We estimate the elasticity of vehicle entries with respect to price using a variant of equation 1.1, replacing

[14]We standardize entries at the portal-15 minute-lane level, then average the resulting values during rush hour periods (over days). Portals with lower values (low rush hour throughput relative to the portal average) are likely more congested. We define a congestion dummy equal to 1 for portals in the bottom 10 percent of the distribution.

the dependent variable with log entries and the policy variables with log price. Table 1.5 reports results. Overall a one percent price increase decreases entries by charged vehicles by .3 percent. This estimated response captures the net effect of two changes: the increase in pecuniary cost and the decrease in time cost (from reduced congestion) under the Area C policy. Given that the Area C charge is a relatively small part of total trip cost, which includes time, fuel, and depreciation, this demand response is large. For passenger and commercial vehicles elasticities are -.17 and -.47, respectively, with both estimates significant at the five percent level.[15] The estimate for other vehicles is a placebo test estimating the effect of passenger prices on entries by exempt municipal vehicles; we find no effect. Our elasticity estimates are necessarily local and may not obtain outside the range of prices observed in our data. They suggest, however, that a modest price increase, e.g. from d'5 to d'6, might produce substantial additional reductions in Area C entries.

If the change in price from Ecopass to Area C is conditionally exogenous, our estimate captures an internally valid causal effect. The question of external validity remains, however. Theory predicts that demand elasticity will vary with income, preferences, the availability of substitutes, and other factors. To provide a qualitative sense of such factors, Table 1.9 puts our estimated elasticity in the context of estimates from other locations. At -.3 our overall estimate is modestly larger than most previous findings for cordon charges, which generally range from -.2 to -.1. Estimates for expressway and bridge tolls exhibit more variation, ranging from -.56 to -.06. While some are similar to our result (e.g. Small et al. (2006) for California State Route 91), others are substantially larger or smaller. For example, Odeck and Brathen (2008) find an average price elasticity of -.56 on Norwegian toll roads. Such larger estimates may reflect the often-greater availability of close substitutes for single-facility tolls than for cordon charges.

[15]We do not separate passenger and commercial vehicles in our primary analysis because under the Area C policy they both face the same d'5 price.

Some cities manipulate parking prices, rather than road prices, in order to optimize travel demand. Studies of such policies typically estimate parking demand, rather than demand for travel on a given road or demand for cordon crossing. While these estimates are not directly comparable to ours, they may provide an instructive benchmark. In a survey of research findings, the US Transit Cooperative Research Program found a range of parking demand elasticities from -.6 to -.1, with -.3 the mode (Vaca and Kuzmyak, 2005). Using data from exogenous changes in San Francisco parking prices, Pierce and Shoup (2013) estimate an average demand elasticity of -.4.

1.5.4 Pollution

Table 1.6 reports the pollution effect of charge suspension on weekdays, estimated using equation 1.4. We focus on CO, PM10, and PM2.5 because these pollutants have direct, negative health effects (Seaton et al., 1995) and all are closely associated with vehicle emissions (Gallego et al., 2013). Estimates show statistically significant increases in CO and PM10, both inside and outside Area C, in the 6 to 17 percent range. These magnitudes are similar to those from our traffic models. The point estimate for PM2.5 outside Area C is greater at 21 percent, but the standard error is large and the estimate is significant only at the ten percent level. This imprecision may stem from the much shorter period over which PM2.5 data have been collected. For CO we can also estimate the effect for monitors located on the ring roads (Circonvallazione Esterna). This estimate is near zero, which roughly accords with our traffic results in Table 1.3. As the half-lives of commonly regulated air pollutants are measured in hours or days (Seinfeld and Pandis, 2012), the observed pollution increases likely derive from additional trips and mode shifting, rather than from trip rescheduling.

These pollution effects are large, particularly given that the priced area is small (5 percent of the city) and Milan has an unusually clean vehicle fleet. Milan's earlier

Ecopass policy, which applied from 2008 through 2011, created incentives for drivers to purchase cleaner vehicles and many did so (Rotaris et al., 2010). This means that for a given number of foregone trips, the effect on pollution would have been smaller in 2012 than in 2007. Like our traffic estimates, our pollution estimates are lower bounds on long-run effects because of the potential for vehicle portfolios to change over the long run.

In order to evaluate the welfare effects of these air pollution changes, we require an estimate of willingness to pay for reductions in PM10 pollution. We adopt estimated annual willingness to pay of \$148.70 per person per $\mu g/m^3$ (in 1982-1984 dollars) from Bayer et al. (2009), who use data from US metropolitan statistical areas. By accounting for migration costs and instrumenting for ambient pollution, this study overcomes several important identification challenges. For comparison, note that the meta-analysis by Smith and Huang (1995) finds a mean marginal willingness to pay of \$110 (in 1982-1984 dollars) per $\mu g/m^3$ TSP reduction in US cities. While this estimate is meaningfully smaller than the one from Bayer et al. (2009), this is unsurprising for two reasons: 1) the downward biases in the OLS hedonic specifications analyzed by Smith and Huang (1995); and 2) the higher real income of the US population in the data used by Bayer et al. (2009). The rough similarity between the Bayer et al. (2009) and Smith and Huang (1995) estimates provides some reassurance that our choice is reasonable.

Naturally the use of a willingness to pay estimate from the United States raises benefit transfer concerns. Kaul et al. (2013) find that transfer errors are typically smaller for function transfers than for value transfers. While a full function transfer is beyond the scope of the present exercise, we can scale the Bayer et al. (2009) estimate to account for local income in Milan. As suggested by Ready and Navrud (2006), we employ a PPP-adjusted exchange rate and find that average income in Milan was roughly 85 percent of US income in 2007 (Hammitt and Robinson, 2011; OECD, 2011). Both Smith and Huang (1995) and Hammitt and Robinson (2011) find that the income elasticity of willingness to pay for air pollution reduction is small in wealthier nations, with the latter noting that US agencies often use an income elasticity of approximately .5. Multiplying $\$149 * (1 - (.15 * .5))$ yields approximately \$138. Converted to 2014 dollars, this becomes \$327.

With this figure in hand, we can compute the aggregate welfare effects of the PM10 changes in Milan from the Area C policy. Mean PM10 concentration in our data is $48 \mu g/m^3$ inside Area C, $44 \mu g/m^3$ outside. The concentration changes implied by our estimates are $1.9 \mu g/m^3$ and $7.5 \mu g/m^3$, respectively. Approximately 77,000 people live in Area C and 1.2 million outside. The implied welfare gain from the Area C policy is approximately \$48 million inside Area C and \$2.94 billion outside, for a total of \$3 billion. This estimate is very large relative to the annual PM10 benefit figures used by transportation researchers in cost-benefit analysis of the Ecopass policy, which have typically been in the range of d'0.4-1.3 million (Rotaris et al., 2010; Danielis et al., 2011). We note that the median transfer error identified by Kaul et al. (2013) is 39 percent, and this does suggest some caution. Even allowing for the possibility of large transfer error, however, our welfare estimate is an order of magnitude larger than those in Rotaris et al. (2010) and Danielis et al. (2011).

The finding that pricing reduce air pollution both inside and outside Area C speaks to an important distributional question. Opponents of the Area C policy have claimed that it improves air quality in an affluent area while doing nothing to address the remainder of the city (Danielis et al., 2011). The estimates in Table 1.6 provide evidence against this claim. While spatial substitution may reduce air quality near ring roads, the policy improves air quality in other locations, both inside and outside Area C. Voting behavior is consistent with such a widespread improvement. In a 2011 referendum on road pricing, more than 79 percent voted in favor (Danielis et al., 2011), an outcome that would be unlikely if only Area C residents benefited from the policy. The successful referendum is somewhat surprising in view of the generally unfavorable public attitudes toward road pricing. As in Stockholm, residents experienced the benefits of road pricing during a trial period before voting and this may have been influential (Eliasson, 2008; Harsman and Quigley, 2010).

1.5.5 Robustness checks

Traffic

We estimated all models with the following trends in date: 1) no trend; 2) linear trend; 3) 4th-degree trend; 4) 7th-degree trend. In nearly all cases the choice of trend had negligible influence on the sign, magnitude and significance of the estimates.

In addition, we compare our primary estimates to those from the interim period between the end of the Ecopass policy and the start of the Area C policy (January 1-15, 2012). During this time drivers could enter the city center without paying, but this period raises identification concerns. First, it was not randomly timed and therefore the potential for unobserved confounders (like changes in bus service) is greater than for the period of our natural experiment. Second, because the return of pricing was assured, questions of habit formation and switching costs are more problematic. Third, because this period lasted only two weeks, these models have less statistical power. Nonetheless the interim period provides a rough benchmark against which to evaluate our main results. Table 1.10 shows that the estimated effects on Area C entries for all vehicles and charged vehicles are similar in magnitude and statistically significant at the one percent level. The estimated effect on motorcycle entries is positive and significant for the interim period, which differs from our primary result and does not accord with theory. The positive sign could reflect the fact that riding a motorcycle in a lower-traffic environment is both safer and more fun.

Table 1.11 reports spatial substitution results for the interim period. The pattern of results accords with those from our natural experiment, but the decrease in traffic on ring roads is no longer statistically significant. Similarly, Table 1.12 shows how effects on Area C entries during the interim period vary with public transit availability. (These estimates come from the same model as those in Table 1.4.) Again the pattern of results is strongly similar to those from our natural experiment, with portals lacking public transit seeing greater increases in traffic.

Taken together, Tables 1.10, 1.11, and 1.12 demonstrate that estimates from the interim period (in January) are quite similar to those from our natural experiment (July-September). This pattern suggests that seasonality in the elasticity of demand for Area C trips is not a first-order concern. Thus our primary estimates provide evidence on driver responses to pricing that generalizes beyond the time of year at which the natural experiment happened to occur. There remains the possibility of bias from seasonal trends in the level of demand. As discussed in Section 1.5.1 and illustrated in Figure 1.3, our time fixed effects and polynomial trend in date appear to effectively control for such trends. Nonetheless we describe seasonal trends in more detail here. Figure 1.8 shows the seasonal pattern of entries into Area C. The period of the natural experiment includes Italy's traditional vacation season, which sees far fewer Area C entries in three August weeks. The five remaining weeks of the experiment, however, include some of the busiest weeks of the year (in September). Given the pattern in Figure 1.8, any failure of our seasonality controls will bias the magnitude of our estimates downward.

Figure 1.5 displays the "effect" of a placebo charge suspension each year 2008-2011 on Area C vehicle entries. There is no evidence of an increase in Area C entries during the placebo periods; if anything they show slight decreases. Similarly, Table 1.13 reports estimated effects of a placebo suspension for the same dates in 2011 (rather than

2012). Estimated magnitudes are much smaller than those in our main results and not one is significant.

Pollution

Table 1.7 reports the estimated effects of placebo suspensions for the same dates 2008-2011 (rather than 2012). Half the estimates are negative and most are statistically insignificant, which aligns with the placebo tests from our traffic models and suggests that our main results are not driven by misspecification. The estimated placebos for CO inside Area C (2011) and PM10 outside Area C (2009) are positive and statistically significant, which recommends some caution in interpreting our corresponding primary estimates.

1.6 Conclusion

Our analysis uses a natural experiment to examine behavioral responses and recover causal effects of Milan's Area C road pricing policy. We find the policy reduces traffic and pollution considerably. Drivers respond with intertemporal substitution toward unpriced times and spatial substitution toward roads outside the charge area. In addition, we show that the effect of pricing on traffic depends on the availability of public transportation. Routes without public transit experience large traffic changes from the Area C charge, while those with public transit experience much smaller changes. We also use long-run changes in Milan's pricing policy to estimate elasticities of traffic with respect to the charge: entries by charged vehicles decrease .3 percent in response to a one percent price increase. This estimate captures the net effect of an increase in the charge and the resulting decrease in time cost from reduced congestion.

Our findings are relevant for policy design. Theory predicts that the substitution behaviors we observe would occur under both optimal and second-best policies, but cities can tailor policy to manage their magnitudes. Cities like Milan, with fixed cordon charges, might reduce intertemporal substitution and move closer to the theoretical optimum by charging a lower but non-zero price for "shoulder" periods adjacent to peak periods. Some drivers might still choose the shoulder period, or switch back to the peak period, but others might switch to public transit or carpool. A city might reduce spatial substitution by expanding the geographic area subject to pricing, such that driving around the priced area would be impractical. Alternatively it might improve roads likely to see policy-driven traffic increases, as London did prior to introducing its cordon charge (Santos, 2004).

Our public transit results also have policy implications. Because responses to pricing vary with transit availability, welfare impacts from pricing will be spatially heterogeneous. Policymakers may wish to consider these distributional impacts when designing a road pricing policy. More generally, our results suggest that road pricing and public transit may be substitutes, at least within cities. In areas that already have high levels of public transit, there may be limited scope for reducing traffic via road pricing.

We find suspension of the charge increased weekday concentrations of CO by 6 percent and PM10 by 17 percent. This is a remarkable change in air quality, given: 1) the charge area represents only 5 percent of the city, and a smaller fraction of the broader metropolitan area; and 2) it is a lower bound on the potential long-run increase. Our estimate is still more surprising in light of Milan's relatively clean vehicle fleet. Previous welfare analyses of Milan's Ecopass policy have found net benefits of approximately d'7-12 million per year, even placing extremely low values on air pollution reductions (Danielis et al., 2011). We estimate that the Area C policy produces a \$3 billion welfare gain from air pollution reductions alone.

Pollution effects from similar policies in cities with dirtier fleets could well be larger. More congested cities would also tend to see larger welfare gains. Among the most congested large world cities are Istanbul, Mexico City, and Rio de Janeiro (TomTom, 2014). Among large US cities, New York, Los Angeles, and Chicago see the highest welfare losses from congestion (Lomax et al., 2012). Our results suggest there is scope for road pricing, even in second-best form, to produce very large welfare gains in such environments.

Acknowledgements: The authors thank Prashant Bharadwaj, Jennifer Burney, Richard Carson, Andrew Chamberlain, Julie Cullen, Gordon Dahl, Jamie Mullins, Kevin Roth, Lanfranco Senn, and two anonymous referees for valuable advice. We also thank AMAT and the Settore Pianificazione e Programmazione Mobilità e Trasporto Pubblico Comune di Milano for data and assistance. Chapter 1, in full, has been submitted for publication of the material as it may appear in the Journal of Urban Economics, 2015, Gibson, Matthew; Carnovale, Maria, Elsevier, 2015. The dissertation author was the primary investigator and author of this paper.

References

Anderson, Michael L (2014). "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion". In: *The American Economic Review* 104.9, pp. 2763–2796.

Arnott, Richard and Marvin Kraus (1998). "When are anonymous congestion charges consistent with marginal cost pricing?" In: *Journal of Public Economics* 67, pp. 45–64.

Arnott, Richard, André De Palma, and Robin Lindsey (1993). "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand". In: *The American Economic Review* 83.1, pp. 161–179.

Azienda Transporti Milanesi (2013). *Company Profile*. Tech. rep.

Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins (July 2009). "Migration and hedonic valuation: The case of air quality". In: *Journal of Environmental Economics and Management* 58.1, pp. 1–14.

Börjesson, Maria, Jonas Eliasson, Muriel B. Hugosson, and Karin Brundell-Freij (Mar. 2012). “The Stockholm congestion charges 5 years on. Effects, acceptability and lessons learnt”. In: *Transport Policy* 20, pp. 1–12.

Carra, Ilaria (2012). *Il sindaco: Salveremo Area C*. la Repubblica.

Carra, Ilaria and Alessia Gallione (2012). *Milano sconfitta dal parcheggiatore sospesa l' area anti-traffico*. la Repubblica.

City of Milan (2012). *Area C*. Accessed December 1, 2012. URL: http://www.comune.milano.it.

Corriere della Sera (Sept. 2012a). *Area C, torna il ticket anti traffico da 5 euro Al giovedì telecamere spente alle 18*.

— (July 2012b). *Il Consiglio di Stato sospende Area C Da questa mattina telecamere spente*.

Danielis, Romeo, Lucia Rotaris, Edoardo Marcucci, and Jérôme Massiani (2011). “An economic, environmental and transport evaluation of the Ecopass scheme in Milan: three years later”.

Davis, Lucas (Feb. 2008). “The Effect of Driving Restrictions on Air Quality in Mexico City”. In: *Journal of Political Economy* 116.1, pp. 38–81.

De Borger, Bruno and Stef Proost (2001). *Reforming transport pricing in the European Union: A modelling approach*. Edward Elgar Publishing.

DeCorla-Souza, Patrick (2004). “Recent U.S. Experience: Pilot Projects”. In: *Research in Transportation Economics* 9.04, pp. 283–308.

Eliasson, Jonas (Nov. 2008). “Lessons from the Stockholm congestion charging trial”. In: *Transport Policy* 15.6, pp. 395–404.

Eliasson, Jonas, Lars Hultkrantz, Lena Nerhagen, and Lena Smidfelt Rosqvist (Mar. 2009). “The Stockholm congestion-charging trial 2006: Overview of effects”. In: *Transportation Research Part A: Policy and Practice* 43.3, pp. 240–250.

Finkelstein, Amy (2009). “E-ZTax: Tax Salience and Tax Rates”. In: *The Quarterly Journal of Economics* 124.3, pp. 969–1010.

Foreman, Kate (2013). “Crossing the Bridge: The Effects of Time-Varying Tolls on Curbing Congestion”.

Friedman, Michael S, Kenneth E Powell, Lori Hutwagner, LeRoy M Graham, and W Gerald Teague (2001). “Impact of changes in transportation and commuting behaviors during the 1996 Summer Olympic Games in Atlanta on air quality and childhood asthma”. In: *Journal of the American Medical Association* 285.7, pp. 897–905.

Gallego, Francisco, Juan-Pablo Montero, and Christian Salas (Nov. 2013). "The effect of transport policies on car use: Evidence from Latin American cities". In: *Journal of Public Economics* 107, pp. 47–62.

Goh, Mark (2002). "Congestion management and electronic road pricing in Singapore". In: *Journal of Transport Geography* 10, pp. 29–38.

Hammitt, James K. and Lisa A. Robinson (2011). "The Income Elasticity of the Value per Statistical Life: Transferring Estimates between High and Low Income Populations". In: *Journal of Benefit-Cost Analysis* 2.1, pp. 1–29.

Harsman, Bjorn and John M. Quigley (2010). "Political and Public Acceptibility of Congestion Pricing: Ideology and Self-Interest". In: *Journal of Policy Analysis and Management* 29.4, pp. 854–874.

Invernizzi, Giovanni, Ario Ruprecht, Roberto Mazza, Cinzia De Marco, Griša Močnik, Costantinos Sioutas, and Dane Westerdahl (July 2011). "Measurement of black carbon concentration as an indicator of air quality benefits of traffic restriction policies within the ecopass zone in Milan, Italy". In: *Atmospheric Environment* 45.21, pp. 3522–3527.

Jones, Peter and Arild Hervik (1992). "Restraining car traffic in European cities: an emerging role for road pricing". In: *Transportation Research Part A: Policy and Practice* 26.2, pp. 133–145.

Kaul, Sapna, Kevin J. Boyle, Nicolai V. Kuminoff, Christopher F. Parmeter, and Jaren C. Pope (2013). "What can we learn from benefit transfer errors? Evidence from 20 years of research on convergent validity". In: *Journal of Environmental Economics and Management* 66.1, pp. 90–104.

la Repubblica (Jan. 2008). *Milano scatta l 'ora dell'Ecopass multe salate per chi sgarra*.

Lévy-Lambert, H (1968). "Tarification des Services à Qualité Variable–Application aux Péages de Circulation". In: *Econometrica: Journal of the Econometric Society* 36.3, pp. 564–574.

Lindsey, C. Robin and Erik T. Verhoef (2000). "Traffic Congestion and Congestion Pricing".

Lomax, Tim, David Schrank, and Bill Eisele (2012). *2012 Urban Mobility Report*. Tech. rep. Institute, Texas Transportation.

Marchand, Maurice (1968). "A Note on Optimal Tolls in an Imperfect Environment". In: *Econometrica* 36.3, pp. 575–581.

Martino, Angelo (2012). *Milano: From Pollution Charge to Congestion Charge*. Tech. rep. TRT Trasporti e Territorio.

Meland, Solveig (1995). “Generalised and advanced urban debiting innovations: the GAUDI Project. III: The Trondheim toll ring”. In: *Traffic Engineering & Control* 36.3, pp. 150–155.

Milan Tourism (2012). *Area C*. Accessed December 1, 2012. URL: http://www.turismo.milano.it.

Mun, Se-il, Ko-ji Konishi, and Kazuhiro Yoshikawa (2003). “Optimal cordon pricing”. In: *Journal of Urban Economics* 54, pp. 21–38.

Odeck, James and Svein Brathen (2008). “Travel demand elasticities and users attitudes: A case study of Norwegian toll projects”. In: *Transportation Research Part A: Policy and Practice* 42, pp. 77–94.

OECD (2011). *OECD Regions at a Glance*. Accessed April 9, 2015. OECD. URL: http://www.oecd-ilibrary.org/sites/reg_glance-2011-en/03/06/index.html.

Olszewski, Piotr and Litian Xie (2002). “Traffic Demand Elasticity with Respect to Road Pricing – Some Evidence from Singapore”. In: *Proceedings of the International Conference on Seamless and Sustainable Transport*. November, pp. 217–228.

— (Aug. 2005). “Modelling the effects of road pricing on traffic in Singapore”. In: *Transportation Research Part A: Policy and Practice* 39.7-9, pp. 755–772.

Pierce, Gregory and Donald Shoup (2013). “Getting The Prices Right”. In: *Journal of the American Planning Association* 79.1, pp. 67–81.

Polak, J and S Meland (1994). “An assessment of the effects of the Trondheim Toll Ring on travel behaviour and the environment”. In: *Towards an intelligent transport system. Proceedings of the first world congress on applications of transport telematics and intelligent vehicle-highway systems, November 30 - 3 December 1994, Paris, volume 2.*

Povoledo, Elisabetta (Aug. 2012). *Effort to Cut Milan Traffic Halted as Court Favors Garage*.

Ramjerdi, Farideh, Harald Minken, and Knut Ostmoe (2004). “Norwegian Urban Tolls”. In: *Research in Transportation Economics* 9.04, pp. 237–249.

Ready, Richard and Stå le Navrud (2006). “International benefit transfer: Methods and validity tests”. In: *Ecological Economics* 60.2, pp. 429–434.

Rotaris, Lucia, Romeo Danielis, Edoardo Marcucci, and Jerome Massiani (June 2010). “The urban road pricing scheme to curb pollution in Milan, Italy”. In: *Transportation Research Part A: Policy and Practice* 44.5, pp. 359–375.

Santos, Georgina (2004). “Urban Road Pricing in the U.K.” In: *Research in Transportation Economics* 9.04, pp. 251–282.

Santos, Georgina (2008). “London Congestion Charging”. In: *Brookings-Wharton Papers on Urban Affairs*, pp. 177–234.

Santos, Georgina and Gordon Fraser (2006). “Road pricing: lessons from London”. In: *Economic Policy* April, pp. 263–310.

Seaton, Anthony, D Godden, W MacNee, and K Donaldson (1995). “Particulate air pollution and acute health effects”. In: *The Lancet* 345.8943, pp. 176–178.

Seinfeld, John H and Spyros N Pandis (2012). *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons.

Small, Kenneth A. and Jose A. Gomez-Ibanez (1998). “Road Pricing for Congestion Management: The Transition from Theory to Policy”. In: *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*. Vol. 2, pp. 107–123.

Small, Kenneth A. and Erik T. Verhoef (2007). *The Economics of Urban Transportation*.

Small, Kenneth A, Clifford Winston, and Jia Yan (2005). “Uncovering the Distribution of Motorists’ Preferences for Travel Time and Reliability”. In: *Econometrica* 73.4, pp. 1367–1382.

Small, Kenneth A., Clifford Winston, and Jia Yan (2006). “Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous Preferences in Policy Design”. In: *Brookings-Wharton Papers on Urban Affairs*, pp. 53–96.

Smith, V Kerry and Ju-chin Huang (1995). “Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models”. In: *Journal of Political Economy* 103.1, pp. 209–227.

TomTom (2014). *TomTom Traffic Index*. Accessed April 11, 2015. URL: http://www.tomtom.com/en_gb/trafficindex/#/list.

Toronto Board of Trade (2011). *Toronto as a Global City: Scorecard on Prosperity*. Tech. rep.

Transport for London (2005). *Environment Report 2005*. Tech. rep.

— (2008). *Environment Report 2008*. Tech. rep.

Tretvik, T (2003). “Urban road pricing in Norway: Public acceptibility and travel behavior. In: Acceptability of transport pricing strategies”. In: *MC-ICAM Conference, Acceptability of Transport Pricing Strategies*.

Vaca, Erin and J. Richard Kuzmyak (2005). *TCRP Report 95*. Tech. rep.

Verhoef, Erik T. (2005). “Second-best congestion pricing schemes in the monocentric city”. In: *Journal of Urban Economics* 58, pp. 367–388.

Verhoef, Erik T., Peter Nijkamp, and Piet Rietveld (1996). “Second-best congestion pricing: the case of an untolled alternative”. In: *Journal of Urban Economics* 40, pp. 279–302.

Viard, V. Brian and Shihe Fu (2014). “The Effect of Beijing’s Driving Restrictions on Pollution and Economic Activity”.

Vickrey, William (1963). “Pricing in Urban and Suburban Transport”. In: *The American Economic Review* 53.2, pp. 452–465.

Wang, Lanlan, Jintao Xu, and Ping Qin (2014). “Will a Driving Restriction Policy Reduce Car Trips? The Case Study of Beijing, China”. In: *Transportation Research Part A: Policy and Practice* 67.September, pp. 279–290.

Wolff, H. and L. Perry (Aug. 2010). “Trends in Clean Air Legislation in Europe: Particulate Matter and Low Emission Zones”. In: *Review of Environmental Economics and Policy* 4.2, pp. 293–308.

Wolff, Hendrik (2014). “Keep Your Clunker in the Suburb: Low-emission Zones and Adoption of Green Vehicles”. In: *The Economic Journal* 124.578, F481–F512.

Xie, Chunying (2013). “Dynamic Decisions to Enter a Toll Lane on the Road”.

1.7 Figures & tables

1.7.1 Figures

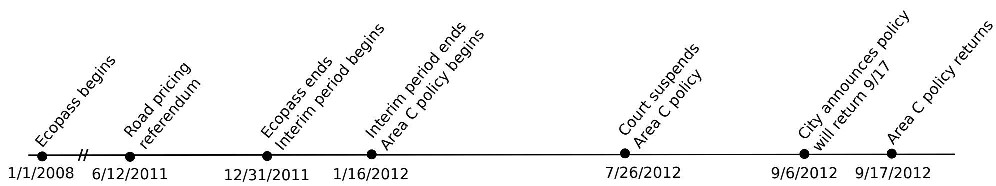


Figure 1.1: Timeline of road pricing in Milan

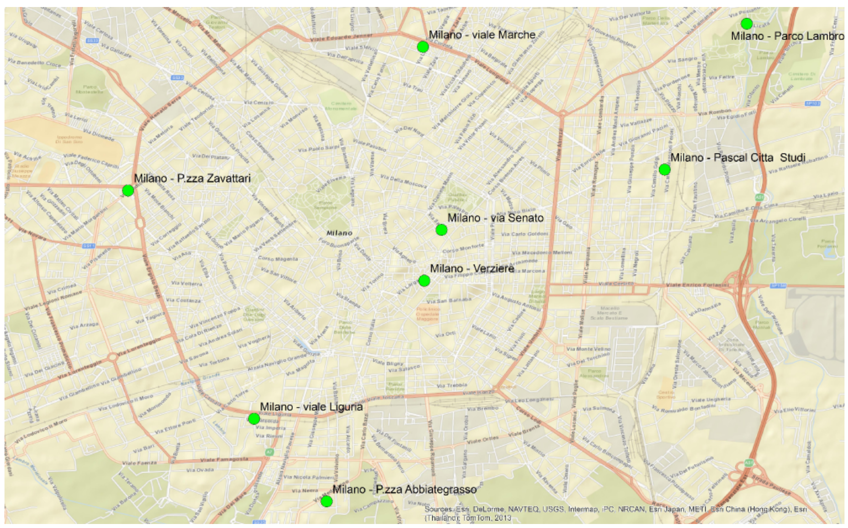


Figure 1.2: Air pollution monitoring stations in Milan

Green circles represent the locations of pollution monitoring stations operated by ARPA Lombardia. Not all stations monitor all pollutants. The via Senato and Verziere stations are inside Area C. The Piazza Zavattari, viale Marche, and viale Liguria stations are on the ring road. The Piazza Abbiategrasso, Pascal Citta Studi, and Parco Lambro stations are outside the ring road.

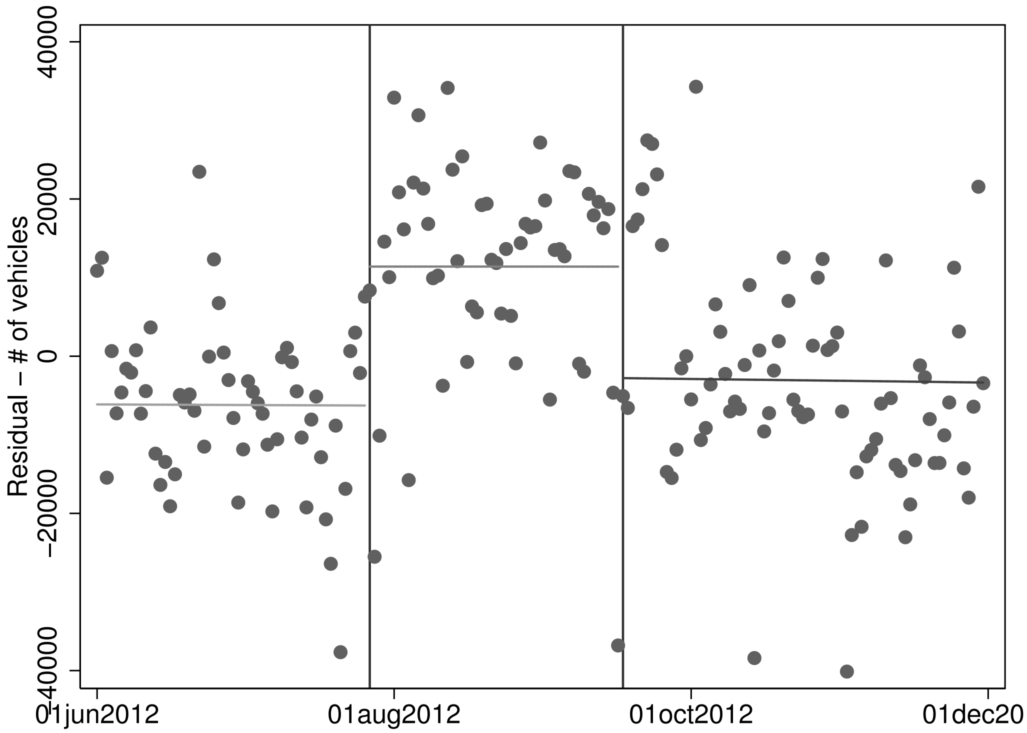


Figure 1.3: Effect of Area C charge suspension on vehicle entries

Residuals are from equation (1.1), with the charge suspension dummy variable excluded. The dependent variable is daily entries into Area C by vehicles subject to the charge. Each point represents one observation (one day). Fitted lines are based on separate degree-zero local polynomial regressions for pre-suspension, suspension, and post-suspension periods. Time controls include 4 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, a 7th-degree time trend in date, and a dummy for the unpriced January 2012 interim period. Weather controls comprise ten-piece linear splines in temperature and positive precipitation.

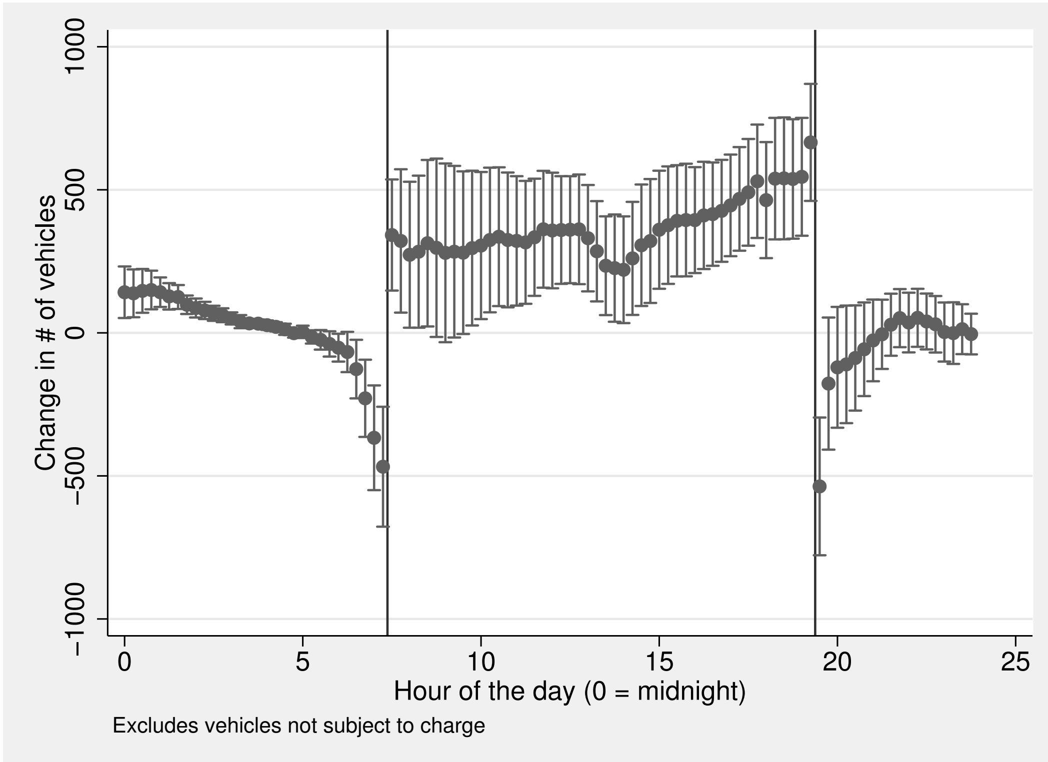


Figure 1.4: Effect of Area C charge suspension on vehicle entries, by 15-minute interval

Dependent variable is Area C entries by vehicles subject to charge. Estimates are from equation (1.1), estimated separately for each 15-minute interval. Whiskers represent Newey-West standard errors multiplied by 1.96. The lag length is 35 for hours 23.5-5.25, 7 otherwise.

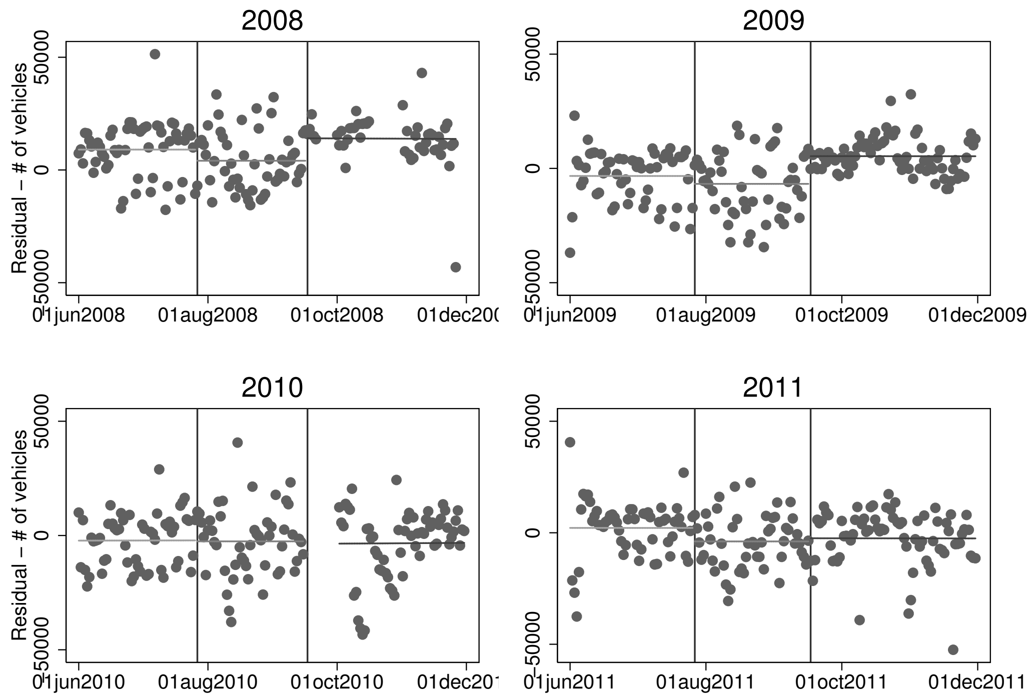


Figure 1.5: Effect of placebo suspensions on Area C vehicle entries

Residuals are from equation (1.1). Dependent variable is daily entries into Area C by vehicles subject to charge. Each point represents one observation (one day). Fitted lines based on separate degree-zero local polynomial regressions for pre-placebo, placebo, and post-placebo periods (which correspond to the dates of the 2012 natural experiment). Time controls include 0-4 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, a 7th-degree time trend in date, a dummy for the unpriced January 2012 interim period, and a dummy for the June-Sept. 2012 charge suspension. Weather controls comprise ten-piece linear splines in temperature and positive precipitation.

1.7.2 Tables

Table 1.1: Descriptive statistics, daily level

	Units	Mean	Std. dev.	Min	Max	N	EU standard
Area C entries	-	169,743.7	47,627.8	3,905	261,172	1,737	-
Passing vehicles	-	2,585,316	1,348,275	37,412	5,918,492	1,754	-
CO	mg/m^3	1.26	.67	0	7.6	17,625	10
PM10	$\mu g/m^3$	47.66	31.07	0	276	9,091	40
PM2.5	$\mu g/m^3$	33.74	26.56	0	177	2,550	25
Precipitation	mm	2.07	6.91	0	121.2	30,929	-

All statistics calculated over daily means. The EU standard for CO is based on a rolling 8-mean, while those for PM10 and PM2.5 are based on annual means.

Table 1.2: Weekday effect of Area C charge suspension on vehicle entries

	All vehicles	Charged vehicles	Motorcycles	Other vehicles
Charge suspension	26725.2***	29266.1***	-1920.9	-62.69
	(5059.5)	(3275.8)	(2447.3)	(54.50)
Charge suspension*weekend	-19590.1**	-23094.3***	3295.2	171.4*
	(9090.5)	(5865.1)	(3566.0)	(98.93)
Year, month, week, DoW FEs	Yes	Yes	Yes	Yes
7th-deg. trend in date	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes
Observations	1737	1737	1720	1737
R^2	0.808	0.805	0.785	0.901

Dependent variable is daily Area C entries. "Other" vehicles are primarily public vehicles like police cars and ambulances, which are exempt from the charge. Each column is a single model corresponding to equation (1.1). All specifications include 4 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, a 7th-degree time trend in date, and a dummy for the unpriced January 2012 interim period. Weather controls comprise ten-piece linear splines in temperature and positive precipitation. Newey-West standard errors with 7 lags in parentheses. Significance denoted by: $*$ $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The effect on entries by all vehicles is approximately 14.5 percent of the mean.

Table 1.3: Weekday effect of Area C charge suspension on sensor-level traffic volume, by distance outside Area C boundary

	Vehicle count	Vehicle count	Vehicle count
All roads	469.8***		
	(131.3)		
Area C		1063.2***	
		(337.1)	
0-1km outside boundary		-1061.1*	
		(587.0)	
1-2km outside boundary		-161.5	
		(361.0)	
2-4.2km outside boundary		606.8**	
		(258.3)	
>4.2km outside boundary		515.3	
		(391.2)	
Non-ring roads			469.2***
			(158.4)
Ring roads			-2433.7**
			(1020.0)
Year, month, week, DoW FEs	Yes	Yes	Yes
7th-deg. trend in date	Yes	Yes	Yes
Weather controls	Yes	Yes	No
Observations	803086	801442	801442
R^2	0.085	0.093	0.093

Dependent variable is daily count of vehicles passing over sensor. Each column is a single model corresponding to equation (1.2). All specifications include 4 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, a 7th-degree time trend in date, and a dummy for the unpriced January 2012 interim period. Weather controls comprise ten-piece linear splines in temperature and positive precipitation. Distance measured from the outside of the Area C boundary. Distance dummies set at the 25th, 50th, and 75th percentiles. Ring roads allow drivers to avoid charge area. Standard errors clustered at sensor level in parentheses. Significance denoted by: $*$ $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The overall effect in the leftmost column is approximately 8 percent of the mean.

Table 1.4: Weekday effect of Area C charge suspension on portal-level vehicle entries, by public transit availability

	Vehicle count	Vehicle count	Vehicle count	Vehicle count
No metro	883.9^{***}			
	(160.4)			
Metro	259.9			
	(266.5)			
No bus		905.6^{***}		
		(148.6)		
Bus		374.3^{*}		
		(211.7)		
No tram			788.6^{***}	
			(152.3)	
Tram			586.1^{**}	
			(247.1)	
No public trans.				1063.2^{***}
				(208.9)
Public trans.				518.3^{***}
				(133.4)
Yr, mo, wk, DoW FEs	Yes	Yes	Yes	Yes
7th-deg. trend in date	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes
Observations	71862	71862	71862	71862
R^2	0.407	0.407	0.406	0.407

Dependent variable is daily Area C entries through a given portal. Each column is a single model corresponding to equation (1.3). All specifications include 4 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, a 7th-degree time trend in date, and a dummy for the unpriced January 2012 interim period. Weather controls comprise ten-piece linear splines in temperature and positive precipitation. Standard errors clustered at portal level in parentheses. Significance denoted by: $*$ $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.5: Price elasticity of Area C vehicle entries

	All charged	Passenger	Commercial	Other
log(Price)	-0.304***	-0.171***	-0.467**	-0.0222
	(0.0944)	(0.0330)	(0.221)	(0.0341)
Month, week, DoW FEs	Yes	Yes	Yes	Yes
7th-deg. trend in date	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes
Observations	1147	1147	1147	1147
R^2	0.458	0.418	0.662	0.557

Dependent variable is log daily Area C entries. Each column is a single model corresponding to equation (1.1), but with the policy variables replaced by the log of weekday average price: d'0.72 for passenger vehicles and d'3.52 for commercial vehicles under Ecopass, d'5 under Area C. Estimated elasticities reflect traffic response to two changes: 1) a 1 percent increase in the Area C cordon charge; 2) the resulting cost decrease from reduced congestion and travel time. Weekends were unpriced under both policies and are excluded from the sample. The January 2012 interim period and the July-September 2012 natural experiment period were also unpriced. All specifications include 11 month, 51 week, and 4 day of week FEs, a holiday dummy, and a 7th-degree time trend in date. Co-linearity with price variation prevents the inclusion of year dummies. Weather controls comprise ten-piece linear splines in temperature and positive precipitation. "Other" vehicles are primarily public vehicles like police cars and ambulances, which are exempt from charge. Standard errors clustered at the year-week level. Significance denoted by: $*$ p < 0.10, ** p < 0.05, *** p < 0.01.

Table 1.6: Weekday pollution effect of Area C charge suspension, by location

	ln(CO)	ln(PM10)	ln(PM2.5)
Area C	0.0606**	0.0404	
	(0.0248)	(0.0407)	
Ring roads	0.0182		
	0.0205		
Outside		0.1696**	0.2139*
		(0.0676)	(0.1210)
Lagged pollution	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes
Year, month, week, DoW FEs	Yes	Yes	Yes
7th-deg. trend in date	Yes	Yes	Yes

Dependent variable is daily log average pollution in a given area of Milan. Pollution normalized for temperature and pressure by ARPA. Each estimate comes from a different regression corresponding to equation (1.4). Specifications include 10 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, a 7th-degree trend, a dummy for the unpriced January 2012 interim period, and 1 lag of log average pollution. Weather controls include 4-knot cubic splines in humidity, wind speed, solar radiation, and precipitation, plus a dummy for positive precipitation. Newey-West standard errors with 1 lag in in parentheses. Significance denoted by: $*$ $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.7: Weekday pollution effects of placebo suspensions

	CO	PM10	PM25
2008			
Area C	0.0463	-0.0642	
	(0.0247)	(0.0472)	
Ring roads	-0.0358*		
	(0.0206)		
Outside		-0.0495	0.0474
		(0.0681)	(0.1399)
2009			
Area C	-0.0095	0.0464	
	(0.0198)	(0.0428)	
Ring roads	0.0433**		
	(0.0182)		
Outside		0.1231**	0.1492
		(0.0591)	(0.0917)
2010			
Area C	-0.0182	-0.0718	
	(0.0254)	(0.0493)	
Ring roads	0.0063		
	(0.0184)		
Outside		-0.0258	-0.0498
		(0.0736)	(0.1054)
2011			
Area C	0.1007***	-0.0516	
	(0.0177)	(0.0492)	
Ring roads	-0.0506***		
	(0.0190)		
Outside		0.0142	0.0697
		(0.0849)	(0.1138)

Dependent variable is daily log average pollution pollution in a given area of Milan. Pollution normalized for temperature and pressure by ARPA. Each estimate comes from a different regression corresponding to equation (1.4). Placebo suspension runs July 27-September 16. Specifications include 10 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, 7th-degree trend, a holiday dummy, a dummy for the unpriced January 2012 interim period, a dummy for the June-Sept. 2012 charge suspension, and 1 lag of log average pollution. Weather controls include 4-knot cubic splines in humidity, wind speed, solar radiation, and precipitation, plus a dummy for positive precipitation. Newey-West standard errors with 1 lag in parentheses. Significance denoted by: $*$ $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.8 Supplementary material

1.8.1 Public transit results

The results in Table 1.8 provide evidence of some preference heterogeneity (sorting). They are inconsistent with an explanation solely based on cost differences, which would predict smaller responses on more congested routes, but do not exclude the possibility that cost differences drive some of the responses to charge suspension.

To see this, consider the framework of Anderson (2014), who derives a condition for choosing rail over driving (simplified here): $c_{rail} - c_{drive} \leq P_0$, where c_{rail} and c_{drive} denote the time costs of driving and the subway. P_0 is the fiscal cost difference between modes converted to units of time. Suppose three driver types, A, B, and C as illustrated in Figure 1.6, all of whom initially take the subway.[16] Drivers B and C face higher subway time cost because they must commute circumferentially to the subway line. Assume B and C are close enough that this cost is the same ($c_{rail,C} = c_{rail,B}$). Then we have $c_{rail,C} - c_{drive,C} < c_{rail,B} - c_{drive,B} \leq P_0$. Provided $c_{drive,A}$ is not too small, $c_{rail,A} - c_{drive,A} < c_{rail,C} - c_{drive,C} < c_{rail,B} - c_{drive,B} \leq P_0$. This framework partially reproduces the public transit results of Table 1.4; it predicts the marginal drivers will be of types B and C, not A. But it implies more infra-marginal drivers on less congested roads (more type B than C), which does not match the pattern of Table 1.8.

[16] While this example is obviously stylized, the radial layout mimics the actual pattern of roads in Milan.

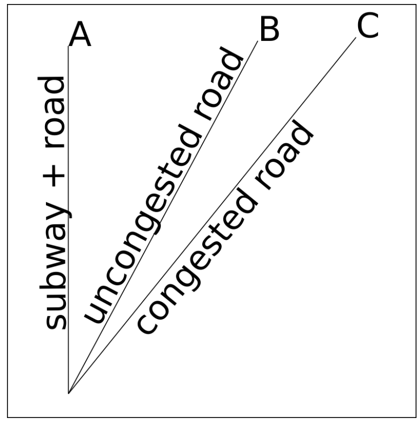


Figure 1.6: Marginal drivers

Table 1.8: Weekday effect of Area C charge suspension on portal-level vehicle entries, by public transit availability and congestion

	Vehicle count
No public trans.	1012.6***
	(211.4)
Public trans.	541.5***
	(149.3)
No public trans. * congested	843.9***
	(159.7)
Public trans. * congested	-189.1
	(534.9)
Year, month, week, DoW FEs	Yes
7th-deg. trend in date	Yes
Weather controls	Yes
Observations	71862
R^2	0.407

Dependent variable is daily Area C entries through a given portal. Each observation is a portal-day. Each column is a single model corresponding to equation (1.3). Congested dummy equals 1 for portals where avg standardized peak (8-9:30AM, 5:45-8PM) volume is below 10th percentile. All specifications include 4 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, a 7th-degree time trend in date, and a dummy for the unpriced January 2012 interim period. Weather controls comprise ten-piece linear splines in temperature and positive precipitation. Standard errors clustered at portal level in parentheses. Significance denoted by: $*$ $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.8.2 Supplemental figures

Figure 1.7: Milan's Area C

Area C is the area of the Milan city center that has been priced under the Ecopass and Area C policies. Numbered circles represent entry portals where the charge is applied using license plate-reading cameras.

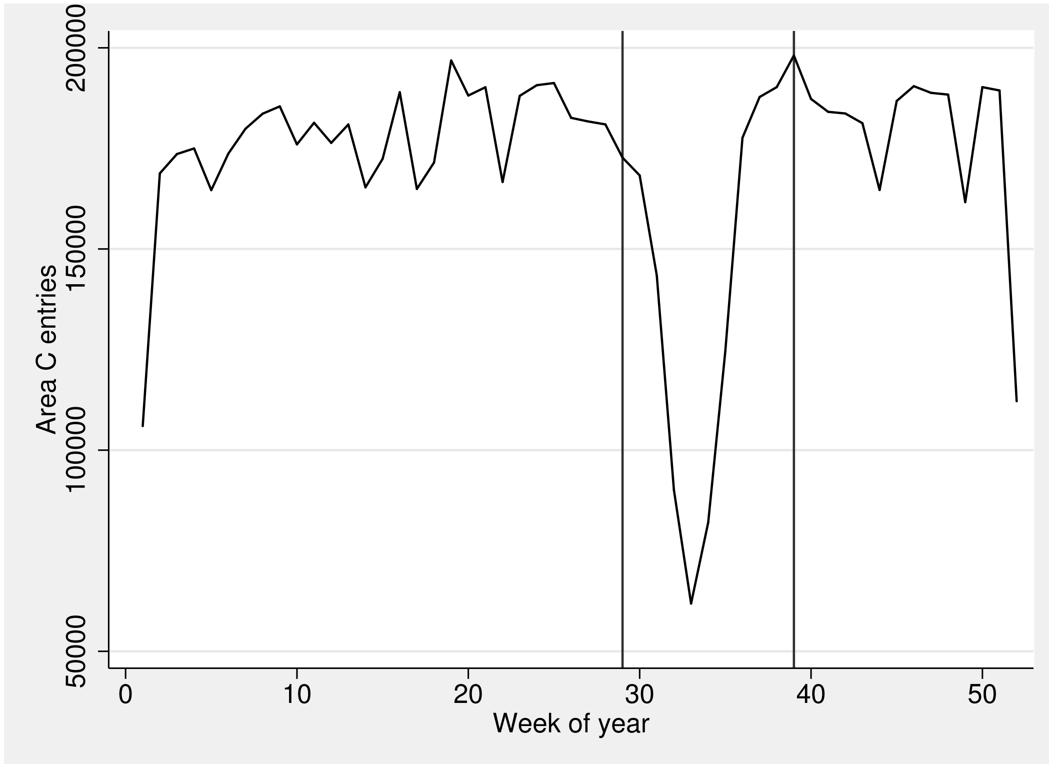


Figure 1.8: Seasonal pattern of Area C vehicle entries

Figure shows average number of Area C entries by week of year, as measured by license plate readers. Vertical lines delimit the period of the 2012 charge suspension.

1.8.3 Supplemental tables

Table 1.9: Comparison of traffic volume effects and elasticities to other empirical road pricing studies

Paper	Location	Policy	Volume change	Price elasticity
Gibson & Carnovale 2015	Milan	Cordon charge	+14.5%	-.30
Jones and Hervik (1992)	Alesund	Toll	-	-.45
Jones and Hervik (1992); Ramjerdi et al. (2004)	Oslo	Cordon charge	-10% to 0%	-.22, -.03
Polak and Meland (1994); Meland (1995)	Trondheim	Cordon charge	-10%	-.10
Small and Gomez-Ibanez (1998)	Autoroute A1 Paris-Lille	Variable toll	-4%	-.16
Small and Gomez-Ibanez (1998)	Singapore Restricted Zone	Cordon charge (1975)	-44%	-
Goh (2002); Olszewski and Xie (2002)	Singapore Expressways	Toll (1995)	-16%	-.22 to -.15
Olszewski and Xie (2002)	Singapore Restricted Zone	Cordon charge (1976, 1989)	-52% to -10%	-.22 to -.32
Tretvik (2003); Ramjerdi et al. (2004)	Bergen	Cordon charge	>-3%	-
Small et al. (2006)	Orange County SR91	Variable toll	-	-.36
Odeck and Brathen (2008)	Norway, various	Toll	-	-.56 to -.82
Santos (2008)	London	Cordon charge	-18%	-
Eliasson et al. (2009)	Stockholm	Cordon charge	-22%	-
Finkelstein (2009)	Various US	Toll	-	-.06
Foreman (2013)	SF Bay Bridge	Variable toll	-9%, -4%	-.08
Xie (2013)	Minneapolis I-394	Variable toll	-	-.14

With the exception of our result, papers are listed in order of publication. Volume changes are from the introduction of pricing or a price change. The volume estimate from our study is positive because we examine the removal of a pricing policy, rather than the imposition of one. Note that our volume change estimate is from a natural experiment, while our price elasticity relies on long-run policy variation. Small and Verhoef (2007) suggest the very large effect of Singapore's initial pricing policy was due to the extremely high charge, which they characterize as far above the second-best optimum level. We separate estimates based on the initial 1977 introduction of pricing in Singapore (Small and Gomez-Ibanez, 1998) from those based on later price changes (Goh, 2002; Olszewski and Xie, 2002). Foreman (2013) provides two estimates: -9% from a regression discontinuity identification, and -4% from a difference-in-differences identification. The rightmost column contains elasticities of traffic volume (cordon crossings or volume on a segment) with respect to price. Unless otherwise noted, they reflect both the direct effect of a toll increase (negative) and the rebound effect from reduced congestion (positive). They generally do not correspond to structural parameters. Elasticities are not available for all of the policies. In some cases authors do not provide sufficient information to calculate them. In others the relevant toll rises from zero to a positive number, leaving the percentage change undefined. In particular, the elasticity estimate of Olszewski and Xie (2002) comes from later price variation, not the initial introduction of the Singapore cordon charge. The Finkelstein (2009) estimate is an average over 33 US facilities.

Table 1.10: Comparison of Area C charge suspension to unpriced January 2012 interim period

	All vehicles	Charged vehicles	Motorcycles	Other vehicles
Charge suspension	26725.2***	29266.1***	-1920.9	-62.69
	(5059.5)	(3275.8)	(2447.3)	(54.50)
Interim period	38692.4***	31237.0***	7924.1**	13.11
	(8308.6)	(6166.6)	(3978.2)	(68.76)
Year, month, week, DoW FEs	Yes	Yes	Yes	Yes
7th-deg. trend in date	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes
Observations	1737	1737	1720	1737
R^2	0.808	0.805	0.785	0.901

Charge suspension estimates are identical to those in main results (Table 1.2). Dependent variable is daily Area C entries. Each column is a single model corresponding to equation (1.1). All specifications include 4 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, and a 7th-degree time trend in date. Weather controls comprise ten-piece linear splines in temperature and positive precipitation. Other vehicles are primarily public vehicles like police and ambulances, which are exempt from charge. The interim period January 1-15, 2012, between the Ecopass and Area C policies, was unpriced. Newey-West standard errors with 7 lags in parentheses. Significance denoted by: ∗ p < 0.10, ** p < 0.05, *** p < 0.01.

Table 1.11: Weekday effect of unpriced January 2012 interim period on sensor-level traffic volume, by distance outside Area C boundary

	Vehicle count	Vehicle count	Vehicle count
All roads	1067.6***		
	(127.4)		
Area C		1597.4***	
		(437.7)	
0-1km outside boundary		583.8	
		(494.3)	
1-2km outside boundary		892.9***	
		(344.6)	
2-4.2km outside boundary		978.0***	
		(308.1)	
>4.2km outside boundary		1402.0***	
		(310.5)	
Non-ring roads			1330.2***
			(150.8)
Ring roads			-231.0
			(697.1)
Year, month, week, DoW FEs	Yes	Yes	Yes
7th-deg. trend in date	Yes	Yes	Yes
Weather controls	Yes	Yes	No
Observations	803086	801442	801442
R^2	0.085	0.093	0.093

Dependent variable is daily count of vehicles passing over sensor. Each column is a single model corresponding to equation (1.2). All specifications include 4 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, a 7th-degree time trend in date, and a dummy for the July-September 2012 charge suspension. Weather controls comprise ten-piece linear splines in temperature and positive precipitation. Distance measured from the outside of the Area C boundary. Distance dummies set at the 25th, 50th, and 75th percentiles. Ring roads allow drivers to avoid charge area. The interim period January 1-15, 2012, between the Ecopass and Area C policies, was unpriced. Standard errors clustered at sensor level in parentheses. Significance denoted by: $*$ p < 0.10, ** p < 0.05, *** p < 0.01.

Table 1.12: Weekday effect of unpriced January 2012 interim period on portal-level vehicle entries, by public transit availability

	Vehicle count	Vehicle count	Vehicle count	Vehicle count
No metro	800.9***			
	(118.9)			
Metro	577.1***			
	(153.9)			
No bus		822.4***		
		(131.2)		
Bus		571.6***		
		(130.3)		
No tram			791.5***	
			(128.1)	
Tram			598.6***	
			(133.8)	
No public trans.				913.3***
				(148.2)
Public trans.				633.5***
				(109.2)
Year, month, week, DoW FEs	Yes	Yes	Yes	Yes
7th-deg. trend in date	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes
Observations	71862	71862	71862	71862
R^2	0.407	0.407	0.406	0.407

Dependent variable is daily Area C entries through a given portal. Each column is a single model corresponding to equation (1.3). All specifications include 4 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, a 7th-degree time trend in date, and a dummy for the July-September 2012 charge suspension. Weather controls comprise ten-piece linear splines in temperature and positive precipitation. The interim period January 1-15, 2012, between the Ecopass and Area C policies, was unpriced. Standard errors clustered at portal level in parentheses. Significance denoted by: $*$ p < 0.10, ** p < 0.05, *** p < 0.01.

Table 1.13: Effect of 2011 placebo suspension on Area C vehicle entries

	All vehicles	Charged vehicles	Motorcycles	Other vehicles
Placebo suspension (2011)	2296.6	419.8	1815.7	-77.80
	(5826.7)	(3992.3)	(2297.9)	(48.72)
Year, month, week, DoW FEs	Yes	Yes	Yes	Yes
7th-deg. trend in date	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes
Observations	1737	1737	1720	1737
R^2	0.808	0.805	0.785	0.901

Dependent variable is daily Area C entries. "Other" vehicles primarily public vehicles like police cars and ambulances, which are exempt from charge. Each column is a single model corresponding to equation (1.1). Placebo suspension runs July 27-September 16, 2011. All specifications include 4 year, 11 month, 51 week, and 5 day of week FEs, a weekend dummy interacted with year, a holiday dummy, a 7th-degree time trend in date, a dummy for the unpriced January 2012 interim period, and a dummy for the June-Sept. 2012 charge suspension. Weather controls comprise ten-piece linear splines in temperature and positive precipitation. Newey-West standard errors with 7 lags in parentheses. Significance denoted by: $*$ p < 0.10, ** p < 0.05, *** p < 0.01.

Chapter 2

Regulation-induced pollution substitution

Abstract

Regulations may cause firms to re-optimize over pollution inputs, leading to unintended consequences. By regulating air emissions in particular counties, the Clean Air Act (CAA) gives firms incentives to substitute: 1) toward polluting other media, like landfills and waterways; and 2) toward pollution from plants in other counties. Using EPA Toxic Release Inventory data, I examine the effect of CAA regulation on these types of substitution. Regulated plants increase their ratio of water to air emissions by 44 percent. Regulation of an average plant increases air emissions at unregulated plants owned by the same firm by 17 percent, resulting in a net emissions increase.

2.1 Introduction

Pollution regulations can have large benefits (Chay and Greenstone, 2003b; Currie and Neidell, 2005). For example, EPA estimated the 1970-1990 benefits of the Clean Air

Act (CAA) at $22 trillion (Environmental Protection Agency, 2011). Previous research has examined some of the costs of the CAA, including labor transition costs (Walker, 2011; Walker, 2013) and firm entry and exit decisions (Henderson, 1996; Becker and Henderson, 2000), but little is known about the costs of firms re-optimizing over pollution inputs. The CAA regulates particular pollutants in particular counties, which creates incentives for firms to substitute among different forms of pollution. This paper tests two variants of this hypothesis: 1) Do firms respond to regulation by polluting other channels, like landfills and waterways? (cross-media substitution); and 2) Do multi-plant firms substitute toward pollution from other plants? (spatial leakage). The existence and magnitude of such responses is important both for welfare analysis of existing policies and for design of future policies.

If one of a county's air pollution monitors exceeds the CAA standard, the EPA designates the county as "non-attainment." The state then issues regulations to reduce that county's air pollution, including emissions requirements for industrial plants. Simple economic theory suggests that firms will respond to such regulations by substituting toward unregulated or less-regulated forms of pollution. In practice, many air emissions abatement strategies collect rather than destroy harmful material, generating secondary waste streams that must be discharged into other media (Farnsworth, 2011).

While there is anecdotal evidence of such behavior (Duhigg, 2009), previous empirical studies have not found much evidence of cross-media substitution. Sigman (1996) tests for substitution in chlorinated solvent releases by metals and manufacturing plants. The author finds no substitution driven by the CAA, but does find substitution driven by hazardous disposal prices. Greenstone (2003) tests for CAA-induced substitution in releases from the iron and steel industry and finds no evidence for it. Gamper-Rabindran (2009) models emissions of volatile organic compounds (VOC) by chemical manufacturers as a function of CAA non-attainment, proxying for output changes with employment changes. She finds no increased emissions into other media. Both Greenstone and Gamper-Rabindran model emissions differences as a function of county non-attainment.

My approach builds on this work along several dimensions. First, I outline a simple theoretical model that assumes constant returns to scale (CRS). Motivated by this model, I show that one can use emissions ratios to recover consistent estimates of substitution elasticities among pollution inputs. Ratio estimation avoids confounding substitution and output effects.

Second, by estimating in ratio levels rather than differences, I account for two important facts: 1) the CAA allows states and firms to respond slowly (over three years) to a non-attainment designation; and 2) many abatement decisions are discrete, producing a one-time change in the level of emissions, not a change in the growth rate. These facts will bias difference-based estimates of substitution responses toward zero.

Third, I control for spatial heterogeneity in regulation. My study follows the implications of Auffhammer et al. (2009), which finds that the effect of CAA non-attainment on the average monitor in a non-attainment county is zero, but that the effect on monitors above the CAA standard is -11 to -14 percent. Similarly, Bento et al. (2014) find that non-attainment affects home prices near non-attainment monitors, but not farther away. These findings suggest that regulators respond to non-attainment by focusing on problematic areas, rather than requiring uniform changes across a county. I demonstrate that only plants near non-attainment monitors are treated under the CAA. This pattern is consistent with a regulator whose objective function involves minimization of enforcement costs, either pecuniary or political (Amacher and Malik, 1996), rather than socially efficient abatement. My analysis of substitution accounts for this and so avoids averaging changes at treated plants with null responses from untreated plants in non-attainment counties. Finally, by pooling across industries I improve statistical power.

Using EPA Toxic Release Inventory (TRI) data, this study tests the substitution hypotheses outlined above by comparing regulated ("treated") plants in particulate non-attainment counties to unregulated plants. My identification relies on the exogeneity of non-attainment status and monitor locations with respect to time-varying plant characteristics. The exogeneity of non-attainment status derives largely from the small share of point sources in particulate emissions (25 percent; Auffhammer et al., 2011). The exogeneity of monitor placement derives from EPA placement rules, which are based on population characteristics (e.g. average age) rather than industrial characteristics, and the prohibitively high cost of relocating a plant in response to monitor placement (Raffuse et al., 2007).

Both cross-media substitution and spatial leakage occur and responses can be large. Regulated plants increase their ratio of water to air emissions by 44 percent. Regulation of an average plant increases air emissions at unregulated plants owned by the same firm by 17 percent, resulting in a net emissions increase. Not all substitution responses constitute unintended consequences; regulated plants increase their ratio of recycling to air emissions by 41 percent.[1]

One might worry that estimation in ratios assumes the hypothesis of cross-media substitution: if a firm fixes water emissions and reduces air emissions, for example, the ratio will increase. This is not the pattern observed in the data. Regulated plants increase their *level* of water emissions by 25 percent. While this estimate is not statistically significant at any conventional level ($p = .25$), it is large and practically important. My analysis focuses on emissions *ratios* rather than levels because this approach allows me to infer substitution elasticities, which facilitate both evaluation of existing policies and simulation of proposed policies.

These findings are important not only for air pollution regulation, but for pollution control policy generally. If firms substitute among various forms of pollution, an optimal

[1] Here and throughout the paper, I refer to changes in log values as percentage changes. While this is not strictly correct for the large changes I estimate, I follow this practice for simplicity of exposition.

policy must consider not just a *plant's* emissions into a particular medium, but rather a *firm's* emissions across all media, in all locations. Optimal policy would set a firm's emissions price for each medium and location equal to the marginal damage from emissions (leaving no medium or location unpriced). While such an optimal policy might not be feasible or consonant with policymaker goals, elasticities of substitution among pollutants are nonetheless a vital input into policy design.

This analysis contributes to the literature on regulation in the presence of mispriced substitutes (e.g. Campbell, 1991). It is the first work to document regulation-induced cross-media pollution substitution. My findings are consistent with the theoretical work of Fullerton and Karney (2014) on pollution substitution. This study also contributes to the literature on pollution leakage. To date this literature has focused on international leakage (Levinson and Taylor, 2008; Davis and Kahn, 2010; Hanna, 2010) and simulated carbon leakage (Fowlie, 2009; Bushnell and Mansur, 2011). Both Henderson (1996) and Becker and Henderson (2000) find the CAA makes firms more likely to enter attainment counties, which might be considered a form of leakage. Fowlie (2010) demonstrates reallocation of NO_x emissions across plants in response to the NO_x Budget Program, but in that case such reallocation was among the aims of the policy. To the best of my knowledge, mine is the first study to find evidence of unintended emissions leakage across existing domestic plants, and the first to show spatial heterogeneity in CAA-driven air pollution reductions at the plant level.

The rest of the paper is organized as follows. Section 2.2 provides background on regulations and abatement technology important to cross-media substitution. Section 2.3 discusses a simple theoretical model that informs my estimation. Section 2.4 describes the data, Section 2.5 presents estimating equations, Section 2.6 presents results, and Section 2.7 explores their robustness. Section 2.8 concludes.

2.2 Background

2.2.1 The Clean Air Act

Under the Clean Air Act, the EPA sets air quality standards for six criteria pollutants: carbon monoxide (CO), nitrogen dioxide (NO_2), particulate matter (PM), lead (Pb), sulfur dioxide (SO_2), and volatile organic compounds (VOC). For detailed information on particulate standards, which are the focus of this paper, see Appendix Table 2.17. A county violates the standard for a particular pollutant if at least one monitor exceeds the CAA standard in a given year.[2] In what follows, I refer to a monitor that exceeds the annual standard as a non-attainment monitor. A monitor violation triggers the following sequence of events (author's interview notes; Environmental Protection Agency, Undated):

1. Together EPA and the state go through a process to designate a county as non-attainment. This may take up to two years.

2. Non-attainment designation begins a process by which states submit a State Implementation Plan (SIP) to EPA. This may take 18 to 36 months.

3. SIPs are not federally enforceable until EPA approves them, but state authorities may enforce them prior to such approval. As a result actual regulation sometimes begins concurrent with a non-attainment designation, but often begins after a delay of a year or more.

As a result of such lags, states have often drafted or even submitted State Implementation Plans before one of their constituent counties officially receives a non-attainment designation (see for example Missoula County Environmental Health Division, 1999).

[2]While EPA sometimes regulates smaller areas within counties, this far less common than county-level regulation (author's interview notes).

SIPs detail steps that will bring the county into attainment. These typically include "lowest achievable emissions rates" (LAER) equipment requirements and plant-specific emissions limits (Becker and Henderson, 2000; Becker and Henderson, 2001; Walker, 2013). SIPs may prescribe a specific control technology for a plant, but they often allow a plant to choose an abatement strategy (discussed in Section 2.2.3). State and EPA enforcement mechanisms include fines, inspections, and withholding of federal highway funds (Becker and Henderson, 2000; Chay and Greenstone, 2005).

Under the CAA, EPA periodically revises standards to reflect new research on the health effects of air pollution. For example, the agency finalized new PM_{10} standards in 1987, but did not designate a county in non-attainment of the new standard until 1990.[3] Revisions of CAA standards typically cause large numbers of counties to fall into non-attainment simultaneously. In my data (described in Section 2.4), *PM* non-attainment lasts for an average of approximately 7 years. Conditional on *PM* non-attainment in at least one year, I observe on average 1 entry into non-attainment and .16 exits; most of these counties remain in non-attainment through 2010, the last year of my data. For 9 counties I observe two entries into non-attainment.

2.2.2 Regulation of water and land emissions

CAA-induced substitution will reduce welfare, relative to an optimal policy, only if substitute emissions are unpriced or underpriced. Such is the case for many TRI pollutants and many emissions channels. The Safe Drinking Water Act (SDWA) and the Pollutant Priority List (PPL) for the Clean Water Act do not cover many TRI chemicals (Gamper-Rabindran, 2009). For example, my TRI data contain 580 chemicals. The PPL lists 126 chemicals (Environmental Protection Agency, 2013). In addition, two recent Supreme Court decisions have limited the scope of the CWA. *Solid Waste Authority of*

[3] Particulate matter 10 microns or less in diameter.

Northern Cook County v. U.S. Army Corps of Engineers removed CWA protection from "isolated" water bodies, including many wetland areas. *Rapanos v. United States* removed CWA protection from waterways that are not navigable year-round (Environmental Protection Agency, 2008). (Note that the Mississippi River would arguably have met such a definition in 2012.)

The Resource Conservation and Recovery Act (RCRA) governs many forms of toxic disposal on land. Coal combustion residuals are currently exempt from many provisions of the RCRA, though the EPA is attempting to regulate them (Environmental Protection Agency, 2010). Some mining and petrochemical wastes are also exempt (Environmental Protection Agency, 1999). Regulation of TRI-listed air pollutants that do not fall into one of the six CAA criteria categories varies by industry. Under the 1990 CAA Amendments, EPA develops industry-specific regulations governing the air release of 187 toxic chemicals ("air toxics"). EPA "...does not prescribe a specific control technology, but sets a performance level based on a technology or other practices already used by the better-controlled and lower emitting sources in an industry" (Environmental Protection Agency, Undated). While the incomplete regulations governing water and land emissions suggest cross-media substitution may reduce welfare relative to the first-best case, a full welfare analysis is beyond the scope of this paper.

2.2.3 Abatement strategies and variable costs

If abatement costs were entirely fixed or if abatement were costless, plants would have no incentive to substitute in response. While abatement technologies usually have fixed costs, they also have large operating costs. Pollution control devices typically require substantial energy and may yield secondary wastes that require costly disposal. Processes that employ catalysts require periodic replacement of the catalyst. These variable costs range from 33 to 100 percent of capital cost for most abatement technologies

(Environmental Protection Agency, Undated; Vatavuk et al., 2000; Farnsworth, 2011). For other abatement options like fuel switching and coal washing, the new fuel must be weakly more expensive than the old, or the plant would have been using it before. Such costs mean that CAA non-attainment changes the relative price of air emissions for regulated plants. I catalog the most common particulate air emissions abatement strategies in Table 2.1.

Many abatement strategies produce secondary waste streams. For example, incineration decreases toxic air emissions but increases carbon emissions. Wet scrubbers "...can lead to water and solid waste pollution problems" (EC/R Incorporated, 1998). Theory predicts that firms will not consider the external costs of such secondary waste and so their abatement strategies may not be socially efficient. Even when a SIP prescribes a particular strategy, this may not correspond to the social optimum. In discussing potential environmental harm from air emissions abatement, EPA argues, "Such well-established adverse effects and their costs are normal and assumed to be reasonable and should not, in most cases, justify nonuse of the control technology" (Domike and Zacaroli, 2011). If states do not consider such adverse effects in formulating SIPs, then they may generate inefficiently high levels of pollution into landfills and waterways.

Abatement strategies often produce abrupt reductions in emissions once implemented. This is one of the reasons I estimate in levels rather than differences, as described in Section 2.5. Note that the set of available technologies may partially reflect the US policy environment, as the Clean Air and Clean Water acts have been in place, in approximately their current form, since the early 1970s.

2.3 Theory

The following simple model informs my estimating equations for cross-media substitution. Suppose a firm produces a single good, which trades at exogenous price p_O, using two pollution inputs A and W. (The following argument also holds in a three-input case; see Appendix Section 2.10.1.) For discussion, let A be air emissions and W be water emissions. The CAA may be viewed as shift in relative input prices $\frac{p_A}{p_W}$, with the increased price of air emissions having two components: 1) pecuniary cost, like the variable abatement cost described in Section 2.2.3; and 2) non-pecuniary cost, for example the cost of incurring the displeasure of a regulator. The object of policy interest is unconditional factor demand W^*, incorporating firms' possible output response to regulation. Suppose a CES production function, so the firm problem becomes:

$$\max_{A,W} p_O \left(c_A A^\rho + c_W W^\rho\right)^{1/\rho} - p_A A - p_W W$$

The constants c_A and c_W reflect a firm's technology. I adopt a CES functional form for expositional convenience, but the key assumption of my model is constant returns to scale (CRS). In their important theoretical work on pollution substitution, Fullerton and Karney (2014) likewise employ a CRS assumption. One might worry that the fixed costs associated with some abatement choices (e.g. scrubbers) violate a CRS assumption. If such costs are financed over relatively long time horizons, however, the non-convexity in the production set is minimal and a CRS assumption is reasonable. This is roughly analogous to modeling a scrubber as an increase in a third input, capital (see Appendix Section 2.10.1).

The choice of CES does not impose any strong assumptions on the nature of substitution.[4] Taking first order conditions, one obtains an optimality condition:

[4] If $\rho = 1$ the production function is linear (perfect substitution); if $\rho \to 0$, it is Cobb-Douglas; if $\rho \to -\infty$, it is Leontief (perfect complements).

$$\left(\frac{c_W}{c_A}\right)\left(\frac{W^{*\rho-1}}{A^{*\rho-1}}\right) = \frac{p_W}{p_A}$$

Taking logs gives a ratio of unconditional factor demands:

$$\ln\left(\frac{W^*}{A^*}\right) = \frac{1}{1-\rho}\ln\left(\frac{c_W}{c_A}\right) + \frac{1}{1-\rho}\ln\left(\frac{p_A}{p_W}\right) \tag{2.1}$$

If ρ is finite and $\rho \leq 1$, the inputs are substitutes and the coefficient $\sigma = \frac{1}{1-\rho}$ on the price ratio is positive. In a two-input model σ is the Hicks elasticity of substitution, which measures the curvature of an isoquant. In a model with three or more inputs, the derivative of the log optimal input ratio with respect to the log price ratio is instead the Morishima elasticity of substitution, which can be negative. If A and W are substitutes, σ is positive and theory predicts CAA regulation will produce an increase in the ratio of water to air pollution $\frac{W^*}{A^*}$.[5] Modeling W^* as a function of prices alone will not recover the elasticity of substitution because of the omitted variable A^*. Rearranging equation 2.1 to put A^* on the right hand side makes this apparent.

$$\ln(W^*) = \frac{1}{1-\rho}\ln\left(\frac{c_W}{c_A}\right) + \frac{1}{1-\rho}\ln\left(\frac{p_A}{p_W}\right) + \ln(A^*) \tag{2.2}$$

In the context of the CAA, suppose a plant is located in a county that falls into non-attainment. The plant has two emissions reduction options: 1) substitute toward another form of pollution W^* (e.g. by switching fuels or using existing pollution-control capital more intensively); 2) produce less output. If the plant does both, the level of W^* may fall even though the ratio $\frac{W^*}{A^*}$ has increased. The output reduction disguises the regulation-induced substitution. Avoiding this confusion requires controlling for A^*, Y^* (output), or both. I model the input ratio, thereby controlling for changes in A^*. Under a CRS

[5]For a model that treats CAA non-attainment as a limit on the quantity of air emissions, please see Appendix Section 2.10.1. The qualitative predications from that model are the same as those presented here.

assumption this is sufficient to recover a scalar function of σ, as the curvature of a firm's isoquants does not vary with output. This approach avoids the potential endogeneity from including choice variables A^* or Y^* on the right-hand side of a regression equation.

One might worry that this framework will capture a "mechanical" substitution effect. After all, if the CAA causes plants in non-attainment counties to reduce their air emissions and leave water emissions unchanged, the ratio $\frac{W^*}{A^*}$ will increase. But this is actually evidence of substitution, as apparent from Figure 2.1. In the left-hand panel, the price of air emissions rises from p_{A0} to p_{A1}. Holding total cost TC and water emissions fixed, the firm's new input bundle is (W_1, A_1) at lower output Y_1. Water emissions are unchanged (by construction), but air emissions are lower. This change, however, incorporates both output and substitution effects. The right-hand panel removes the output effect by drawing a cost line (in green) at the new prices and the original output level Y_0. The input bundle is now (W_2, A_2), where $W_2 > W_0$. Holding output fixed, water emissions have actually increased.

The preceding discussion assumes a static production technology, with input substitution driven by exogenous price changes. This assumption may be incorrect if firms respond to regulation with both technology changes (e.g. installation of new pollution-control capital) and input substitution. Such a firm response might be thought of as an increase in a third factor (capital), or instead as a technological change in the constants c_W and c_A. Under a CES functional form assumption, the latter case is unproblematic. Factoring equation 2.1 yields the following.

$$\ln\left(\frac{W^*}{A^*}\right) = \frac{1}{1-\rho}\left[\ln\left(\frac{c_W}{c_A}\right) + \ln\left(\frac{p_A}{p_W}\right)\right] \tag{2.3}$$

Given a proxy for the quantity $\left[\ln\left(\frac{c_W}{c_A}\right) + \ln\left(\frac{p_A}{p_W}\right)\right]$, it is still possible to recover a scalar function of the elasticity of substitution $\sigma = \frac{1}{1-\rho}$. If the CES assumption fails in this case, my estimates are no longer informative about the substitution elasticity σ. Instead they will capture the full effect of CAA regulation on the log optimal input ratio, which may be a function of multiple underlying structural parameters.

2.4 Data

My plant-level emissions and location data come from the EPA Toxic Release Inventory (TRI) 1987-2010. TRI records annual emissions of more than 500 chemicals by mass (in pounds or grams). TRI data encompass a broad set of industries, from electric power to soybeans. The top ten industries by total TRI-reportable emissions are listed in Table 2.2. The database also includes the Dun & Bradstreet DUNS number for the parent company of each plant.

These data have several shortcomings, discussed in Hamilton (2005). Only large facilities are required to participate.[6] Firms typically report estimates derived from engineering models, rather than direct measurements. Gamper-Rabindran (2006) finds that the location variables are sometimes inaccurate. Under TRI there are penalties for false reporting, but not high emissions, which should ameliorate firm incentives to under-report emissions. The EPA has fined firms up to $27,000 per day for reporting problems in the past (Gamper-Rabindran, 2009). In the early years of TRI data collection, reporting requirements changed dramatically. For example, reported pollution increased sixfold between 1990 and 1991 due to reporting changes required by the Pollution Prevention Act (Environmental Protection Agency, 2012). To avoid confounding such reporting changes with genuine emissions changes, I exclude the period 1987-1992 from my analysis.

A subset of TRI chemicals are classified as particulates (PM).[7] The TRI data

[6]Reporting thresholds have varied over time and by chemical. Typically firms must report if they use or process more than 10,000 pounds of a TRI-listed chemical per year.

[7]Professor Michael Greenstone generously shared his mapping from TRI chemicals to CAA criteria pollutants. Details are available in Greenstone (2003). These data also include mappings to lead and VOC,

capture emissions in great detail, distinguishing for example between different types of underground wells. To simplify presentation and analysis I aggregate up to the categories described in Table 2.3 by adding the mass of each chemical emitted (in pounds).

Data on county attainment status come from the EPA Green Book. Monitor-level data on pollutant concentrations come from the EPA Air Quality System (AQS) 1993-2010. For descriptive statistics see Appendix Table 2.15.

2.5 Estimation

2.5.1 Estimating equations

To estimate treatment effects on air emissions, I use the following specification, with i indexing plant and t year.

$$\ln(A_{it}) = \overline{\alpha_i} + \overline{\delta_t} + \beta treated_{it} + \varepsilon_{it} \tag{2.4}$$

The dependent variable is the log of a plant's air emissions. The equation includes plant fixed effects and year dummies, with the latter capturing secular forces influencing emissions. As discussed in Section 2.5.2 below, the variable $treated_{it}$ equals 1 for plants that were within two kilometers of a non-attainment monitor in year $t-1$. If CAA regulations are effective in reducing air emissions, I expect estimates of β to be negative. Because this specification does not control for output or other inputs, it captures the full effect of the policy, including both output and substitution effects.

---

which I do not employ. I do not analyze lead emissions because of the small number of treated plants. The VOC mapping is problematic because VOC are not directly regulated under the CAA. They are one of two primary precursors (the other is NO_x) of ozone, which is a CAA criteria pollutant. While one would expect ozone non-attainment to affect VOC emissions, the link is much less clear than for particulates, as not all VOC contribute substantially to ozone formation. EPA regulates PM_{10} (particles <10 microns in diameter) and $PM_{2.5}$ (<2.5 microns in diameter) separately, but the Greenstone data do not allow me to separately identify these categories. TRI does not include emissions of CO, NO_2, or SO_2.

In addition, to investigate the time pattern of effects, I estimate an event-study specification.

$$\ln(A_{it}) = \overline{\alpha_i} + \overline{\delta_t} + \sum_j \tau_j + \varepsilon_{ipt} \tag{2.5}$$

The variables τ_j are indicators for a time index defined relative to treatment. I include dummies for $\tau = -3$, $\tau = -2$, $\tau = -1$, $\tau = 0$, $\tau = 1$, $\tau = 2$, and $\tau \geq 3$, so the reference category is the average of years for which $\tau < -3$. Tau equals 0 in the first treated year. This means that if a county receives a non-attainment designation in year $\tau = -1$, some of its plants enter treatment the following year.

To test for cross-media substitution, I estimate the following.

$$\ln\left(\frac{W_{it}}{A_{it}}\right) = \overline{\alpha_i} + \overline{\delta_t} + \beta treated_{it} + \varepsilon_{it} \tag{2.6}$$

As before, I include plant fixed effects and year dummies. The quantity $\ln\left(\frac{W_{it}}{A_{it}}\right)$ is the plant's log emissions ratio, with the numerator emissions into another medium (e.g. water or land) and the denominator air emissions. The estimating equation closely parallels the ratio of unconditional factor demands from equation 2.1 above. The treatment dummy proxies for the unobservable shift in the price ratio $\frac{p_A}{p_W}$ or, alternatively, the combination of price changes and technology changes in $\frac{c_W}{c_A}$. The coefficient $\beta = \nu\sigma$ is a scalar function of the elasticity of substitution $\sigma = \frac{1}{1-\rho}$, where ν is the percentage change in relative prices produced by treatment. If air and water emissions are substitutes, theory predicts the CAA will induce cross-media substitution and estimates of β will be positive. If instead air and water emissions are complements, estimates of β will approach zero.[8]

To test for within-firm leakage, I estimate the following specification using all plants in attainment counties.

[8] If the true aggregate production function includes three or more inputs, β is a function of the Morishima elasticity and negative values are possible.

$$\ln(A_{it}) = \overline{\alpha_i} + \overline{\delta_t} + \gamma multiplant_{it} + \beta multiplant_{it} * other\_treated_{it} + \varepsilon_{it} \tag{2.7}$$

Again I include plant fixed effects and year dummies. The variable $multiplant_{it}$ is a dummy for being part of a multi-plant firm. The variable $other\_treated_{it}$ is a dummy for one or more treated plants within the same firm, year, and 6-digit NAICS code. If the CAA induces spatial leakage, estimates of β will be positive.

Finally, in order to test indirectly for general-equilibrium leakage to untreated plants, I estimate the following equation, again using plants in attainment counties.

$$\ln(A_{it}) = \overline{\alpha_i} + \overline{\delta_t} + \beta total\_treated_{jt} + \varepsilon_{it} \tag{2.8}$$

In this specification the variable $total\_treated_{jt}$ is a count of treated plants at either the state-year level or the state-year-NAICS6 level. If general-equilibrium leakage is in fact occurring, estimates of β should be positive.

All of the above specifications involve logs or log ratios and thus exclude plants reporting zero emissions into a given medium. Plants that do not emit into the air are beyond the scope of this work. Plants that do not emit into a given non-air medium (e.g. water) before treatment, but begin emitting under treatment, are potentially interesting. As a practical matter there are very few such plants–fewer than 25 for any medium. Nonetheless I am investigating hurdle models that better account for such behavior.

2.5.2 Defining treatment

Past research on cross-media substitution has typically defined treatment as presence in a non-attainment county, but this conceals important spatial heterogeneity.

Auffhammer et al. (2009) find the effect of county non-attainment status on an average monitor is zero, but the effect on a non-attainment monitor is negative 11 to 14 percent. This suggests that regulators treat plants near non-attainment monitors intensively, while treating plants farther away lightly or not at all. I present evidence in support of this hypothesis. First I estimate a simple regression of a plant's air emissions on plant fixed effects and year dummies:

$$\ln(A_{it}) = \overline{\alpha_i} + \overline{\delta_t} + \varepsilon_{it} \tag{2.9}$$

In this equation A denotes air emissions, while i indexes plant and t year. Figure 2.2 is a local linear regression fit to plant residuals from non-attainment counties against the distance to the nearest non-attainment monitor. It provides evidence that regulators indeed treat plants near non-attainment monitors intensively, while treating more distant plants lightly or not at all. This pattern is consistent with the hedonic results from Bento et al. (2014).

Based on this pattern, I define a variable $treated_{it}$ equal to $Nonattain_{it-1} * 1\left\{Distance_{it-1} \leqslant \overline{D}\right\}$. That is, I consider a plant *treated* in year t if in the prior year its county was in non-attainment and the plant was located "close" to a non-attainment monitor. Based on Figure 2.2, I use a threshold distance $\overline{D}$ of two kilometers.[9] I use lagged rather than contemporaneous non-attainment status because: 1) state regulations may not take effect in the first non-attainment year (see Section 2.2); and 2) some firm responses plausibly require substantial time to implement (e.g., existing contracts might limit fuel switching). This treatment variable forms the basis for all subsequent results. Defining treatment in this way invokes an additional identifying assumption, exogeneity of monitor placement with respect to plant-level scope for abatement and substitution,

[9]While this pattern holds on average, it need not hold for all industries and pollutants. Stack height provides one source of heterogeneity. If a plant has tall stacks, it exerts more influence on distant monitors than on those nearby (author's interview notes). In such a case, even if regulators focus on particular plants, they may not be the plants adjacent to non-attainment monitors.

which I discuss in Section 2.5.3. This spatial pattern is consistent with a regulator whose objective function involves minimization of enforcement costs, either pecuniary or political (Amacher and Malik, 1996). The qualitative evidence presented by Becker and Henderson (2000) on regulator-firm negotiations is also consistent with such an explanation.

2.5.3 Treatment exogeneity

I cannot recover the causal effects of treatment unless it is exogenous to my plant-level outcomes of interest. Concretely, I assume exogeneity of: 1) county-level attainment status; and 2) distance to the nearest non-attainment monitor. As for the first assumption, past literature has typically argued that county non-attainment is exogenous.[10] Chay and Greenstone (2003a); Chay and Greenstone (2003b); Chay and Greenstone (2005) document that PM_{10} non-attainment counties do not differ systematically from attainment counties on observable dimensions (including economic shocks), either in levels or in changes. Appendix Table 2.15 shows that the emissions profiles of plants in *PM* attainment and non-attainment counties are not statistically different in my data.

Non-attainment is plausibly exogenous if a given firm produces a small portion of the ambient air pollution in a county. For the *average* plant in a non-attainment county, this is a tenable assumption. Motor vehicles typically account for the majority of *PM* pollution, especially in urban areas. The California Air Resources Board estimates that 74 percent of PM_{10} emissions come from non-point sources like road dust and from residential fuel combustion (Auffhammer et al., 2011).

The spatial heterogeneity documented in Section 2.5.2, however, calls into question the exogeneity of CAA regulation for *treated* plants (plants actually affected by

[10] Examples include Henderson (1996); Becker and Henderson (2000); Greenstone (2002); Auffhammer et al. (2011); Walker (2011).

regulation). CAA regulations primarily affect plants within two kilometers of a non-attainment monitor. It might be that past emissions by a given plant were pivotal in pushing its county above the CAA standard. If that were the case, CAA regulation would be endogenous to past emissions by treated plants. For example, if a plant experienced particularly strong demand for its output in a given year, it might have emitted more air pollution than usual and pushed the nearby monitor above the CAA standard.

This potential problem provides additional motivation for my use of emissions ratios, rather than emissions levels, in my analysis of cross-media substitution. If the endogeneity of non-attainment with respect to past emissions stems from output shocks, then treatment should remain exogenous to emissions ratios. It is still possible, however, that a plant might push its county into non-attainment because of shocks to emissions ratios. This form of endogeneity is perhaps less plausible, but impossible to exclude in principle. For example, a plant's scrubber might fail in a given year, increasing its ratio of air to water emissions and pushing its county into non-attainment. The sign of the bias in such a case would depend on the autocorrelation in the shocks to emissions ratios. Endogenous past output could also bias my estimates of CAA treatment effects on the level of air emissions. For example, if output shocks were negatively autocorrelated, my estimates might overstate CAA treatment effects. If instead output shocks were positively autocorrelated, it might understate them.

Figure 2.3 investigates the possibility of endogenous entry into treatment using an event study framework (estimates from equation 2.5). I define a new time index τ relative to treatment. A county receives a non-attainment designation in year $\tau = -1$ and plants within 2km of a non-attainment monitor enter treatment in the following year ($\tau = 0$). If the figure showed either higher air emissions or a lower ratio of other emissions to air emissions at $\tau = -1$, that would be evidence of endogenous entry into treatment. Instead the figure shows the opposite pattern. Air emissions fall in the final pre-treatment year and the ratio of other emissions to air emissions increases. I attribute this pattern to two sources: 1) by time $\tau = -1$, firms have potentially known of their state's implementation plan (SIP) for several years and may have begun to respond; and 2) some provisions of the SIP may take effect during year $\tau = -1$, and some firms may be able to respond within the year. Any such anticipatory behavior or rapid adjustment by firms will bias the magnitudes of my estimates downward. Note however that for most plants I observe long pre- and post-treatment periods. By estimating in levels with 18 years of data, I largely mitigate potential biases from anticipatory behavior or rapid firm responses.

My second identifying assumption is exogeneity of distance to the nearest non-attainment monitor. Violations of this assumption could spring from two sources: firm location decisions and state monitor placement decisions. Given the relatively low cost of new monitors, firms are unlikely to profit by strategically locating away from existing monitors. The state monitor location decision warrants more discussion. States design monitoring networks, which must follow EPA rules and which EPA must approve (CFR, 2015). EPA may also suggest changes to planned networks. Importantly in this setting, the agency's placement rules largely depend on population characteristics, not firm characteristics. For example, EPA requires monitors in areas of high population density (Bento et al., 2014) and near large sensitive populations (e.g. asthmatic children Raffuse et al., 2007). Two types of monitoring sites raise potential endogeneity concerns: "Sites located to determine the impact of significant sources or source categories on air quality" and "Sites located to determine the highest concentrations expected to occur in the area covered by the network" (CFR, 2015). Monitors placed under these two rules could be correlated with unobservable time-varying characteristics of plants, as discussed below. States are prohibited from putting monitors in locations that do not meet scientific criteria. In most cases it is illegal for a state to move a monitor, and EPA allows relocation only if the new site is better under its scientific criteria. Should a state fail to follow these rules, EPA may file suit against it (Chay and Greenstone, 2005).

Note that my identifying assumption is *exogeneity of distance to the nearest non-attainment monitor,* not distance to the nearest monitor. The former is a weaker assumption, particularly given the event-study evidence that the plants in my data are not pivotal in putting their counties into non-attainment. Nonetheless, to investigate potential endogeneity, I first regress the distance to the nearest non-attainment monitor on a set of 317 dummies for six-digit NAICS codes, omitting the constant term. Figure 2.4 displays the probability density function of the coefficient estimates. While the distribution is roughly normal around a mean of 11.7 kilometers, some coefficients are statistically distinguishable from that mean in both the positive and negative directions. Industries in the right tail show no clear pattern. They include, for example, beet sugar manufacturing, prisons, and national defense. The R^2 from the regression is .6, indicating that industry explains a substantial fraction of the variation in distance to the nearest non-attainment monitor. This suggests that plant fixed effects are necessary to my identification strategy, but even with plant fixed effects the possibility of non-zero covariance between time-varying plant unobservables and monitor distance remains. To evaluate this threat to identification, I regress the distance to the nearest non-attainment monitor on a vector of year dummies and the changes in log emissions into various media for untreated plant-years (pre-treatment or farther than two kilometers from the nearest non-attainment monitor). A negative coefficient is consistent with states strategically placing monitors near faster-growing emissions sources. Table 2.4 shows that the estimates are positive in five of eight cases and generally not statistically significant. Emissions growth rates in untreated plant-years generally do not systematically predict distance from eventual non-attainment monitors. The estimate for "onsite other" is negative and significant, however. If states are placing monitors closer to plants with faster growth in their emissions to waste piles, this will potentially bias my estimates for the "onsite other" channel. (For a version of this specification with emissions levels instead of growth rates, see Appendix Table 2.18.)

2.6 Results

2.6.1 Air emissions

Table 2.6 presents my estimate of the CAA treatment effect on airborne particulate emissions. Treated plants decrease their air emissions by 23 percent. This is larger than the 11 to 14 percent effect on non-attainment monitors reported by Auffhammer et al. (2011) because: 1) plant emissions become diluted as they mix with surrounding air; and 2) the treated plants in my sample are not the only factor influencing ambient air pollution. Column 2 adds state linear time trends. This reduces the magnitude of the estimate modestly, from 23 to 19 percent, but it remains statistically significant at the five percent level. These estimates reflect both output and substitution effects. If there is substantial general-equilibrium leakage to untreated plants, my estimated effects on air emissions will be biased upward in magnitude. I investigate this possibility in Section 2.7.1 and provide evidence this is not a substantial concern. Bento et al. (2014) show these air quality improvements disproportionately benefit low-income people.

Column 3 presents the results from an event-study specification (equation 2.5), where again the dependent variable is log air emissions. The time pattern suggests that most of the emissions reductions occur when τ is 0 or 1. (For discussion of the pre-treatment decline, see Section 2.5.2.) This helps motivate my use of fixed-effects models in levels. Estimates based on changes in treatment status would be biased toward zero because of the emissions decline at $\tau = -1$. With a relatively long pre-treatment period, however, a model in levels averages this pre-treatment decline with other untreated years, mitigating the bias. At approximately -36 percent, the event-study estimates are close in magnitude to my primary result (-23 percent). Together these results demonstrate that treated plants do indeed reduce airborne particulate emissions.

2.6.2 Cross-media substitution, all industries

Panel A in Table 2.7 shows estimated treatment effects from equation 2.6, by medium across all industries. The dependent variable is a log emissions ratio, with emissions into a given medium (indicated in the column heading) in the numerator, and air emissions in the the denominator. Positive estimates are evidence of cross-media substitution. There is evidence of statistically significant substitution toward onsite water emissions $\left(\hat{\beta} = .44\right)$, offsite water emissions $\left(\hat{\beta} = .26\right)$, and recycling $\left(\hat{\beta} = .41\right)$. ("Onsite other" emissions include waste piles, leaks, and spills.) The negative estimate for onsite land emissions demonstrates that the ratio approach does not assume positive elasticities of substitution.

Panel B in Table 2.7 adds state linear time trends. The estimates for onsite water $\left(\hat{\beta} = .49\right)$, offsite water $\left(\hat{\beta} = .26\right)$, and recycling $\left(\hat{\beta} = .45\right)$ are essentially unchanged from my primary results in panel A. The EPA is aware of the potential for substitution and in some cases has taken steps to mitigate it (Environmental Protection Agency, 2001). In particular, EPA is currently developing rules to restrict water emissions from power plants (Environment News Service, 2013). The potential for pollution substitution remains, however, in a wide variety of industries.

The large increase in recycling highlights the fact that not all substitution responses work against the intent of the CAA. The increased water emissions, however, impose social costs. The magnitude of those costs is difficult to quantify, given the relative scarcity of well-identified studies on the health and productivity effects of water pollution. On a net basis, cross-media substitution need not reduce welfare. Suppose a plant responds to non-attainment by reducing output and substituting toward water emissions. Gross water emissions may end up below their initial level. In such a case the CAA may still improve welfare, but substitution attenuates the gains if water emissions are mispriced, relative to a counterfactual world with an optimal policy.

2.6.3 Cross-media substitution, by industry

It is difficult to analyze substitution patterns at the industry level due to the small number of treated plants: recall that not all plants in non-attainment counties are treated. Moreover not all plants report emissions into all media. Nonetheless, to illustrate the heterogeneity in substitution responses, Table 2.9 presents estimates for the three industries with the largest treated sample sizes: primary metals, wood products, and utilities. (Appendix Table 2.25 presents more disaggregated estimates at the 3-digit NAICS level.) Estimates again come from equation 2.6. In the discussion that follows, note that I cannot reject the null hypothesis of equal coefficients in most cases; the evidence of heterogeneity is merely suggestive. Wood products and utilities show large decreases in air emissions, 47 and 36 percent respectively, while primary metals show only a 8 percent decrease. Similarly, wood products and utilities increase their ratios of water to air emissions by 59 and 118 percent, while primary metals increase this ratio by only 33 percent. The primary metals industry increases the use of waste piles ("onsite other") by more than 180 percent. Only utilities substantially increase their ratio of offsite land disposal to air emissions, but the effect is very large at 174 percent. Both wood products and utilities increase their ratio of recycling to air emissions by much more than does the primary metals industry, again suggesting the former two industries are more intensively regulated or have greater scope for substitution.

2.6.4 Leakage

Intuition predicts that firms might respond to treatment of a plant in one county by shifting output to a plant in another county. Table 2.11 provides evidence they do so. Estimates correspond to equation 2.7. For the average plant in an attainment county, treatment of another plant within the same firm and 6-digit NAICS code increases air emissions by 17 percent. Column (2) adds state linear time trends and the estimate is slightly smaller at 15.5 percent. Treating the number of other treated plants as a continuous variable (column 3), estimated leakage is 12 percent per treated plant. With the addition of state linear time trends in column 4, the estimate is again slightly smaller at 11 percent, but remains statistically significant at the ten percent level. This leakage has associated health, mortality, and productivity costs. As a robustness check I estimate the same model, grouping plants by firm and 2-digit NAICS code, and report results in Table 2.29. Estimates are modestly smaller than in my preferred specification, though still positive and significant. This is reasonable, as the coarser classification groups plants that may not be close substitutes for each other. I take the TRI parent company identifiers as given. If they are defined at a level below the ultimate corporate parent, my estimates will likely understate the true amount of leakage. Likewise, if there is general-equilibrium leakage to plants owned by other firms in attainment counties, my estimates will be biased downward (see Section 2.7.1).

The identifying assumptions for this model are modestly stronger than for my model of cross-media substitution and warrant brief discussion. Limiting the sample to attainment-county plants changes the interpretation of the estimates, but is not in itself problematic, especially if attainment status is exogenous. Interpreting the estimates in Table 2.11 as causal, however, also requires that the leakage plants do not differ from other attainment-county plants in time-varying, unobservable ways. Appendix Table 2.16 shows that emissions profiles for leakage and non-leakage plants are not significantly different, which is reassuring but does not exclude the possibility of endogeneity.

The average treated plant in my data is the only treated plant within its firm, and on average that firm includes approximately three plants that are candidates for leakage-driven increases: they share the same six-digit NAICS code and are located in attainment counties.[11] Average air emissions at eventually treated plants prior to treatment are 4394 pounds, while average baseline emissions at leakage candidates are 2318 pounds. The estimate from column one of Table 2.11 implies the following net change in emissions from treating an average plant. The treated plant reduces emissions by $.23 * 4394 = 1010$ pounds. The three candidate plants together increase emissions by $3 * .17 * 2318 = 1182$ pounds. On net, then, the average CAA plant treatment increases firm-level particulate emissions by $1182 - 1010 = 172$ pounds. This result should be interpreted with several important caveats in mind. First, the TRI data cover only large plants, which may be more likely to belong to multi-plant firms and thus may have more scope for within-firm leakage. Second, these estimates describe only TRI-reportable particulate emissions. Third, leakage patterns might differ for other CAA-regulated pollutants (e.g. SO2). Fourth, industrial sources account for approximately 25 percent of particulate emissions in an average county (Auffhammer et al., 2011), so the implied changes in ambient pollution are much smaller than the emissions changes I estimate at the plant level.

Leakage reduces the welfare gains from CAA regulation, relative to a first-best policy, because attainment-county emissions are unpriced (unregulated). This leakage need not imply a net welfare loss, however. Leakage-driven emissions increases occur in attainment counties, which by definition have lower ambient air pollution. In addition, the average attainment county population is approximately 1/3 of the average non-attainment

[11] This is the average number of leakage candidates over all treated firms, including single-plant firms that have zero leakage candidates by definition.

county population.[12] Particularly if the social damage function for air pollution is convex, the net welfare effect from CAA treatment of the plants in my data may be positive. Leakage does present a potential problem in using difference-in-differences designs to evaluate the CAA, as it is a spillover from the treatment group (typically non-attainment counties) to the control group (attainment counties). The spillovers identified in Table 2.11 imply that such analyses overstate CAA benefits in non-attainment counties and fail to account for some of the costs in attainment counties. This provides additional motivation for my use of an emissions ratio, rather than a level, in my analysis of cross-media substitution.

2.7 Additional results, robustness & placebos

2.7.1 Air emissions

It is possible CAA regulation induces general-equilibrium leakage, with output reallocated from treated plants to attainment-county plants not owned by the same firm. If this is the case, my estimated effects on air emissions at treated plants will be biased upward in magnitude. My estimated within-firm leakage effects will be biased downward in magnitude. Under a CRS assumption, general-equilibrium effects will not bias my estimates of cross-media substitution. It is impossible to test directly for general-equilibrium leakage, since all plants are potentially affected by CAA regulation through general-equilibrium mechanisms. One can however test indirectly for general-equilibrium leakage by modeling the air emissions at untreated plants[13] as a function of the number of treated plants "nearby." To that end I estimate equation 2.8 and report results in Appendix Table 2.20. In the first specification, the estimated coefficient on the

[12]Author's calculation from 2010 Census data.

[13]i.e. Attainment-county plants and plants farther than 2km from a non-attainment monitor in a non-attainment county

number of treated plants in the same state-year is negative, insignificant, and zero to two decimal places. In the second, the estimated coefficient on the number of treated plants in the same state, year, and six-digit NAICS code is .048 and significant at the ten percent level. Together these estimates suggest that if general-equilibrium leakage is occurring, it is relatively small in magnitude.

It is possible that intra-firm leakage causes my treatment model to overestimate the air emissions reductions undertaken by treated plants. To evaluate this possibility, I estimate a variant of my air emissions model (equation 2.4), controlling for spillovers as in equation 2.7. Reported in Appendix Table 2.21, the estimates are unchanged.[14]

Appendix Table 2.22 presents results based on toxicity-weighted air emissions. The treatment effect is larger in magnitude, at $-.51$, and significant at the five percent level. I do not employ toxicity weights in my preferred specifications for the following reasons: 1) toxicity weights for a given chemical can vary by three orders of magnitude, depending on the method used (Hertwich et al., 1998); and 2) toxicity weights rely on assumptions that some chemicals are not carcinogenic, but epidemiological evidence suggests such assumptions may not hold (Hendryx et al., 2012).

Both Henderson (1996) and Becker and Henderson (2000) show that CAA non-attainment influences plant entry and exit decisions, and this is a potential source of bias. A Heckman correction would be inappropriate, as I do not have any variables that would enter the selection equation but not the outcome equation. Instead I restrict the sample to plants present throughout the study period and estimate treatment effects on air emissions (see Appendix Table 2.23). While the estimates are no longer statistically significant due to the much smaller sample, at $-.26$ and $-.23$ they are remarkably close to the results in Table 2.6. This suggests selection does not meaningfully bias my main results.

[14]The potential bias in my main specification would come from the influence of the spillover plants on the estimates for year dummies. The failure to control for spillovers has no practical import because only a small number of plants are affected by spillovers. Identification of the year dummies comes primarily from plants that do not receive spillovers.

Lastly I estimate a specification where the dependent variable is the growth rate of air emissions (the difference in logs), similar to the specifications employed by Greenstone (2003) and Gamper-Rabindran (2009). Results are in Appendix Table 2.24. The estimated treatment effect is $-.03$, roughly similar to the Greenstone and Gamper-Rabindran results, and not statistically significant. This demonstrates the importance of modeling air emissions in levels, rather than growth rates.

2.7.2 Cross-media substitution

Table 2.12 moves from a ratio specification based on equation 2.1 to a specification in levels. Only the recycling estimate remains statistically significant.[15] The estimated 24 percent increase in the level of water emissions, while not statistically significant ($p = .25$), is practically important. It suggests that plants may be substantially increasing water emissions in response to CAA regulation. If such is indeed the case, the estimated elasticity of substitution from my preferred specification is an important consideration for policy design.

To test whether spillovers influence my cross-media results, I estimate my leakage model using an emissions ratio as the dependent variable and report results in Appendix Table 2.26. Estimates are generally near zero and statistically insignificant, with one exception: the estimate for “offsite other” is negative 66 percent. This spillover will tend to bias my cross-media model toward finding evidence of substitution toward the “offsite other” channel.

In addition, Appendix Table 2.27 shows estimates for toxicity-weighted emissions ratios, indicating firms shift their most toxic emissions into on-site waste heaps and recycling facilities. As discussed in Section 2.7.1, however, toxicity weights are

[15]This result indicates recycling is not a good output proxy, as firms typically face higher costs under environmental regulation and therefore reduce output (Becker and Henderson, 2000).

problematic in several respects. Finally, Appendix Table 2.28 presents results aggregated across all non-air media, for specifications in both ratios and levels. The elasticity of substitution from my preferred ratio specification is .27 and the estimate is significant at the five percent level.

2.7.3 Placebos

Treatment should have no direct effect on plants that do not emit any air pollution, and the results from Appendix Table 2.20 suggest that general-equilibrium treatment effects are negligible. Table 2.13 tests this hypothesized null effect by estimating a variant of equation 2.6 with two changes: 1) the dependent variable is log emissions into a given medium (the lack of air emissions precludes a ratio); and 2) treatment is interacted with a dummy indicating zero air emissions. If my model is well specified, it should find no effect of CAA regulation on these plants. The estimates are indeed insignificant, with the exception of the one for onsite land emissions. While the latter is statistically significant, it is negative. If there were some omitted variable decreasing land emissions at plants near non-attainment monitors, it would work against finding cross-media substitution (it would bias my primary estimates downward).

Table 2.14 reports results from a placebo test of my leakage model. I construct variables based on placebo "treated" plants: plants within the same firm and 6-digit NAICS code that are located in non-attainment counties, but farther than eight kilometers from the nearest non-attainment monitor. As these plants are not treated and general-equilibrium effects are not apparent, one should not see increased air emissions by attainment-county plants in the same firm and NAICS code. If my leakage model is capturing, for example, changes in the geographic distribution of output that happen to be correlated with treatment, this placebo test should return large positive estimates. Instead the estimates in Table 2.14 are in the one to three percent range and are not statistically significant. This suggests that the leakage results in Table 2.11 do not spring from an omitted variable problem.

2.8 Conclusion

While economists have long recognized the potential for substitution responses to single-medium pollution regulation, empirical studies have found little evidence of such effects. The paucity of available data and the difficulty of controlling for scale have made firm responses hard to detect. Using specifications motivated by classical firm optimization theory, this study provides evidence of regulation-induced pollution substitution in response to the Clean Air Act. Estimates from EPA Toxic Release Inventory data show that CAA-regulated plants increase their ratio of water to air emissions by 44 percent. Particulate regulation of an average plant increases air emissions at unregulated plants owned by the same firm by 17 percent. This results in a net emissions increase. Responses of this magnitude plausibly have social costs and should be considered in policy design. The welfare effects of such substitution present an interesting subject for future research. Additionally, I document spatial heterogeneity in regulatory intensity, which suggests that regulators seek to minimize costs (political or pecuniary) in implementing the CAA.

My ratio estimation approach could be used to recover other policy-relevant substitution elasticities. Given current policy focus on carbon abatement, well-identified estimates of the carbon-labor and carbon-capital substitution elasticities would be valuable. The *PM* elasticities estimated in this study might also helpfully inform the design of future pollution control policy. A maximally efficient policy, with emissions into every medium and location priced according to marginal damage, would be difficult to achieve and might not be desirable for normative reasons. The primary goal of the Clean Air Act is safeguarding human health (Environmental Protection Agency, 2011), not economic efficiency. Given any set of policy goals, however, it is easier to formulate effective policy when policymakers have well-identified estimates that allow prediction of firm responses. For example, my cross-media results suggest that restricting water emissions or increasing water quality monitoring in CAA non-attainment counties might be important for protecting public health. My estimates of within-firm leakage imply that applying PM-style regulations to carbon emissions would be largely ineffective. Legislators might also want to consider the regulator behavior implied by my spatial heterogeneity result when designing future policy.

Such improvements in policy design would likely have economically significant consequences. While environmental economics research initially focused on the mortality effects of air pollution, especially for infants and the elderly, there is growing evidence that air pollution has costly effects on healthy adults. Isen et al. (2014), for example, find that in-utero and early childhood air pollution exposure depresses earnings for workers ages 29-31. Zivin and Neidell (2012) find air pollution decreases worker productivity. Given these large costs, the returns to improved pollution regulation may be large.

Acknowledgements: I would like to thank Maximilian Auffhammer, Prashant Bharadwaj, Richard Carson, Andrew Chamberlain, Julie Cullen, Joshua Graff Zivin, Michael Greenstone, Kelsey Jack, Mark Jacobsen, Lynn Russell, Jeremy Schreifels, Ron Shadbegian, Glenn Sheriff, Jeffrey Shrader, Larry Sorrels, Junjie Zhang and participants in the UC San Diego environmental seminar for invaluable assistance with this project. Chapter 2, in full, is currently being prepared for submission for publication of the material.

References

Amacher, Gregory S and Arun S Malik (1996). “Bargaining in environmental regulation and the ideal regulator”. In: *Journal of Environmental Economics and Management* 30.2, pp. 233–253.

Auffhammer, Maximilian, Antonio M. Bento, and Scott E. Lowe (July 2009). “Measuring the effects of the Clean Air Act Amendments on ambient concentrations: The critical importance of a spatially disaggregated analysis”. In: *Journal of Environmental Economics and Management* 58.1, pp. 15–26.

Auffhammer, Maximilian, Antonio M Bento, and Scott E Lowe (2011). “The City-Level Effects of the 1990 Clean Air Act Amendments”. In: *Land Economics* 87.1, pp. 1–18.

Becker, Randy and Vernon Henderson (2000). “Effects of Air Quality Regulations on Polluting Industries”. In: *Journal of Political Economy* 108.2, pp. 379–421.

Becker, Randy A and J Vernon Henderson (2001). “Costs of air quality regulation”. In: *Behavioral and distributional effects of environmental policy*. University of Chicago Press, pp. 159–186.

Bento, Antonio, Matthew Freedman, and Corey Lang (2014). “Who Benefits from Environmental Regulation? Evidence from the Clean Air Act Amendments”. In: *Review of Economics and Statistics* 0.

Blackorby, Charles and R Robert Russell (1989). “Will the Real Elasticity of Substitution Please Stand Up ? (A Comparison of the Allen/Uzawa and Morishima Elasticities)”. In: *The American Economic Review* 79.4, pp. 882–888.

Bushnell, James B and Erin T Mansur (May 2011). “Vertical Targeting and Leakage in Carbon Policy”. In: *American Economic Review* 101.3, pp. 263–267.

Campbell, Harry Fleming (1991). “Estimating the elasticity of substitution between restricted and unrestricted inputs in a regulated fishery: a probit approach”. In: *Journal of Environmental Economics and Management* 20.3, pp. 262–274.

CFR (2015). *40 CFR Part 58, Appendix D to Part 58 - Network Design Criteria for Ambient Air Quality Monitoring*.

Chay, Kenneth Y and Michael Greenstone (2003a). *Air quality, infant mortality, and the Clean Air Act of 1970*. Tech. rep. National Bureau of Economic Research.

— (2003b). “The Impact Of Air Pollution On Infant Mortality: Evidence From Geographic Variation In Pollution Shocks Induced By A Recession”. In: *The Quarterly Journal of Economics* 118.3, pp. 1121–1167.

— (2005). “Does Air Quality Matter? Evidence from the Housing Market”. In: *Journal of Political Economy* 113.2.

Currie, Janet and Matthew Neidell (2005). "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?" In: *The Quarterly Journal of Economics* 120.3, pp. 1003–1030.

Davis, Lucas W and Matthew E Kahn (2010). "International Trade in Used Vehicles: The Environmental Consequences of NAFTA". In: *American Economic Journal: Economic Policy* 2.4, pp. 58–82.

Department of Energy (2014). *Cleaning Up Coal*. Accessed January 16, 2014.

Domike, Julie R and Alec C Zacaroli, eds. (2011). *The Clean Air Act Handbook*. American Bar Association.

Duhigg, Charles (2009). *Cleansing the Air at the Expense of Waterways*.

EC/R Incorporated (1998). *Stationary Source Control Techniques Document for Fine Particulate Matter*. Tech. rep. EPA Contract No. 68-D-98-026, Work Assignment No. 0-08. EC/R Incorporated.

Environment News Service (2013). *Court Orders EPA to Impose Power Plant Water Pollution Rule*.

Environmental Protection Agency (1999). *Land Disposal Restrictions for Hazardous Wastes*. Tech. rep. EPA530-F-99-043. Environmental Protection Agency.

— (2001). *Hazardous Air Pollutant Emissions for Miscellaneous Organic Chemical Manufacturing: Supplementary Information Document for Proposed Standards*. Tech. rep. Emission Standards Division, Environmental Protection Agency.

— (2008). *Clean Water Act Jurisdiction Following the U.S. Supreme Court's Decision in Rapanos v. United States & Carabell v. United States*. Tech. rep. Environmental Protection Agency.

— (2010). *Disposal of Coal Combustion Residuals From Electric Utilities; Proposed Rule*. Tech. rep. Federal Register 75. 118.

— (2011). *The Benefits and Costs of the Clean Air Act from 1990 to 2020 Final Report, Rev . A.* Tech. rep. Environmental Protection Agency.

— (2012). *Factors to Consider When Using Toxics Release Inventory Data*. Tech. rep. Environmental Protection Agency.

— (2013). *CWA Methods*. Accessed June 8, 2013.

— (Undated). *Air Pollution Control Technology Fact Sheet*. Tech. rep. EPA -452/F-03-032. Environmental Protection Agency.

Farnsworth, David (2011). *Preparing for EPA Regulations: Working to Ensure Reliable and Affordable Environmental Compliance*. Tech. rep. Regulatory Assistance Project.

Fowlie, Meredith (2009). “Incomplete Environmental Regulation, Imperfect Competition, and Emissions Leakage”. In: *American Economic Journal: Economic Policy* 1.2, pp. 72–112.

— (2010). “Emissions Trading, Electricity Restructing, and Investment in Pollution Abatement”. In: *The American Economic Review* 100.3, pp. 837–869.

Fullerton, Don and Daniel H Karney (2014). “Multiple Pollutants, Unregulated Sectors, and Suboptimal Environmental Policies”.

Gamper-Rabindran, Shanti (July 2006). “Did the EPA’s voluntary industrial toxics program reduce emissions? A GIS analysis of distributional impacts and by-media analysis of substitution”. In: *Journal of Environmental Economics and Management* 52.1, pp. 391–410.

— (2009). “The Clean Air Act and volatile organic compounds: Did plants reduce their health-indexed air emissions or shift their emissions into other media?”

Greenstone, Michael (2002). “The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures”. In: *Journal of Political Economy* 110.6, pp. 1175–1219.

— (2003). “Estimating Regulation-Induced Substitution: The Effect of the Clean Air Act on Water and Ground Pollution”. In: *American Economic Review: AEA Papers & Proceedings* 93.2, pp. 442–449.

Hamilton, James (2005). *Regulation through revelation: the origin, politics, and impacts of the Toxics Release Inventory Program*. Cambridge University Press.

Hanna, Rema (2010). “US environmental regulation and FDI: evidence from a panel of US-based multinational firms”. In: *American Economic Journal: Applied Economics* 2.3, pp. 158–189.

Henderson, J Vernon (1996). “Effects of Air Quality Regulation”. In: *The American Economic Review* 86.4, pp. 789–813.

Hendryx, Michael, Jamison Conley, Evan Fedorko, Juhua Luo, and Matthew Armistead (Jan. 2012). “Permitted water pollution discharges and population cancer and non-cancer mortality: toxicity weights and upstream discharge effects in US rural-urban areas”. In: *International Journal of Health Geographics* 11.1, p. 9.

Hertwich, Edgar G., William S. Pease, and Thomas E. McKone (1998). “Evaluating Toxic Impact Assessment Methods: What Works Best?” In: *Environmental Science and Technology*.

Isen, Adam, Maya Rossin-Slater, and W Reed Walker (2014). *Every Breath You Take-Every Dollar You’ll Make: The Long-Term Consequences of the Clean Air Act of 1970*. Tech. rep. National Bureau of Economic Research.

Levinson, Arik and M. Scott Taylor (Feb. 2008). “Unmasking the Pollution Haven Effect”. In: *International Economic Review* 49.1, pp. 223–254.

Missoula County Environmental Health Division (1999). *History of Missoula’s Air Quality Program*. Accessed October 10, 2014.

Raffuse, S., D. Sullivan, M. McCarthy, B. Penfold, and H. Hafner (2007). *Analytical Techniques for Technical Assessments of Ambient Air Monitoring Networks*. Tech. rep. Environmental Protection Agency.

Sigman, Hilary (1996). “Cross-Media Pollution: Responses on Chlorinated Solvent Releases to Restrictions”. In: *Land Economics* 72.3, pp. 298–312.

Vatavuk, William, Donald van der Vaart, and James Spivey (2000). *VOC Controls*. Tech. rep. EPA/452/B-02-001. Environmental Protection Agency.

Walker, W Reed (2011). “Environmental regulation and labor reallocation: Evidence from the Clean Air Act”. In: *The American Economic Review* 101.3, pp. 442–447.

— (2013). “The Transitional Costs of Sectoral Reallocation: Evidence From the Clean Air Act and the Workforce”. In: *The Quarterly Journal of Economics* 128.4, pp. 1787–1835.

Zivin, Joshua Graff and Matthew Neidell (2012). “The Impact of Pollution on Worker Productivity”. In: *The American Economic Review* 102.7, pp. 3652–3673.

2.9 Figures and tables

2.9.1 Figures

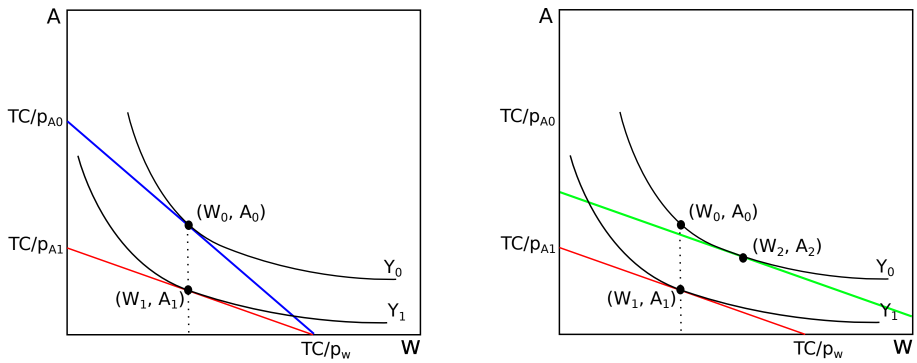


Figure 2.1: Pollution changes, holding output fixed

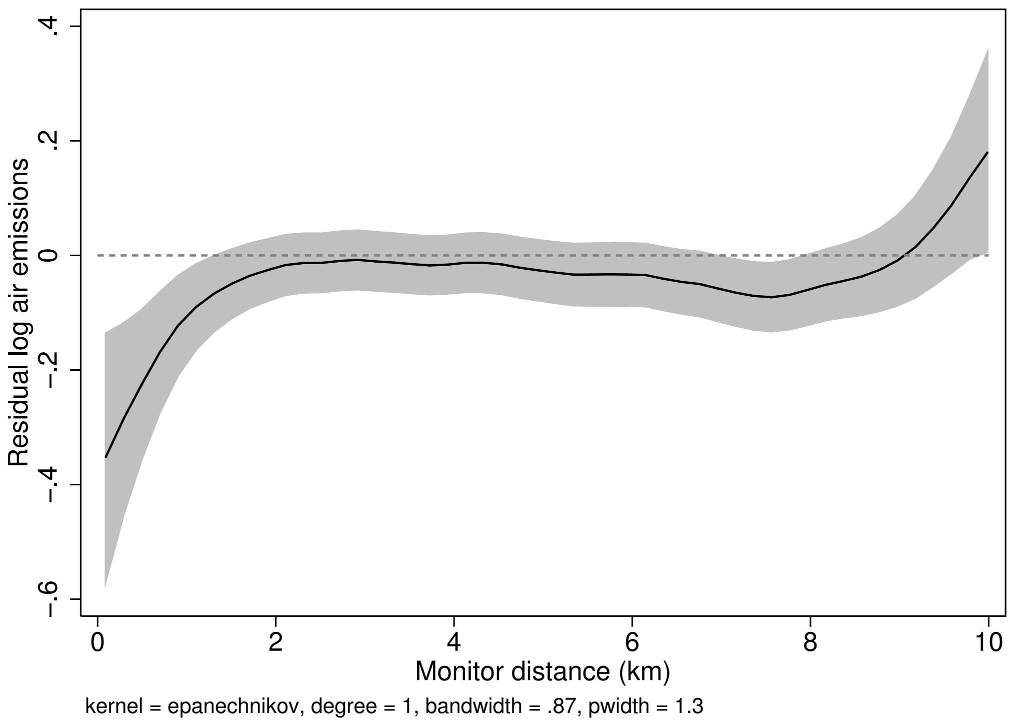


Figure 2.2: Residual air emissions by distance from nearest non-attainment monitor

Underlying residuals from equation 2.9, a panel model of log air emissions (lbs) with year dummies and plant fixed effects. The fitted line represents a local linear regression run on residuals for plants in non-attainment counties. Shaded area is the 95% confidence interval.

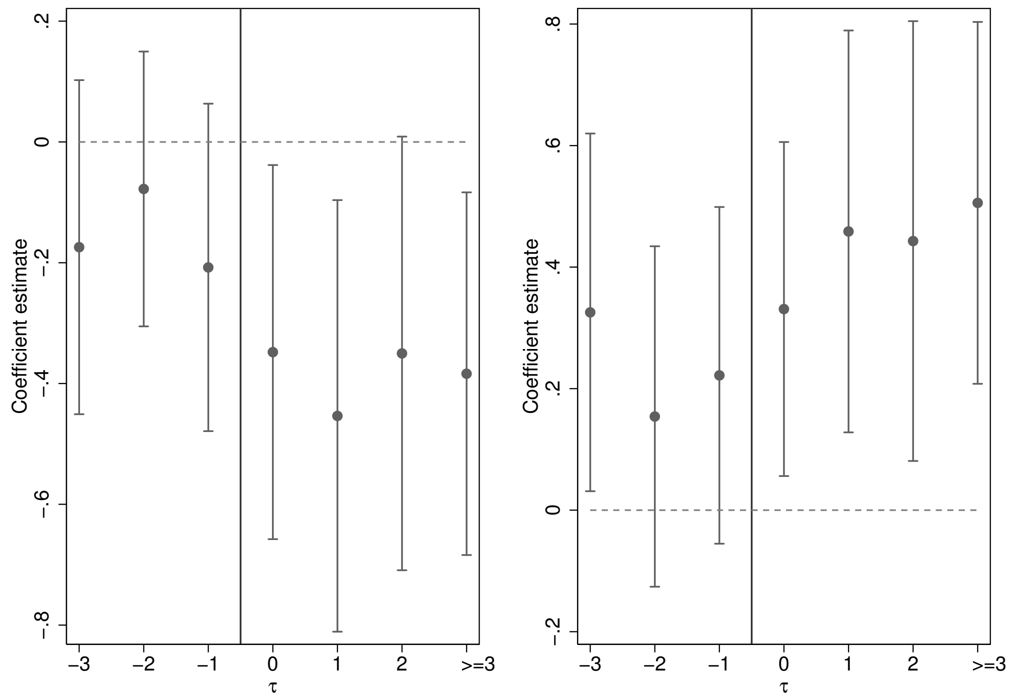


Figure 2.3: Event study estimates

Estimates from equation 2.5, also reported in column 3 of Table 2.6. Dependent variable is log air emissions (lbs). Reference category is years for which $\tau < -3$. A county enters non-attainment in year tau=-1 and plants within 2km of a non-attainment monitor enter treatment in the following year (tau=0). Dependent variable is log air emissions. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year.

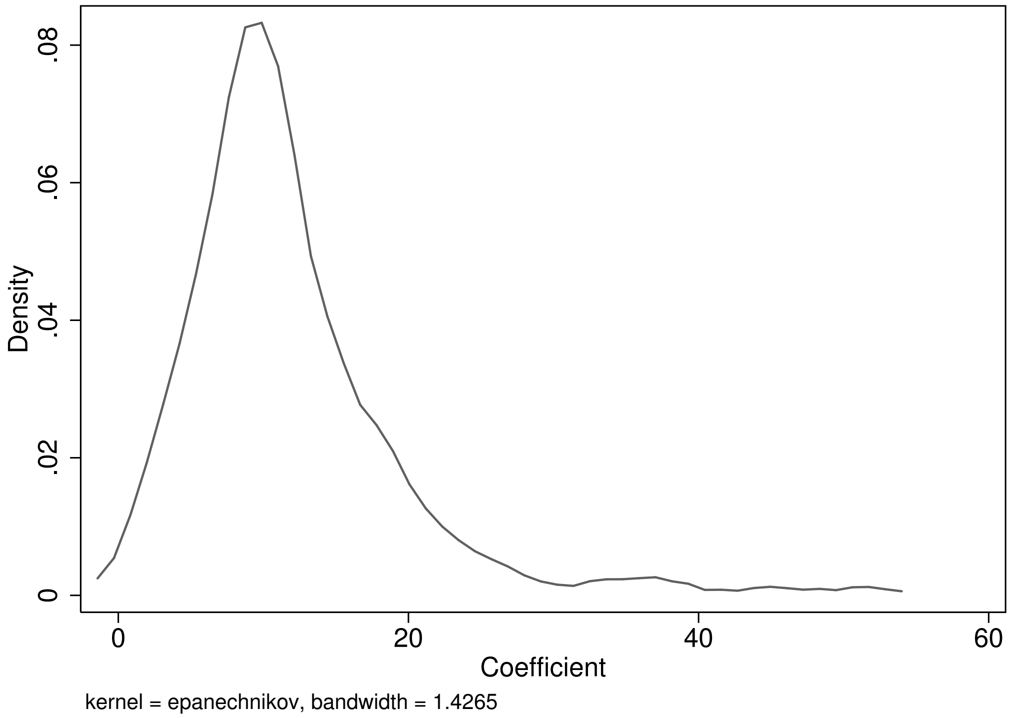


Figure 2.4: PDF of NAICS6 coefficients

Probability distribution function of estimates from a regression of distance to nearest non-attainment monitor on year dummies and 317 dummies for six-digit NAICS codes. Regression does not include a constant. $R^2 = .6$. Industries in the right tail show no clear pattern. They include, for example, beet sugar manufacturing, prisons, and national defense.

2.9.2 Tables

Table 2.1: Particulate abatement strategies

Name	Category	Description	Variable Costs	Secondary wastes
Output reduction	-	-	-	-
Reduce exhaust temp./pressure	-	Lower reaction temperature generates fewer particulates	Efficiency loss	-
Fuel switching	-	Switch to washed coal, oil, or natural gas	Added fuel cost	Coal slurry (offsite)
Process modification	-	e.g. Changing furnace type or cooling system	-	-
Flue gas conditioning	Pretreatment	Chemistry/temp./moisture modified to aid collection	Absorbant, electricity	Sulfates
Precollection	Pretreatment	Collectors use gravity/intertia to gather particles	Electricity	Solid waste
Electrostatic precipitation	End-of-pipe	Field charges particles, collected by electrode	Electricity, water	Liquid/solid waste
Fabric filters	End-of-pipe	Tightly woven fabric and dust layer trap particles	Electricity, filters	Solid waste
Wet scrubbers	End-of-pipe	Liquid (often sprayed) traps particles	Electricity, water	Liquid/solid waste
Incineration	End-of-pipe	Emissions burned at 300-2000^{o}F, sometimes catalyzed	Fuel, catalyst	CO2, N2, H2O
Ventilation	Fugitive control	e.g. Vacuum hoods, building enclosue	Electricity	Solid waste
Road paving	Fugitive control	-	Maintainance	-
Water spraying	Fugitive control	Wet down sources of fugitive emissions, e.g. coal piles	Water	Coal slurry

Sources: Department of Energy (2014); EC/R Incorporated (1998); Environmental Protection Agency (Undated); Farnsworth (2011); Vatavuk et al. (2000). Variable costs range from 33 to 100 percent of capital cost for most "end-of-pipe" abatement technologies. Incineration is typically used only for waste streams containing both PM and VOCs.

Table 2.2: Top ten industries, by TRI-reportable emissions

Rank	NAICS code	Industry
1	221112	Fossil electric power
2	325188	Inorganic chemicals
3	212231	Pb & Zn mining
4	212234	Cu & Ni mining
5	212221	Au mining
6	331111	Iron & steel
7	325199	Organic chemicals
8	322121	Paper
9	562211	Hazardous waste
10	324110	Petroleum Refining

Table 2.3: Aggregated TRI emissions categories

Aggregated category	Included TRI components
Onsite air	Fugitive air, stack air
Onsite water	Onsite water
Onsite land	Landfills, impoundment ponds, underground wells
Onsite other	Waste piles, leaks, spills
Offsite water	Public/private water treatment
Offsite land	Landfills, impoundment ponds, underground wells
Offsite other	Residual emissions, waste brokers, incinerators and storage facilities
Recycled or treated	Recycled, recovered, treated

Table 2.4: Monitor distance and emissions growth rates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Distance	Distance	Distance	Distance	Distance	Distance	Distance	Distance
D.Onsite air	0.185							
	(0.131)							
D.Onsite water		0.0303						
		(0.219)						
D.Onsite land			0.296					
			(1.022)					
D.Onsite other				-1.284**				
				(0.575)				
D.Offsite water					-0.0165			
					(0.0800)			
D.Offsite land						0.0742		
						(0.0855)		
D.Offsite other							0.123	
							(0.0852)	
D.Recycled or treated								-0.0146
								(0.0794)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7824	1384	391	288	2578	2619	1899	3836

Standard errors in parentheses

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

Estimates from a regression of distance to the nearest non-attainment monitor (in km) on changes in log emissions. Sample is untreated plant-years. SEs clustered at the county level, which is the level of exogenous variation. Observation counts differ across columns because not all plants report emissions into all media. “Onsite other” emissions include waste piles, leaks, and spills.

Table 2.6: Effect on air emissions

	(1)	(2)	(3)
	Onsite air	Onsite air	Onsite air
Treated	-0.225**	-0.192**	
	(0.0968)	(0.0964)	
Tau=-3			-0.174
			(0.141)
Tau=-2			-0.0778
			(0.116)
Tau=-1			-0.208
			(0.138)
Tau=0 (1st treated year)			-0.348**
			(0.158)
Tau=1			-0.454**
			(0.182)
Tau=2			-0.350*
			(0.183)
Tau>=3			-0.384**
			(0.153)
State linear trends	No	Yes	No
Year dummies	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes
Observations	97621	97621	97621

Standard errors in parentheses

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

Estimates in columns 1-2 correspond to equation 2.4, while estimates in column 3 correspond to equation 2.5. Dependent variable is log air emissions (lbs). SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year.

Table 2.7: Effect on emissions ratios

Panel A: Main specification							
	Onsite water	Onsite land	Onsite other	Offsite water	Offsite land	Offsite other	Recycled or treated
Treated	0.436**	-0.363	0.889	0.257*	0.101	0.271	0.412***
	(0.186)	(0.309)	(0.562)	(0.152)	(0.146)	(0.282)	(0.128)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24458	12088	5233	28260	40286	24026	49007
Panel B: State linear trends							
	Onsite water	Onsite land	Onsite other	Offsite water	Offsite land	Offsite other	Recycled or treated
Treated	0.487***	-0.380	1.141	0.259*	0.176	0.261	0.447***
	(0.179)	(0.345)	(0.767)	(0.141)	(0.149)	(0.275)	(0.139)
State linear trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24458	12088	5233	28260	40286	24026	49007

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

Estimates correspond to equation 2.6. Dependent variable is log emissions ratio (lbs), with the numerator indicated atop the column and the denominator air emissions in all columns. Specification includes year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Observation counts differ across columns because not all plants report emissions into all media. “Onsite other” emissions include waste piles, leaks, and spills.

Table 2.9: Effect on emissions ratios, by 2-digit NAICS code

	Onsite air	Onsite water	Onsite land	Onsite other	Offsite water	Offsite land	Offsite other	Recycled or treated
Primary metals	-0.0776	0.334	-0.560	1.880***	0.340*	-0.0649	0.245	0.219
	(0.129)	(0.232)	(0.408)	(0.384)	(0.183)	(0.179)	(0.386)	(0.147)
Observations	51678	11228	2090	1545	20900	22943	14202	36568
Wood products	-0.470***	0.585	0.0615		0.203	0.157	0.302	0.983**
	(0.158)	(0.358)	(0.234)		(0.248)	(0.267)	(0.392)	(0.411)
Observations	33612	8668	5447	1531	5906	12851	7003	9538
Utilities	-0.356	1.180	-0.998	-0.159		1.742*	2.036	11.10***
	(0.438)	(0.890)	(1.139)	(0.591)		(0.975)	(1.445)	(0.287)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5399	3207	3317	838	368	2445	1451	1188

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

Includes the three 2-digit NAICS industries with the largest treated sample sizes. Column 1 (onsite air) corresponds to equation 2.4, remaining columns to equation 2.6. Dependent variable is log air emissions (lbs) in column 1, otherwise log emissions ratio (lbs), with the numerator indicated atop the column and the denominator air emissions in all columns. All specifications include year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Observation counts differ across columns because not all plants report emissions into all media. “Onsite other” emissions include waste piles, leaks, and spills.

Table 2.11: Leakage effect, within firm & 6-digit NAICS code

	(1)	(2)	(3)	(4)
	Onsite air	Onsite air	Onsite air	Onsite air
1+ other treated plants	0.169**	0.155**		
	(0.0718)	(0.0700)		
Count other treated			0.120**	0.107*
			(0.0594)	(0.0579)
State linear trends	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Multiplant dummy	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes
Observations	90688	90688	90688	90688

Standard errors in parentheses

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

Estimates correspond to equation 2.7, where "other treated plant" is a treated plant within the same firm and 6-digit NAICS code. Dependent variable is log air emissions (lbs). Specification includes year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Sample restricted to plants in attainment counties. Parent firm identifiers come from TRI data.

Table 2.12: Effect on emissions levels

	Onsite water	Onsite land	Onsite other	Offsite water	Offsite land	Offsite other	Recycled or treated
Treated	0.234	-0.367	0.485	0.0619	-0.0900	0.174	0.184*
	(0.202)	(0.277)	(0.493)	(0.135)	(0.163)	(0.242)	(0.0954)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24458	12088	5233	28260	40286	24026	49007

Standard errors in parentheses

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

Dependent variable is log emissions (lbs), with the medium indicated atop the column. All specifications include year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Observation counts differ across columns because not all plants report emissions into all media. “Onsite other” emissions include waste piles, leaks, and spills.

Table 2.13: Placebo effect on emissions levels

	Onsite water	Onsite land	Onsite other	Offsite water	Offsite land	Offsite other	Recycled or treated
Treated*no air emissions	-0.647	-0.761***	0.171	0.243	0.0984	-0.774	-0.127
	(0.454)	(0.0474)	(0.139)	(0.376)	(0.294)	(0.581)	(0.162)
Treated*air emissions	0.273	0.620	0.877***	-0.0544	0.0724	0.144	0.172*
	(0.192)	(0.666)	(0.323)	(0.185)	(0.154)	(0.200)	(0.0899)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	34544	16393	7307	51753	69004	39451	85433

Standard errors in parentheses

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

Estimates correspond to equation 2.6, but with 2 changes: 1) the dependent variable is log emissions in pounds (not a ratio); and 2) estimates for “Treated*no air emissions” report the effect of placebo treatment (being near a non-attainment monitor) on plants with no air emissions, which should not be affected by the CAA. Estimates for “Treated*air emissions” are for actually treated plants; they are not placebos. The medium is indicated atop the column. All specifications include year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Observation counts differ across columns because not all plants report emissions into all media.

Table 2.14: Placebo leakage effect, within firm & 6-digit NAICS code

	(1)	(2)	(3)	(4)
	Onsite air	Onsite air	Onsite air	Onsite air
1+ other placebo plants	0.0293	0.0245		
	(0.0443)	(0.0446)		
Count placebo plants			0.0142	0.0115
			(0.0234)	(0.0239)
State linear trends	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Multiplant dummy	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes
Observations	99580	99580	99580	99580

Standard errors in parentheses

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

Estimates correspond to equation 2.7, but using variables based on placebo treated plants: plants within the same firm and 6-digit NAICS code, located in non-attainment counties, but farther than 8km from the nearest non-attainment monitor. Dependent variable is log air emissions (lbs). Specification includes year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Sample restricted to plants in attainment counties. Parent firm identifiers come from TRI data.

2.10 Supplementary material

2.10.1 Alternative theoretical models

Modeling the CAA as a quantity restriction

Suppose two pollution inputs: A ~ air emissions, W ~ water emissions. Treat the CAA as an exogenous quantity restriction $\overline{A}$ on air emissions. The object of policy interest is unconditional factor demand W^*, incorporating firms' possible output response to regulation. Suppose a CES production function, so the firm problem becomes:

$$\max_{A,W} p_o \left(c_A A^\rho + c_W W^\rho\right)^{1/\rho} - p_A A - p_W W + \lambda \left[\overline{A} - A\right]$$

Taking FOCs, one obtains an optimality condition:

$$\left(\frac{c_W}{c_A}\right)\left(\frac{W^{*\rho-1}}{A^{*\rho-1}}\right) = \frac{p_W}{p_A + \lambda}$$

If the constraint does not bind prior to CAA non-attainment, the shadow price λ is zero. Taking logs gives ratio of unconditional factor demands:

$$\ln\left(\frac{W^*}{A^*}\right) = \frac{1}{1-\rho}\ln\left(\frac{c_W}{c_A}\right) + \frac{1}{1-\rho}\ln\left(\frac{p_A + 0}{p_W}\right) \tag{2.10}$$

Treat CAA non-attainment as a decrease in $\overline{A}$ such that it binds. This changes the value of λ from zero to an unknown positive number. The optimality condition then becomes:

$$\ln\left(\frac{W^*}{\overline{A}}\right) = \frac{1}{1-\rho}\ln\left(\frac{c_W}{c_A}\right) + \frac{1}{1-\rho}\ln\left(\frac{p_A + \lambda}{p_W}\right) \tag{2.11}$$

If ρ is finite and $\rho \leq 1$, then the coefficient on the last term is positive. The positive shadow price λ causes an increase in the last term. Theory then predicts an increase in the ratio of water to air pollution $\frac{W^*}{\overline{A}}$. This prediction is the same as the one from the model treating CAA non-attainment as a relative price change. The difference is that under this model, a regression that fails to control for output will not produce biased estimates if $\overline{A}$ is truly exogenous. Rearranging equation 2.11 yields:

$$\ln\left(W^*\right) = \frac{1}{1-\rho}\ln\left(\frac{c_W}{c_A}\right) + \frac{1}{1-\rho}\ln\left(\frac{p_A+\lambda}{p_W}\right) + \ln\left(\overline{A}\right) \tag{2.12}$$

If regulators consider plant characteristics when deciding on the constraint $\overline{A}$, however, the potential for bias in a non-ratio specification returns.

Three production inputs

Suppose a nested CES production function, including a third input L. As in Fullerton and Karney (2014), this input may be regarded as labor or as a composite of non-pollution inputs like labor, land and capital. The firm problem then becomes:

$$\max_{A,W,L} p_O c_2 \left\{ c_P \left[c_1 \left(c_A A^\rho + c_W W^\rho \right)^{1/\rho} \right]^\theta + c_L L^\theta \right\}^{1/\theta} - p_A A - p_w W$$

The constants c_1, c_2, c_A, c_W, c_P and c_L reflect a firm's technology. Taking first order conditions on A and W, then dividing, yields:

$$\frac{p_O}{p_O}\frac{c_2}{c_2}\frac{\{\cdot\}^{1/\theta-1}}{\{\cdot\}^{1/\theta-1}}\frac{c_P}{c_P}\frac{[\cdot]^{\theta-1}}{[\cdot]^{\theta-1}}\frac{c_1}{c_1}\frac{(\cdot)^{1/\rho-1}}{(\cdot)^{1/\rho-1}}\frac{c_W}{c_A}\frac{W^{*\rho-1}}{A^{*\rho-1}} = \frac{p_W}{p_A}$$

This produces the same optimality condition derived in Section 2.3.

$$\left(\frac{c_W}{c_A}\right)\left(\frac{W^{*\rho-1}}{A^{*\rho-1}}\right) = \frac{p_W}{p_A}$$

Intuitively, this is because the firm substitutes over the air-labor and water-labor input pairs in the same way, so changes in the third factor do not affect the ratio of A and W. Under this functional form assumption, the omission of other inputs from my ratio regression specifications will not prevent inference of properties of the parameter $\sigma = \frac{1}{1-\rho}$. Nested CES is not the only functional form with this property, but it illustrates the character of the required assumptions in a three-input case.

Now let us consider a three-input case where production is not nested CES, but remains CRS. My estimates can no longer be interpreted as the Hicks elasticity σ. Instead they will capture the Morishima elasticity of substitution with respect to price p_A (Blackorby and Russell, 1989):

$$M_{AW}\left(Y, p_A, p_W\right) = \varepsilon_{WA} - \varepsilon_{AA}$$

where ε_{WA} and ε_{AA} are cross- and own-price elasticities of factor demand. (While this is the natural generalization of the Hicks elasticity, its asymmetry makes it different in one important respect: the elasticity M_{AW} is informative for changes in $p_A$but not for changes in p_W.) Unlike the sign of σ, the sign of M_{AW} is ambiguous because the sign of ε_{WA} is unknown when there are three or more inputs. If in fact production is non-CES in three or more inputs, my estimates allow the possibility of a negative Morishima elasticity. Note that adding controls for the levels of additional inputs (beyond A and W) would force the tradeoff back into the $A-W$ plane. As Blackorby and Russell (1989) argue, this measure of curvature is interesting but substantially less informative than the Morishima elasticity. In this case the omission of other inputs from the right hand side of my regression equations does not create confounding problems, but rather allows recovery of the Morishima elasticity. The identifying assumption is not that the change in p_A has no effect on other inputs, but rather that only p_A changes and other prices remain constant. If the plants under study are price takers in factor markets and CAA non-attainment does not produce general-equilibrium effects on other factor prices, then this assumption likely holds.

2.10.2 Additional tables

Descriptive tables

Table 2.15: TRI *PM* descriptive statistics, by attainment status

	Attainment counties		Nonattainment counties	
	Mean	Stdev	Mean	Stdev
Onsite air	6757.32	498535.05	1927.42	39160.96
Onsite water	1032.32	13219.98	465.88	9128.17
Onsite land	62328.93	1328520.39	55508.89	687992.66
Offsite other	53327.01	2576281.40	74478.75	1847581.34
Offsite water	476.49	24594.32	1104.67	49007.71
Offsite land	15236.25	155807.26	21324.12	178990.85
Offsite other	5360.06	75463.33	5673.53	78149.45
Recycled or treated	88889.85	740562.92	99512.66	965368.21
Dist. to nonattain monitor (km)	0.00	0.00	11.78	12.27
Treated	0.00	0.02	0.09	0.29
Observations	99580		20529	

Emissions measured in pounds. Unit of observation is a plant-year.

Table 2.16: TRI *PM* descriptive statistics, by leakage dummy

	Other plants		Leakage plants	
	Mean	Stdev	Mean	Stdev
Onsite air	7542.00	532524.13	1370.39	5731.27
Onsite water	1127.86	13903.57	582.96	11175.12
Onsite land	65172.75	1357515.96	12245.54	119960.86
Offsite other	51794.02	2440025.55	70.45	1783.87
Offsite water	523.38	26249.00	363.11	5354.29
Offsite land	15724.30	152502.16	27039.82	308992.27
Offsite other	5626.86	78133.06	8745.91	93685.60
Recycled or treated	95116.95	783390.88	100342.31	446333.85
Observations	87263		3425	

Emissions measured in pounds. Unit of observation is a plant-year. “Other plants” are plants in attainment counties that have no treated plants within the same firm-year. “Leakage plants” are plants in attainment counties that have at least one treated plant within the same firm-year.

Table 2.17: Historical CAA particulate standards

Final rule	Type	Averaging time	Standard (g/m3)	Form
1987	PM_{10}	24hr	150	Not to be exceeded more than once per year on average over a 3-year period
		Annual	50	Annual arithmetic mean, averaged over 3 years
1997	$PM_{2.5}$	24hr	65	98th percentile, averaged over 3 years
		Annual	15	Annual arithmetic mean, averaged over 3 years
	PM_{10}	24hr	150	Not to be exceeded more than once per year on average over a 3-year period
		Annual	50	Annual arithmetic mean, averaged over 3 years
2006	$PM_{2.5}$	24hr	35	98th percentile, averaged over 3 years
		Annual	15	Annual arithmetic mean, averaged over 3 years
	PM_{10}	24hr	150	Not to be exceeded more than once per year on average over a 3-year period

Adapted from http://www.epa.gov/ttn/naaqs/standards/pm/s_pm_history.html. Accessed March 19, 2014.

Monitor distance

Table 2.18: Monitor distance and emissions levels

	Distance	Distance	Distance	Distance	Distance	Distance	Distance	Distance
Onsite air	0.0180							
	(0.0867)							
Onsite water		0.204						
		(0.201)						
Onsite land			1.088***					
			(0.383)					
Onsite other				0.310				
				(0.386)				
Offsite water					-0.278***			
					(0.0773)			
Offsite land						0.0584		
						(0.0798)		
Offsite other							-0.152*	
							(0.0889)	
Recycled or treated								-0.124
								(0.0874)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8227	1519	451	349	2843	2988	2297	4308

Standard errors in parentheses

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

Estimates from a regression of distance to the nearest non-attainment monitor (in km) on log emissions. Sample is untreated plant-years. SEs clustered at the county level, which is the level of exogenous variation. Observation counts differ across columns because not all plants report emissions into all media. "Onsite other" emissions include waste piles, leaks, and spills.

Table 2.20: General-equilibrium spillover test

	(1) Onsite air	(2) Onsite air
Num. treated plants (state)	-0.00140	
	(0.00148)	
Num. treated plants (state and NAICS6)		0.0481*
		(0.0269)
Year dummies	Yes	Yes
Plant FEs	Yes	Yes
Observations	96529	96529

Standard errors in parentheses

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

Estimate corresponds to equation 2.8. Dependent variable is log air emissions (lbs). SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Sample restricted to plants in attainment counties. “Num. treated plants (state)” is the number of treated plants in a given state-year. “Num. treated plants (state and NAICS6)” is the number of treated plants in a given state, year, and six-digit NAICS code.

Air emissions

Table 2.21: Effect on air emissions, intrafirm spillover controls

	(1)	(2)
	Onsite air	Onsite air
Treated	-0.243**	-0.206**
	(0.0987)	(0.0977)
Spillover controls	Yes	Yes
State linear trends	No	Yes
Year dummies	Yes	Yes
Plant FEs	Yes	Yes
Observations	102039	102039

Standard errors in parentheses

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

Estimates in columns 1-2 correspond to equation 2.4, while estimates in column 3 correspond to equation 2.5, but with the inclusion of spillover controls from equation 2.7: a multiplant firm dummy and the interaction of that dummy with the number of treated plants within the same firm. Dependent variable is log air emissions (lbs). SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year.

Table 2.22: Effect on toxicity-weighted air emissions

	(1)	(2)
	Onsite air	Onsite air
Treated	-0.515**	-0.413*
	(0.240)	(0.236)
State linear trends	No	Yes
Year dummies	Yes	Yes
Plant FEs	Yes	Yes
Observations	62119	62119

Standard errors in parentheses

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

Estimates in columns 1-2 correspond to equation 2.4, while estimates in column 3 correspond to equation 2.5. Dependent variable is log toxicity-weighted air emissions (unitless). SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. EPA inhalation toxicity weights applied to air emissions.

Table 2.23: Effect on air emissions, plants open 1993-2010

	(1)	(2)
	Onsite air	Onsite air
Treated	-0.259	-0.228
	(0.158)	(0.164)
State linear trends	No	Yes
Year dummies	Yes	Yes
Plant FEs	Yes	Yes
Observations	21864	21864

Standard errors in parentheses

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

Estimates correspond to equation 2.4. Dependent variable is log air emissions (lbs). SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year.

Table 2.24: Effect on air emissions, difference specification

	(1)	(2)
	D.Onsite air	D.Onsite air
Treated	-0.0341	-0.0329
	(0.0370)	(0.0383)
State linear trends	No	Yes
Year dummies	Yes	Yes
Plant FEs	Yes	Yes
Observations	93443	93443

Standard errors in parentheses

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

Estimates correspond to equation 2.4, but with the dependent variable replaced by the year-on-year difference in logs (the growth rate). This specification parallels those used by Greenstone (2003) and Gamper-Rabindran (2009). SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year.

Cross-media substitution

Table 2.25: Effect on emissions ratios, by 3-digit NAICS code

	Onsite air	Onsite water	Onsite land	Onsite other	Offsite water	Offsite land	Offsite other	Recycled
Primary metals	0.116	0.193	0.194	2.052**	-0.211	-0.0477	0.114	0.0358
	(0.179)	(0.282)	(0.848)	(0.854)	(0.454)	(0.274)	(0.431)	(0.267)
Chemicals	-0.347*	-0.102	4.030	2.826***	0.0284	0.0294	0.146	1.324**
	(0.196)	(0.268)	(3.790)	(0.629)	(0.227)	(0.292)	(0.458)	(0.631)
Fabricated metals	-0.246	0.467			0.338	0.436	-0.444	0.270
	(0.335)	(0.634)			(0.309)	(0.400)	(0.410)	(0.257)
Nonmetallic mineral products	-0.677*		1.939***		4.552***	1.034**	0.611	-0.0928
	(0.350)		(0.488)		(0.246)	(0.449)	(0.656)	(0.229)
Transportation equipment	-0.891**	1.977	-0.493		0.647*	-0.114	2.767**	0.948
	(0.369)	(1.972)	(0.805)		(0.333)	(0.543)	(1.405)	(0.800)
Petroleum and coal	-1.173**	2.506***			6.434**	1.396**	-2.001	0.501
	(0.539)	(0.483)			(3.079)	(0.599)	(1.905)	(0.765)
Utilities	-0.606	0.875	-0.999	-0.132		2.193**	2.061	11.16***
	(0.443)	(0.853)	(1.139)	(0.590)		(1.076)	(1.448)	(0.289)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

Includes the seven 3-digit NAICS industries with the largest treated sample sizes. Column 1 (onsite air) corresponds to equation 2.4, remaining columns to equation 2.6. Dependent variable is log air emissions (lbs) in column 1, otherwise log emissions ratio (lbs), with the numerator indicated atop the column and the denominator air emissions in all columns. All specifications include year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Observation counts differ across columns because not all plants report emissions into all media. “Onsite other” emissions include waste piles, leaks, and spills.

Table 2.26: Intra-firm leakage effect on emissions ratios, within firm & 6-digit NAICS code

	Onsite water	Onsite land	Onsite other	Offsite water	Offsite land	Offsite other	Recycled or treated
1+ other treated plants	-0.0357	-0.0543	-0.0778	0.0877	0.295*	-0.655***	0.00585
	(0.195)	(0.213)	(0.261)	(0.163)	(0.159)	(0.243)	(0.143)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23174	11729	4872	25634	37787	21502	45138

Standard errors in parentheses

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

Estimates correspond to equation 2.7, where “other treated plant” is a treated plant within the same firm and 6-digit NAICS code, but dependent variable is log emissions ratio (lbs). Numerator indicated atop column and denominator is air emissions in all columns. Specification includes year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Observation counts differ across columns because not all plants report emissions into all media. “Onsite other” emissions include waste piles, leaks, and spills. Sample restricted to plants in attainment counties. Parent firm identifiers come from TRI data.

Table 2.27: Effect on toxicity-weighted emissions ratios

Panel A: Main specification							
	Onsite water	Onsite land	Onsite other	Offsite water	Offsite land	Offsite other	Recycled or treated
Treated	0.151	0.0613	2.075***	0.174	0.174	-0.00606	0.797**
	(0.239)	(0.181)	(0.305)	(0.451)	(0.353)	(0.813)	(0.368)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10668	3792	1967	15739	20928	10950	32862

Panel B: State linear trends							
	Onsite water	Onsite land	Onsite other	Offsite water	Offsite land	Offsite other	Recycled or treated
Treated	0.150	-0.0102	2.142***	0.215	0.236	-0.144	0.728**
	(0.284)	(0.187)	(0.371)	(0.406)	(0.333)	(0.832)	(0.367)
State linear trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10668	3792	1967	15739	20928	10950	32862

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

Estimates correspond to equation 2.6. Dependent variable is log toxicity-weighted emissions ratio (unitless), with the numerator indicated atop the column and the denominator air emissions in all columns. Specification includes year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Observation counts differ across columns because not all plants report emissions into all media. "Onsite other" emissions include waste piles, leaks, and spills. EPA inhalation toxicity weights applied to air emissions and ingestion weights applied to all other emissions.

Table 2.28: Aggregate effect on emissions ratios

	(1)	(2)	(3)	(4)
	Other media	Other media	Other media	Other media
Treated	0.271**	0.284**	0.144	0.108
	(0.121)	(0.121)	(0.122)	(0.131)
Log air emissions			0.217***	
			(0.0105)	
State linear trends	No	Yes	No	No
Year dummies	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes
Observations	83218	83218	83218	83218

Standard errors in parentheses

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

Estimates correspond to equation 2.6. In columns 1-2, dependent variable is log emissions ratio (lbs), with the numerator aggregate emissions into all non-air media and the denominator air emissions. In columns 3-4 dependent variable is log emissions into all non-air media. All specifications include year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year.

Leakage

Table 2.29: Leakage effect, within firm & 2-digit NAICS code

	(1)	(2)	(3)	(4)
	Onsite air	Onsite air	Onsite air	Onsite air
1+ other treated plants	0.118**	0.104**		
	(0.0477)	(0.0463)		
Count other treated			0.0644*	0.0540
			(0.0375)	(0.0364)
State linear trends	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Multiplant dummy	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes
Observations	90688	90688	90688	90688

Standard errors in parentheses

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

Estimates correspond to equation 2.7, where "other treated plant" is a treated plant within the same firm and 2-digit NAICS code. Dependent variable is log air emissions (lbs). Specification includes year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Sample restricted to plants in attainment counties. Parent firm identifiers come from TRI data.

Table 2.30: Leakage effect, continuous firm size controls

	(1)	(2)	(3)	(4)
	Onsite air	Onsite air	Onsite air	Onsite air
1+ other treated plants	0.171^{**}		0.156^{**}	
	(0.0714)		(0.0714)	
Count other treated		0.121^{**}		0.108^{*}
		(0.0592)		(0.0593)
Plants in firm	Yes	Yes	No	No
Plants in firm and NAICS	No	No	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Multiplant dummy	Yes	Yes	Yes	Yes
Plant FEs	Yes	Yes	Yes	Yes
Observations	90688	90688	90688	90688

Standard errors in parentheses

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

Estimates correspond to equation 2.7, where "other treated plant" is a treated plant within the same firm and 6-digit NAICS code. "Plants in firm" is a count of all plants in a given firm-year. "Plants in firm and NAICS" is a count of plants within firm-year and 6-digit NAICS code. Dependent variable is log air emissions (lbs). Specification includes year dummies and plant fixed effects. SEs clustered at the county level, which is the level of exogenous variation. Unit of observation is a plant-year. Sample restricted to plants in attainment counties. Parent firm identifiers come from TRI data.

Chapter 3

Time Use and Productivity: The Wage Returns to Sleep

Abstract

While economists have long been interested in effects of health and human capital on productivity, less attention has been paid to the influence of time use. We investigate the productivity effects of the single largest use of time—sleep. Because sleep influences performance on memory and focus intensive tasks, it plausibly affects economic outcomes. We identify the effect of sleep on wages by exploiting the relationship between sunset time and sleep duration. Using a large, nationally representative set of time use diaries from the United States, we provide the first causal estimates of the impact of sleep on wages. A one-hour increase in seasonal weekly sleep increases a worker's wage by 1%. At the location level, a one-hour increase in long-run weekly mean sleep increases mean wage by 4.5%. Our results highlight the economic importance of sleep and pose potentially fruitful questions about the effects of time use on labor market outcomes.

3.1 Introduction

Economists have long been interested in determinants of productivity. Understanding what makes workers more effective is a fundamental question in economics, important for both individual decisions and public policy. While there are traditions of research in human capital (Becker, 1962; Becker, 1964) and health (Grossman, 1972), less attention as been paid to the influence of time use on worker productivity. Many types of time use, from reading to vacationing, plausibly impact productivity on the job. In this study we focus on the time a worker spends sleeping, potentially one of the most important influences on human performance.

Evidence from medical research points to the vital role sleep plays in determining productivity. Tired doctors make more mistakes (Ulmer et al., 2009). Tired students perform worse on tests (Taras and Potts-Datema, 2005). Poor sleep raises total mortality rates (Cappuccio et al., 2010). These results suggest that inadequate sleep lowers productivity, impedes the development of human capital, and imposes large direct costs on society. Moreover for the average individual, sleep takes up more time than any other activity. Despite the manifest importance of sleep, economists have largely treated it as a biological phenomenon outside their purview. We investigate an important question that has been overlooked almost entirely in economics: what are the effects of sleep on wages?

Answering this question poses formidable challenges. First, a pioneering study by Biddle and Hamermesh (1990) shows that higher wages raise the opportunity cost of sleep time, leading individuals to decrease their sleep. This result demonstrates that causal relationships between sleep and wages could run in both directions. Additionally, sleep may be correlated with unobservable worker characteristics, like ability, that also influence wages. Finally, because sleep is a large portion of the time budget and complementary to almost all human activity, it is extremely difficult to isolate exogenous variation in sleep.

Motivated by medical research on circadian rhythm, we resolve this endogeneity using changes in sleep induced by differences in sunset time. Earlier sunset causes workers to begin sleeping earlier. Because work and school start times do not respond to solar cues (Hamermesh et al., 2008), this earlier bed time translates into more sleep. We exploit this mechanism to identify the wage effects of both short and long-run sleep changes. First, we use seasonal variation in sunset time within a location to identify the wage effect of short-run sleep changes: earlier sunset in winter induces longer sleep duration. Second, we use long-run variation in average sunset time across locations, created by time zones, to identify the wage effects of long-run sleep changes. Such timing differences stem from US time zone boundaries drawn in 1883, which stem in turn from the historical accident that placed the Prime Meridian through Greenwich, England. For two locations at the same latitude and in the same time zone, the location farther east will experience sunset sooner than the location farther west. Residents of the eastern location will go to bed earlier and sleep longer on average.

To implement our empirical strategy, we geocode observations from the American Time Use Survey (ATUS), at the county or CBSA level where possible and at the state level otherwise. ATUS provides rich labor market information about individuals, a wealth of control variables, and detailed time use data from daily diaries. Using the diary date and location, we assign each observation a diary-date sunset time and an annual average sunset time. We then use these sunset time instruments to estimate the short and long-run causal effects of sleep on wages.

Our results show that a one-hour increase in average weekly sleep, on a seasonal time scale, increases a worker's wage by 1%. At the location level, a one-hour increase in long-run weekly average sleep increases average wage by 4.5%. Both estimates reflect exogenous sleep changes for all workers in a location, so they potentially include productivity spillovers across workers. In addition, our long-run estimate may include general-equilibrium effects induced by exogenously higher worker productivity. These are, to our knowledge, the first causal estimates of how sleep affects wages.[1]

Our long-run identification strategy naturally raises concerns that workers might sort on sunset time or on its correlates. As part of our large set of robustness checks, we demonstrate that our long-run wage effects are offset by increased home prices, removing the incentive to sort. In addition, we conduct a variety of tests for worker sorting and find no evidence for it.

Our study demonstrates that sleep is not just an economic curiosity, but rather a vital determinant of productivity. A one-hour increase in a location's weekly mean sleep raises wages by roughly half as much as a one-year increase in education (Psacharopoulos and Patrinos, 2004). These results point to the large impact that non-labor market activities can have on labor market performance. They suggest governments and schools must account for the productivity impacts of sleep to design optimal scheduling and time-use policies. By examining the largest use of human time, our study contributes to the time-use literature following Becker (1965). We also contribute to the growing literature on how environmental forces influence worker productivity (Zivin and Neidell, 2012) and to the broader productivity literature on factors like information technology (Bloom et al., 2012) and workplace practices (Black and Lynch, 2001). Future work should extend these results to compare them to non-time intensive changes in leisure or lifestyle attributes.

The rest of the paper proceeds as follows. Section 3.2 presents a time use model with sleep as a choice variable, illustrating challenges associated with identifying the effect of sleep on wages, and discusses related literature. Section 3.3 presents the

[1]Biddle and Hamermesh (1990) includes a regression with wages on the left-hand side and sleep on the right and finds a negative relationship. This is consistent with reverse causality and highlights the difficulty of isolating quasi-experimental variation in sleep.

estimating equations and discusses our identification strategy. Section 3.4 describes the data used in the study. Section 3.5 reports and discusses our short-run results, provides robustness checks, and discusses extensions to the main results, while Section 3.6 does the same for our long-run results. Section 3.7 concludes.

3.2 Identifying the effect of sleep on productivity and wages

3.2.1 Previous research

Existing studies of the relationship between sleep and wages in economics are few and are largely concerned with addressing the question of whether sleep should be treated as a choice variable rather than simply a biological necessity. Biddle and Hamermesh (1990) is the first paper to provide empirical evidence on this issue and remains one of the only empirical investigations of labor market impacts of sleep. The authors lay out a model with agents optimizing over sleep, work, and leisure time in an otherwise standard setting. While their theoretical model allows sleep to affect productivity, Biddle and Hamermesh do not focus on this relationship in their empirical work. Instead they emphasize the causal mechanism operating in the opposite direction, modeling sleep as a function of instrumented wage (see Biddle and Hamermesh (1990) Table 6). Brochu et al. (2012) and Szalontai (2006) also estimate the impact of changes in wage on sleep using more recent data from Canada and South Africa. Finally, Bonke (2012) has examined the impact of two chronotypes—whether the individual is a “morning” or “evening” person—on income. This study provides evidence on the related question of whether sleep quality impacts labor market outcomes.

Daylight savings time (DST) has been used in a variety of settings in economics as a proxy for sleep changes. (See for example Smith (2014).) However, the short-term nature of any sleep change induced by DST limits its use in studying slow-moving outcomes like wages. Moreover, examination of ATUS data shows that the relationship between DST and sleep is complex. Transition into DST reduces sleep by 40 minutes on the day of the change, but transition out of DST is not associated with a noticeable change in sleep time (Barnes and Wagner, 2009).

Medical studies concerned with the effect of long term differences in sleep on health or mortality[2] are closest to our study in terms of time horizon. A series of papers starting with Mckenna et al. (2007) have used laboratory tasks to examine the impact of short-term sleep loss on a variety of outcomes that provide insight into how sleep could impact work performance. Van Dongen et al. (2003) conducted the longest laboratory-controlled study on the effect of sleep levels on cognitive performance. The researchers kept subjects in the lab for two weeks, placing them into groups receiving 4, 6, and 8 hours of sleep. The subjects were given daily tests of attention, memory, and cognition. The research found that the groups subjected to 4 and 6 hours of sleep performed progressively worse on all three tests, relative to the 8-hour group. Intriguingly, the subjects' subjective assessments followed a different pattern, declining for a few days and then leveling off. Observed cumulative effects quickly achieved large magnitudes: after one week, subjects in the 6-hour group performed as badly as subjects who were deprived of sleep entirely for one night. This indicates that small sleep reductions over long periods of time can have very large effects. Van Dongen et al. (2003) provides one of the most compelling pieces of evidence for the negative productivity effects of reduced sleep. Appendix Table 3.12 expresses the results of Van Dongen et al. (2003) and similar medium-term causal studies as elasticities. These studies all manipulated sleep duration by relatively modest amounts, from one to four hours per night, over periods of one to

[2]For instance Cappuccio et al. (2010) and Krueger and Friedman (2009).

three weeks. In almost every case they find very large effects. The typical elasticity of task performance with respect to sleep duration is approximately four.

3.2.2 A productive sleep model

The following analytical model, adapted from Biddle and Hamermesh (1990), illustrates the trade-offs between consumption, leisure, and sleep when sleep affects wages. It also demonstrates the reverse causality from wages to sleep that creates one of the main identification challenges and clarifies how we think about our instrument. Consider a consumer optimizing over sleep time T_S and a composite leisure good Z, which requires inputs of both time T_z and goods X such that $T_z = bZ$ and $X = aZ$. The good X trades at the exogenous price P. The consumer has non-labor income I and time endowment T^*. Denote work time T_w. Let an individual's market wage w_m depend on sleep as follows: $w_m = w_1 + f(T_S)$, with $w_1 > 0$, $f'(T_S) > 0$, and $f''(T_S) < 0$.

Note that this theoretical model could easily be adapted to study other non-work time uses, but the function linking wage to time use would likely be different. We assume that a function of sleep, αT_S, enters the utility function, where α is the relative utility enjoyed by the individual per hour of sleep.[3] The parameter α provides a convenient link between our analytical model and our instrumental variables estimation strategy, as discussed below. The worker optimizes over sleep and composite leisure, subject to time and income constraints, as follows.

$$\max_{Z,T_S,\lambda} U(Z,\alpha T_S) + \lambda \left(I + (w_1 + f(T_S))(T^* - T_S - bZ) - aPZ\right)$$

Combining first-order conditions yields a two by two system of equations that implicitly

[3]Our predictions are qualitatively unchanged if we assume that sleep does not enter the utility function directly, but rather as an input to the production of the composite leisure good Z as in HEREHERE.

describe the worker's optimal choice.

$$U_1 w_m - U_1 f'(T_S)\, T_w - \alpha U_2 (aP + bw_m) = 0$$

and

$$I + (w_1 + f(T_S))(T^* - T_S - bZ) - aPZ = 0$$

Applying the implicit function theorem, we can evaluate several interesting derivatives. First, consider the effect of an exogenous wage increase on sleep time.

$$\frac{\partial T_S}{\partial w_1} = (aP + bw_m)(U_1 - \alpha U_2 b)\, D^{-1} + T_w \frac{\partial T_S}{\partial I}$$

In the previous expression, $D^{-1} < 0$ equals the negative of the Jacobian. This is a variant of the usual Slutsky equation. The first term captures the substitution effect, which differs from the typical form in that it includes $-\alpha U_2 b$. When $\alpha = 1$ the value $(U_1 - \alpha U_2 b) > 0$ and the first term is negative. Increased wages raise the opportunity cost of sleep, decreasing optimal sleep. This means that a naïve regression of wages on sleep will not recover causal effects.

To motivate our later use of an instrument for sleep, consider the effects of an exogenous increase in α. Since α controls the relative attractiveness of sleep, an increase in the parameter will induce agents to want to consume more sleep.

$$\frac{\partial T_S}{\partial \alpha} = U_2 (aP + bw_m)^2 (-D)^{-1} > 0$$

The effect on leisure can operate in either direction.

$$\frac{\partial T_z}{\partial \alpha} = bU_2 (aP + bw_m) \left(f'(T_S - w_m)\, T_w \right) (-D)^{-1} \lesseqgtr 0$$

The ambiguous sign comes from the expression $(f'(T_S - w_m)T_w)$, which is the opportunity cost of an additional leisure hour. More specifically, this expression is the gross opportunity cost of an additional leisure hour, $-w_m$, adjusted for the additional income generated by increased sleep, $f'(T_S)T_w$ (recall that T_S increases in response to an increase in α). Individuals with low wages (low w_1), or a combination of high work hours and low sleep hours, will tend to decrease leisure time in response to decreased α. This is because the income effect dominates the substitution effect and income is a complement of leisure time. For low-wage workers, the substitution effect is small. For high-work, low-sleep workers, the income effect is large; they are in the steep portion of the sleep-wage function and any change in wage applies to many hours. We test these theoretical predictions in Section 3.6.4.

3.3 Empirical strategy

3.3.1 Estimating equations

Our goal is to estimate an equation of the form

$$\text{wage}_{it} = f(T_{S,it}) + \varepsilon_{it}$$

where we expect $\partial f / \partial T_S > 0$, at least for low T_S. Given the reverse causality between wages and sleep, however, we might erroneously find $\partial f / \partial T_S < 0$.[4] To avoid this problem and to account for the wide variety of other omitted variables that might co-vary with sleep and wages, we predict sleep using one of two instruments based on local sunset

[4]The general form is given in the model above, but we can also illustrate the issue with a simple two equation system that will prove useful below. Let the sleep-wage relationship be given by

$$\begin{aligned} w &= \alpha T_S + \varepsilon \\ T_S &= \beta w + \nu \end{aligned}$$

time, then use the instrumented values of sleep to estimate wage impacts.

To estimate short-run seasonal effects, we employ sunset time on the ATUS diary date as our instrument.

$$\begin{aligned} T_{S,ijt} &= \alpha \text{sunset}_{jt} + \gamma_j + \mathbf{x}'_{it}\delta_1 + \nu_{ijt} \\ \ln(\text{wage}_{ij\tau}) &= \beta T_{S,ijt} + \gamma_j + \mathbf{x}'_{it}\delta_2 + \varepsilon_{ij\tau} \end{aligned} \tag{3.1}$$

In the above equations $T_{S,ijt}$ is nighttime sleep for individual i in location j on date t, sunset_{jt} is the sunset time on that date in that location, γ_j is a location fixed effect, $\mathbf{x}_{it}$ is a vector of controls, and $\text{wage}_{i\tau}$ is a measure of wages or earnings at time τ. We distinguish between the time subscripts on wages, τ, and sleep, t, to highlight the fact that we treat sleep on date t as a consistent estimate of average sleep at time τ.[5] Our wage measure is the answer to a question about "usual weekly earnings" rather than wages on the day of the interview, so τ may be thought of as indexing the wage-setting period. We provide more discussion of the interpretation of this estimator in the presence of slow-moving wages, measurement error, and seasonality below. Only if some realized wages adjust seasonally or intra-seasonally do we expect a positive estimate of β. Controls include: race indicators; age; age squared; a full-time indicator; a gender indicator; indicators for holiday, day of week, and year; and detailed occupation code indicators.

To investigate long-run effects, we employ annual average sunset time as our

where ε and ν are random error terms, $E[\varepsilon\nu] = 0$, $E[T_S\varepsilon] = 0$, and $E[w\nu] = 0$. Then if $\beta < 0$ as is argued by the previous literature, the bias from OLS estimation can be signed as follows:

$$\hat{\alpha} = \alpha + \underbrace{\beta \frac{E[\varepsilon w]}{E[T_S^2]}}_{<0}$$

So $\hat{\alpha} < \alpha$. Naive OLS will tend to understate the effect of sleep on wages if this is the dominant source of bias.

[5] We also treat a worker's observable characteristics on date t as consistent estimates of observables at time τ. Since many such characteristics are fixed or vary extremely slowly (for example race, occupation, and education), we believe this assumption is benign.

instrument. Because this instrument does not vary across individuals within a location, we collapse the ATUS data to the location level. This highlights the long-run, group-level nature of our exogenous variation and maximizes first-stage power. Following the recommendation in Solon et al. (2013), we weight location-level observations using counts of the underlying individual ATUS observations to correct for heteroskedasticity. (Appendix Table 3.16 provides evidence of heteroskedasticity from a modified Breusch-Pagan test and Appendix Table 3.17 presents unweighted results.)

$$\begin{aligned} T_{S,j} &= \alpha \text{sunset}_j + \mathbf{x}_j' \delta_1 + \nu_j \\ \ln(\text{wage}_j) &= \beta T_{S,j} + \mathbf{x}_j' \delta_2 + \varepsilon_j \end{aligned} \tag{3.2}$$

In the above equations $T_{S,j}$ is average nighttime sleep in location j, sunset_j is the average sunset time in that location, $\mathbf{x}_j$ is a vector of controls, and wage_j is average wage in that location. Controls include: coastal distance, a 10-piece linear spline in latitude; share full time; median age; race shares; occupation shares; and a 5-piece linear spline in population density.

The ideal experiment aimed at our question would exogenously vary short- or long-run sleep at the individual level, then estimate the relationship between wages and average sleep over the wage-setting period. Our study departs from this ideal in one important respect: our instrument exogenously varies sleep at the location level, not the individual level. This is important for the interpretation of our estimates, as discussed in Section 3.6.

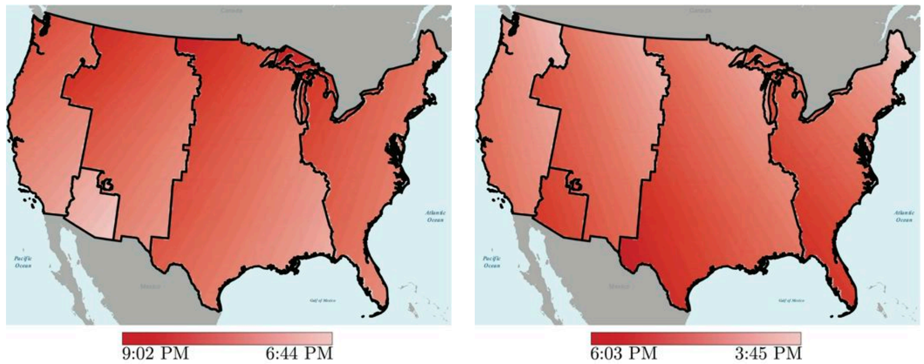
3.3.2 Local sunset time instruments

We would like to estimate the relationship between sleep and wages, but, as discussed above, sleep is plainly endogenous. To isolate exogenous variation in sleep, we instrument for sleep using two measures of local sunset time: sunset time on the day of the ATUS interview (short-run, seasonal variation) and average sunset time in a location (long-run variation). Instrument relevance flows from the same source in both cases. Human sleep patterns and circadian rhythm are synchronized with the rising and setting of the sun through a process known as entrainment. Roenneberg et al. (2007) show that "the human circadian clock is predominantly entrained by sun time rather than by social time." The authors demonstrate that earlier sunset induces workers to begin sleep earlier. The detailed ATUS files enable us reproduce this result: workers experiencing earlier sunset go to bed earlier and this causal connection between sunset and bedtime persists even if the worker goes to bed well after dark. In a vacuum, an earlier sunset time would cause workers to go to bed earlier and rise later, so it would not affect sleep duration. But workers face morning coordination constraints due to work and school scheduling (Hamermesh et al., 2008), so earlier sunset and earlier bedtime increase sleep duration. (We verify that sunrise time does not predict sleep duration in ATUS below.) The arguments for validity are different for the two instruments and we discuss them separately below.

Seasonal variation

Figure 3.1 shows sunset times across the continental United States on the summer solstice (Panel 3.1a), the vernal equinox (Panel 3.2), and the winter solstice (Panel 3.1b) in 2012. Darker reds indicate later sunset times. Sunset time differences at the equinox are equivalent to long-run average differences. On average, locations farther west have later sunset times than locations farther east within each time zone. The exact difference, however, changes seasonally, with locations farther north experiencing later sunset during the summer and the reverse in the winter. This variation in the angle of the sunset gradient is caused by changes in solar declination, or the angle of the sun relative to the equator.

For more discussion of solar mechanics, see Section 3.8.2.



(a) Summer solstice, Jun 20 (b) Winter solstice, Dec 21

Figure 3.1: Seasonal sunset time

Each map shows sunset time for the continental United States in 2012. Panel (a) is for the summer solstice and Panel (b) is for the winter solstice. Sunset times are indicated by color according to the scale under each figure. Darker red indicates later sunset, lighter red indicates earlier. The time zone boundaries are given by bold black lines.

In our models of seasonal effects, validity requires that other wage determinants not co-vary with sunset time within a location. While many wage determinants potentially exhibit seasonal patterns, there is no reason a priori to assume those seasonal patterns match the seasonal pattern in sunset time. Helpfully for our strategy, seasonality in sunset time is different at different latitudes, with northern locations exhibiting much higher variance in daily sunset (earlier in winter, later in summer) compared to southern locations. This allows us to disentangle seasonal variation in sunset from seasonal variation tied to the calendar, for example the December shopping season. Potentially important omitted variables remain, however. We test the robustness of our estimates to additional controls, including weather and finer time dummies, in Section sec:seas-robustn-checks. There is one potential confounder we cannot address. Seasonal variation in sunset time is almost perfectly correlated with seasonal variation in daylight duration. Note, however, that our sleep model predicts higher wages in winter, when sleep is high and daylight duration is low. If low daylight duration leads to poor mood and reduced productivity, this will bias our specifications against finding wage effects from sleep.

Long-run variation

Figure 3.2 illustrates the source of our long-run variation in average sunset time across locations. As the sun sets, eastern locations grow dark earlier than western locations. On average, residents in more easterly locations go to bed earlier and sleep longer. The maximum difference in sunset time within a US time zone is approximately one hour.

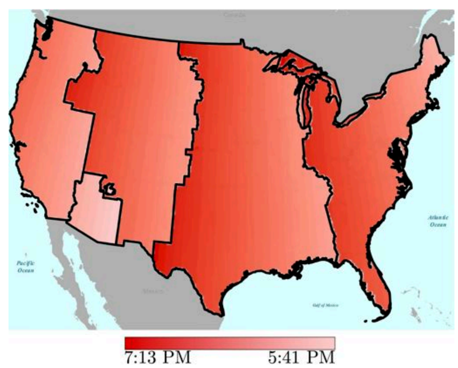


Figure 3.2: Vernal equinox, Mar 20

Map shows sunset time at the vernal equinox for the continental United States in 2012. Darker red indicates later sunset, lighter red indicates earlier. The time zone boundaries are given by bold black lines.

The difference in average sunset time between two locations over the year is plausibly orthogonal to other factors influencing the labor market, making average sunset time a valid instrument. In particular, time zone boundaries break the link between average sunset time and longitude. On average sunset time is, by construction, orthogonal to latitude, however controlling for latitude accounts for the North-South wage gradient in the US. All locations on earth experience the same average daylight duration over the year, so this is not a problematic omitted variable.

The design of US time zones derived primarily from scientific, rather than commercial, considerations. Railroads implemented the first US time zones, called Standard

Railroad Time (SRT), on November 18, 1883. They replaced a patchwork of railroad time standards and were quickly adopted by the US government and Western Union (Allen, 1883; Anonymous, 1883). While railroads were the first adopters, the primary impetus for standard time and the zone plan itself came from scientists concerned with problems like simultaneous observation of the aurora borealis at different points across the US (Bartky, 1989). The width of a zone, 15 degrees of longitude, was chosen to correspond with a one-hour difference in solar time (LOC, 2010). Ultimately, US time zones derive from the speed at which the earth rotates and the historical accident that drew the Prime Meridian through Greenwich, England: King Charles II chose Greenwich as the site for the Royal Observatory in 1675.

Endogenous modifications to time zone borders could have undermined this initial randomization. Indeed, state and local governments may petition the Department of Transportation to switch time zones, which has resulted in a long-run westward movement of boundaries (USNO, 2014). This movement means that the precise location of the boundary is endogenous and research designs based on comparing nearby communities on opposite sides of the boundary might be biased. Note, however, that the westward movement of boundaries is the opposite of what we expect if counties are choosing their time zone based on sleep-driven productivity considerations. Switching from being on the eastern side of a time zone to the western side (which is what has happened to shift the time zone boundaries) moves the county from getting the "best" average sunset treatment to getting the "worst" in terms of sleep duration. Moreover, our design does not depend on the exact location of the boundary, but on the relative longitudes of cities across a time zone; the distance between the easternmost city in our data and the border is common to all observations in the time zone and does not contribute to our coefficients of interest. (In Table 3.7 we show our results are robust to the exclusion of all counties on time zone borders.) To avoid potential endogeneity, we drop locations that do not observe daylight saving time. Finally, while time zone borders often coincide with state borders, they frequently do not, and twelve of the lower 48 US states span multiple time zones (Hamermesh et al., 2008).

Current or past worker sorting on sunset time would also threaten the validity of our average sunset time instrument. We provide empirical evidence against such sorting in Section 3.8.5. Furthermore, visual inspection of Figure 3.1 makes a sorting story difficult to credit. There is no intuitive similarity across locations with the same average local sunset time. Central Kentucky is not obviously like Eastern Colorado, nor is San Francisco like the Ozarks. To test for contemporary worker sorting, we regress county demographics on our average sunset instrument in Appendix Table 3.18. We do find a significant relationship between average sunset time and population density, which motivates our use of a flexible control for this variable. Estimated coefficients for other variables are generally small and not statistically significant. We also investigate the possibility of sorting responses to the 1883 institution of time zones in Section 3.8.5.

Firms might also sort on average sunset time, but simple firm optimization theory suggests that firms do not have strong incentives to do so. If a firm pays its workers their marginal product, managers may not care whether that marginal product is slightly higher or lower. Nonetheless this sorting is a theoretical possibility, and we test for it by regressing total wage bill in a county on sunset time and find no effect. In contrast, per-capita wage bill is influenced by sunset time, as shown in Table 3.9.

Other possible channels for failure of our exclusion restriction are discussed below, and for issues that are amenable to empirical investigation, results are shown in Section 3.6.3. First, if sunrise and sunset shift the timing of activities within a day, this could conceivably influence productivity in ways that are hard to anticipate. In part, this motivates the use of our instrument, which induces changes in sleep small enough to be unlikely to trigger schedule changes but large enough to identify effects. Hamermesh et al., 2008 show that, conditional on hours worked, sunrise and sunset do not alter within-day work schedules.

Second, introspection suggests that average sunset time might have direct effects on mood and thus productivity. This is substantially more difficult to argue given locations do not differ in average daylight duration. Even if average sunset is correlated with mood, this could be the result of changes in sleep duration (Minkel et al., 2011). We would have to believe that conditional on daylight duration and sleep time, average sunset still has direct effects on mood, perhaps through an interaction with schedule. For example, if a worker anticipates eating dinner in darkness, perhaps she is sad and less productive all day. If this were true, it would create downward bias in our estimates: workers closer to the eastern edge of a time zone would be sad (reducing productivity) and sleep more (boosting productivity). There are numerous such possible narratives and we cannot sign the potential net bias.

3.3.3 Wage setting and measurement error

Even in the case where we correctly identify exogeneous changes in short-run sleep using our seasonal instrument, there is an additional identification issue inherent in studying wages rather than productivity: timing mismatch between observations of sleep and wages combined with a potentially low frequency relationship between wages and productivity mean that our seasonal estimate will necessarily be biased. For the survey day, we observe that day's sleep and the income reported by the individual, but if sleep is productive and earnings are a function of productivity, then the wage we observe is actually based on past sleep, not the contemporaneous sleep that we see. Luckily, using seasonal sunset to predict sleep in the first stage of Equation (3.1) provides us, under an assumption about the function that relates productivity to earnings, with a functional form for the relationship between observed sleep and earnings-relevant sleep. In the

Appendix, we exploit this relationship to get exact bounds on the expected bias.

This bias is present in any seasonal estimate, but can be relatively benign. Consider, for instance, a piece-rate worker paid each day. Our observation of this worker's earnings could be based on yesterday's earnings and therefore sleep the previous night. We observe tonight's sleep, however, so the timing of our sleep observation is off by one night. Since the seasonal component of sleep is highly autocorrelated, the error in our estimate will be slight because we are using almost the right variable. In general, however, earnings change more slowly, and the degree of bias in the seasonal estimate can be large.

To calculate the bias, note that the equation for the seasonal component of sunset time is a known expression and therefore induces, in the first stage, a known function for the seasonal variation in sleep. This functional form is a sinosoidal pattern with wavelength equal to one year. In the Appendix we show that if earnings are a linear function of average productivity, then the estimated seasonal coefficient has an asymptotic bias that depends only on the distribution of D, the frequency of earnings changes in the popultaion, and a known trigonometric function. In particular

$$\frac{\hat{\beta}}{\beta} = \sum_{D} D^{-2} \sum_{k=1}^{D} \sum_{j=0}^{D-1} \cos(k+j) \Pr(D) \tag{3.3}$$

where $\hat{\beta}$ is the estimate from the second stage of Equation (3.1).

Barattieri et al. (2010, Figures 12 and 13) provide estimates of the density function for D, allowing us to calculate this expression. The authors provide two sets of estimates: one based on raw, reported earnings and another based on earnings that have been cleaned to remove measurement error. The raw series corresponds to the earnings variable that we use for our headline estimates. For a given individual, these earnings can vary over time due to contractual wage changes, changes in real take-home pay unrelated to wage (like overtime or commision), and measurement error. Using this measure, Equation (3.3) is 0.25, indicating that our estimate is one-quarter the size of the true coefficient.

Ideally, we would like to calculate the distribution of D using only contractual wage changes and other changes in take-home income caused by productivity changes, but due to the presence of measurement error, we view 0.25 as a lower bound on the attenuation of our estimate, with one important caveat. Since Barattieri et al. (2010) provide estimates of earnings changes only at 4 month intervals, this bound could overstate attenuation because it will under-weight changes that occur in less than 4 months ($D \leq 120$). From Figure 3.7, one can see that underweighting these high-frequency changes will substantially increase the bias.

Using the cleaned series from Barattieri et al. (2010), Equation (3.12) is -0.006, indicating that our estimate would be fully attenuated. The cleaned series removes measurement error but also likely removes real take-home pay changes, which would raise the frequency of earnings changes. Thus, we view this as a upper bound on the degree of attenuation. In conclusion, we expect, *a priori* that our seasonal should either be 0 or no more than one-quarter of the long-run estimate, which does not suffer from this source of bias.

3.4 Data

The most recent and largest data set from the United States containing both sleep time and wage information is the American Time Use Survey (ATUS), which asks a subset of Current Population Survey (CPS) participants to fill out a time use diary for one day. ATUS began in 2003 and the most recent data are for 2013. For this study, we use the sample of individuals age 18 or older who report receiving positive weekly wages from a primary or secondary job. Summary statistics for variables of interest are given in

Table 3.1 along with a comparison between early and late sunset time areas. The table shows values for all individuals who report earning a weekly wage. (We discuss data processing in more detail in the appendix.)

Table 3.1: ATUS Summary Statistics

Variable	Early Sunset Mean/(SD)	Late Sunset Mean/(SD)	Difference (SE)	Obs.
Weekly earnings ($)	856.7	848.3	8.4*	76,062
	(633.6)	(623.1)	(4.6)	
Hourly wage ($)	15.7	15.5	0.18*	43,927
	(9.6)	(9.3)	(0.09)	
Sleep (min/week)	3516.0	3463.5	52.6***	76,062
	(889.0)	(861.1)	(6.3)	
Sunset time (24 hr)	17.6	20.1	-2.5***	76,062
	(0.7)	(0.5)	(0.005)	
Work (min/week)	1739.0	1734.6	4.4	76,062
	(1814.1)	(1817.7)	(13.2)	
Female (0/1)	0.53	0.53	0.005	76,062
	(0.50)	(0.50)	(0.004)	
Age (years)	42.3	42.22	0.03	76,062
	(12.5)	(12.6)	(0.09)	
Race, white (0/1)	0.82	0.82	-0.0007	76,062
	(0.39)	(0.38)	(0.0028)	
Race, black (0/1)	0.13	0.13	-0.0002	76,062
	(0.33)	(0.33)	(0.0024)	
Weekend (0/1)	0.51	0.51	0.004	76,062
	(0.50)	(0.50)	(0.004)	
HS or less (0/1)	0.33	0.33	-0.008**	76,062
	(0.47)	(0.47)	(0.003)	
Some college	0.29	0.30	-0.002	76,062
	(0.46)	(0.46)	(0.003)	
College	0.24	0.236	0.003	76,062
	(0.43)	(0.42)	(0.003)	
Number of children	0.96	0.96	0.004	76,062
	(1.12)	(1.12)	(0.008)	
Ever married (0/1)	0.76	0.76	-0.001	76,062
	(0.43)	(0.43)	(0.003)	

Summary statistics for two sub-samples from ATUS are shown. Early sunset is defined as having a sunset time earlier than the median, and late sunset time is later than the median. Significance is determined from a t-test on the difference between means. Total observations are given in the far right column. The early and late sunset time groups are samples with half of these observations each.

Aside from giving basic information on the sample, Table 3.1 also provides initial evidence in support of our main results. One can see that early sunset locations have significantly higher wages and sleep duration than areas with later sunset times. (Dividing the data in this way conflates seasonal and long-run variation in sunset time.) In contrast, other individual characteristics are well balanced across the two groups. Out of 11 other tests, only one difference is significant—the fraction of the population with a high school degree or less. This difference works in the direction of explaining the difference in wages in the two groups, but other (insignificant) differences work in the opposite direction. Results controlling for these characteristics are reported in Sections 3.5.1, 3.5.2, 3.6.1, and 3.6.2.

To assign locations to individuals in ATUS, we began by merging the ATUS data with the corresponding CPS data (the match rate was 100%). For a given individual, the CPS data often contain location at the county level. This variable is censored for individuals living in counties with fewer than $100,000$ residents. When county is available, we assign the county centroid as an individual's location. We have county location for approximately 44% of ATUS observations. For an additional 28% of observations, we observe location at the level of Census CBSA, a small group of counties in the same metropolitan area. For the remaining 28% of observations, ATUS contains location at the state level. We assign the 2010 population-weighted state centroid (computed by the Census) as the location for these individuals. In all cases where we refer to Federal Information Processing Standards (FIPS) codes, we are referring to either the county- (FIPS 6-4) or CBSA-level code, if available, or the state level code (FIPS 5-2) where more detailed location data is unavailable.

Nighttime sleep is our primary sleep measure. We remove any sleep that starts and ends during daylight hours on the date of diary entry. This will exclude naps, which might be an adaptation strategy for some short sleepers. Importantly, it also removes night-shift workers, for whom the sunset instrument should not be relevant. ATUS gathers data on all sleep during the course of a single 24 hour period for each individual, so there are potentially other ways to calculate naps, and our results are robust to alternative definitions.

Our primary wage measure is "usual weekly earnings" as reported in ATUS. This variable is defined for all respondents who have positive labor income and are not self-employed. It is top-coded above \$2,884.61. We also estimate a version of our model including only workers who receive an hourly wage, "hourly earnings at main job" as reported in ATUS. This variable is likewise top-coded at the level such that hourly earnings multiplied by usual weekly hours equals \$2,884.61. Some control variables (e.g. occupation codes) appear in both ATUS and CPS files, with very minor differences across the two versions. Where possible we use ATUS variables. Some variables (e.g. race) are available only in the CPS. Our preferred regression specifications include a set of 22 occupation dummies or shares based on the ATUS "trdtocc1" variable, which categorizes the respondent's main job. Examples include "education, training, and library occupations" and "food preparation and serving related occupations."

The main shortcoming of ATUS is that it asks a new cross section of individuals for time use diaries each year, so we cannot construct an individual-level panel. As the summary statistics make clear, however, it offers a rich set of covariates including education, gender, race, and household characteristics. For a more detailed description of ATUS, see Hamermesh et al. (2005). Importantly, ATUS also releases the exact date that the survey was conducted. Using this date and respondent location, we are able to determine sunset time for each individual in the dataset using solar mechanics algorithms from Meeus (1991). We compute annual average sunset time by computing sunset for each day in an individual's location, then calculating the mean over days of the year.

The Quarterly Census of Employment and Wages (QCEW), collected by the US Bureau of Labor Statistics, includes information on wages and employment (workers, not hours) at the county level. We construct a panel in counties, 1990-2013, in order to investigate the reduced-form effects of our long-run instrument. Appendix Table 3.13 presents summary statistics.

3.5 Short-run results: the effect of sleep on wages

This section examines whether a marginal increase in weekly average sleep will, over a seasonal time horizon, change hourly wage. Here, we present results from ATUS on this question. Estimation methodology is described in Section 3.3.1.

3.5.1 Primary short-run results

Table 3.2: Short-run effects

	First stage	Reduced form	2SLS	OLS
	Sleep	ln(earnings)	ln(earnings)	ln(earnings)
Seasonal sunset time	-24.0***	-0.0039**		
	(2.28)	(0.0017)		
Sleep			0.00016**	-0.000028***
			(0.000074)	(0.0000026)
Individual controls	Yes	Yes	Yes	Yes
Time controls	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes
FIPS FEs	Yes	Yes	Yes	Yes
Observations	76062	76062	76062	76062
F-stat on IV			118.66	
Elasticity			0.60	

The table shows results from estimating Equation (3.1). The first three columns show the first stage, reduced form, and two-stage least squares estimates. The fourth column reports the uninstrumented version of the second stage of Equation (3.1). The dependent variable is indicated at the top of each column. Earnings refers to "usual weekly earnings". Sleep is measured in minutes per week and sunset time in hours. Controls include: location fixed effects; race dummies; age; age squared; a full-time dummy; a gender dummy; dummies for holiday, day of week, and year; and occupation dummies. Standard errors, clustered at the FIPS code (location) level, are reported in parentheses. Significance indicated by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3.2 presents the estimated effect of short-run sleep changes on wages, instrumenting for sleep on the ATUS diary date using the local sunset time on that date. Column 1 reports the first-stage estimates. Consistent with the dicusssion in Section 3.3.2, a one-hour increase in local sunset time causes an individual to sleep about 24 minutes less per week. The raw covariance between sleep and wage (Column 4) is slightly negative, as we would expect given strong reverse causality. Using the sunset time instrument, however, the estimated coefficient of .00016 log points per minute of weekly average sleep is positive and significant. For comparison with our long-run estimates, we can multiply this estimate by 60 and obtain .0096. A one-hour increase in weekly average sleep over a seasonal time horizon increases wages by just under one percent. Because we include location fixed effects as controls, this specification does not use any of the long-run variation in sunset time to identify our effect of interest.

We cluster standard errors at the FIPS code level (county, CBSA, or state) to reflect that the exogenous variation is at the group rather than the individual level. Note that clustering at higher levels does not change the inference. We have clustered up to the state level without any appreciable difference in standard errors. The first-stage F statistic of 109.9 well exceeds the relevant Stock-Yogo critical value of 16.38, so we reject the null hypothesis of weak instruments, where "weak" is defined as true size greater than 10% for a nominal 5% test (Stock and Yogo, 2002). This reassures us that the results of our t-tests are reasonable.

Expressed as an elasticity, our IV estimate is .6. This magnitude is consistent with the experimental evidence summarized in Table 3.12. Medical researchers have typically found elasticities of task performance with respect to sleep duration of approximately four. If wages are equal to a worker's marginal physical product multiplied by output price, we expect such performance effects to produce equally large wage effects. The smaller magnitude of our estimate, relative to the medical literature, may reflect differences between laboratory tasks and actual work tasks or the broader scope for adaptation (e.g.

stimulant use) outside the lab. Many workers' wages do not vary over a seasonal time horizon and such workers will also decrease the magnitude of our estimate (relative to a case with perfectly flexible wages).

Unaided intuition might suggest smaller effects of sleep on performance, but intuition provides a poor sense of this relationship: Van Dongen et al. (2003) showed that subjects' self-reported fatigue quickly stabilized after a few days of sleep reduction, even as their performance continued to decline. Van Dongen et al. (2003) also found that several days of two–hour sleep reductions reduced performance by as much as a night of complete sleep deprivation. Based on this study, the experience of attempting to work after a completely sleepless night likely provides a better sense of the performance effects from reducing short-run weekly sleep by one hour per night.

Several nuances bearing on the interpretation of our estimate warrant discussion. First, our instrument affects all workers in a location identically, which changes the interpretation of our results if there are productivity spillovers across workers. While we do not know if sleep generates such spillovers, Moretti (2004) finds evidence that human capital does. In such a case our estimated β captures not the effect of increasing individual sleep, but rather the effect of increasing sleep for all workers in a location. Second, managers might set wages based on average productivity in a location rather than individual worker productivity. Under this assumption, an increase in sleep by an individual would have no effect on her wage, as it would not appreciably change average productivity. In a case like this, our estimate captures the effect of increased sleep by all workers on average productivity, rather than an individual-level effect. Finally, it is possible our instrument influences both sleep duration and sleep quality. This is true, however, of any exogenous variation in sleep, even in a laboratory setting. In such a case our estimates are still consistent for the effect of an exogenous sleep change, but the interpretation changes slightly.

3.5.2 Short-run robustness checks

We test the sensitivity of our short-run results to a wide variety of robustness checks, including varying controls for worker and location characteristics, and varying controls for seasonal trends. These checks generally indicate that the results are robust to a broad set of specifications.

Table 3.3 demonstrates that while different control sets yield different precision, they have very little influence on the point estimate. Controlling for a quadratic in usual hours worked returns estimates very similar to our primary short-run results. Note that we deliberately do not include usual hours worked as a control variable in the main specification. This allows the worker to take additional sleep time out of either work time or other (non-work, non-sleep) time. By controlling for work time, we would be forcing all changes in sleep to come out of other time, which might bias our estimates. Nonetheless the robustness of our result to this control is encouraging.

Omitting controls entirely, or omitting various subsets, likewise makes little difference. The addition of finer race controls or education controls does not meaningfully change the estimate. Finally, if we cluster at the state level, rather than the location level, the standard errors are approximately the same.

The temperature data is the NCEP/NCAR reanalysis produced by Kalnay et al. (1996). The data is available at a daily frequency on a two-by-two degree spatial grid. We use the daily temperature from the nearest grid point for estimation.

In Table 3.4 we show our results are robust to varying controls for seasonality, including quarter fixed effects and a cubic time trend. We also subset to weekdays and workdays (these differ because some workers work on the weekend) and find similar estimates.

Appendix Table 3.14 reprises the specification from Table 3.2, but only for workers who report being paid hourly. These results are quite similar. In principle the

Table 3.3: Robustness of short-run estimates: Controls

	First stage		Reduced form		2SLS	
	Sleep		ln(earnings)		ln(earnings)	
Usual work hours quadratic						
Seasonal sunset time	-23.8***	(2.28)	-0.0043***	(0.0016)		
Sleep					0.00018**	(0.000071)
No controls						
Seasonal sunset time	-23.1***	(2.46)	-0.0041*	(0.0023)		
Sleep					0.00018*	(0.00010)
Only time controls						
Seasonal sunset time	-23.3***	(2.30)	-0.0044*	(0.0022)		
Sleep					0.00019*	(0.00010)
Only FIPS FEs						
Seasonal sunset time	-23.4***	(2.46)	-0.0028	(0.0022)		
Sleep					0.00012	(0.000100)
Only time controls and FIPS FEs						
Seasonal sunset time	-23.7***	(2.30)	-0.0033	(0.0022)		
Sleep					0.00014	(0.000098)
More controls						
Seasonal sunset time	-23.9***	(2.26)	-0.0041**	(0.0016)		
Sleep					0.00017**	(0.000072)
More controls: Education						
Seasonal sunset time	-24.1***	(2.25)	-0.0035**	(0.0016)		
Sleep					0.00015**	(0.000069)
State clustering						
Seasonal sunset time	-24.0***	(2.29)	-0.0039**	(0.0016)		
Sleep					0.00016**	(0.000070)

The table shows results from estimating Equation (3.1). Dependent variable is indicated at the top of each column. Unless otherwise noted, controls, number of observations, and standard errors are the same as in Table 3.2. Significance indicated by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

coefficients are not directly comparable, since the change in weekly wage includes both wage and hour effects. As we show below, however, our seasonal sunset time differences produce very small changes in work hours, so the two tables represent roughly the same change.

In part, we examine hourly wage earners to address concerns like those raised in Borjas (1980) about the use of constructed hourly wage measures. The hourly wage results also allow us to explore one interesting possible source of heterogeneity between salaried and hourly workers. One might expect that salaried workers are engaged in less routine tasks so attention lapses or other sleep-driven performance changes might be

Table 3.4: Robustness of short-run estimates: Time controls and sample

	First stage Sleep	Reduced form ln(earnings)	2SLS ln(earnings)
Quarter FEs			
Seasonal sunset time	-22.2***(4.56)	-0.0090***(0.0032)	
Sleep			0.00041** (0.00016)
Cubic time trend			
Seasonal sunset time	-24.6***(2.30)	-0.0038** (0.0017)	
Sleep			0.00015** (0.000072)
No weekends			
Seasonal sunset time	-16.7***(3.12)	-0.0035* (0.0021)	
Sleep			0.00021 (0.00014)
Observations	37525	37525	37525
Only workdays			
Seasonal sunset time	-17.6***(2.77)	-0.0059***(0.0022)	
Sleep			0.00033** (0.00014)
Observations	43296	43296	43296

The table shows results from estimating Equation (3.1). Dependent variable is indicated at the top of each column. Unless otherwise noted, controls, number of observations, and standard errors are the same as in Table 3.2. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

more costly. We do not find a substantial or significant difference between the two groups using our preferred specification, but we examine this comparison using QCEW data in Section 3.6.4.

3.6 Long-run results: the effect of sleep on wages

3.6.1 Primary results

We now turn to the long-run effects of average weekly sleep in a location on average wage in that location. Estimation methodology is described in Section 3.3.1.

Table 3.5 presents long-run effects using the average sunset time instrument described in Section 3.3.2. The first column reports the first-stage estimates for average sleep, within a location, as a function of average sunset time. Average weekly sleep in a location where the sun sets one hour later is approximately one hour less. This

Table 3.5: Long-run effects

	First stage Sleep	Reduced form ln(earnings)	2SLS ln(earnings)	OLS ln(earnings)
Avg. sunset time	-64.0***	-0.049***		
	(16.2)	(0.019)		
Sleep			0.00077**	0.00012***
			(0.00033)	(0.000044)
Geographic controls	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Observations	529	529	529	529
Adjusted R^2	0.137	0.809	0.718	0.809
F-stat on IV	15.52			
Elasticity			2.33	

The table shows results from estimating Equation (3.2), with location-level observations weighted by the count of underlying ATUS respondents. The first three columns show the first stage, reduced form, and two-stage least squares estimates. The fourth column reports the uninstrumented version of the second stage of Equation (3.2). The dependent variable is indicated at the top of each column. Earnings refers to "usual weekly earnings". Sleep is measured in minutes per week and sunset time in hours. Controls include: coastal distance, a 10-piece linear spline in latitude; share full time; median age; race shares; occupation shares; and a 5-piece linear spline in population density. White heteroskedasticity-robust standard errors reported in parentheses. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

corresponds to the range of variation produced by the long-run sunset time instrument. Using the IV specification, we again recover a positive, significant effect of average sleep on average wage: .00077. For comparison with our short-run estimates, we can multiply this estimate by 60 and obtain .046. A one-hour increase in weekly average sleep over a seasonal time horizon increases wages by about 4.5%. This is roughly 4.5 times larger than our short-run estimate, consistent with greater wage flexibility in the long run and the difficulty of adapting to long-run sleep reductions with strategies like stimulant use. Expressed as an elasticity, our estimate is 2.3, again smaller than the average value of 4 observed in medical studies summarized in Table 3.12.

We report White heteroskedasticity-robust standard errors. We do not cluster because the data underlying Table 3.5 are collapsed to a cross section (in locations) and our instrument varies exogenously across locations. The first-stage F statistic of 15.5 falls slightly short of the Stock-Yogo critical value (16.38) for a maximum size of 10%

in a nominal 5% test. It is substantially greater than the value (8.96) for a maximum size of 15% in a nominal 5% test. This recommends some modest caution in interpreting the results of our t tests, but rules out gross failures of size control.

The estimated effects in Table 3.5 correspond to long-run, or even permanent, changes in the mean values of variables within a location. This means they potentially incorporate spillovers across workers and general-equilibrium effects accumulated over time. While they demonstrate sleep has large long-run effects on wage, they may not provide a good estimate of, for example, the effect from increased long-run sleep by a single worker, or even all the workers in a single business.

Taking average values for wages and assuming 50 work weeks per year, one can calculate the annual income effects implied by our long-run estimates. If mean weekly sleep in a location increased by one hour per week and work time remained unchanged, mean annual income would rise by about $1,950. In reality, extra sleep comes out of both work and non-work time. If workers took roughly half of the extra sleep hour out of work time, as we find in Table 3.10, then a one-hour increase in weekly mean sleep in a location would increase mean annual income by about $1,340. If extra sleep came solely at the expense of work time, the income increase would be $880. That figure naturally leads one to ask why workers don't work less and sleep more. One possible explanation lies in the spillovers and general-equilibrium effects our estimate incorporates. Because our estimates are based on location means, they likely overstate the effects an individual worker would experience from changing her sleep in isolation.

3.6.2 Long-run robustness checks

As before, we test the sensitivity of our primary results to a wide variety of robustness checks. Broadly, we examine the inclusion or exclusion of controls and changes to the estimation sample. We also conduct a deeper exploration of geographic sorting that might invalidate our instrument in Section 3.6.3. Together our checks indicate that the results reported above are robust to varying assumptions and changes in estimation technique.

We first show the linear results hold under alternative control variable specifications in Table 3.6. The first pair of rows show that including a quadratic in usual hours worked does not move the coefficient estimate appreciably. (As discussed above, we exclude hours worked from the primary specification to avoid bias from forcing sleep increases to come at the expense of leisure.) Inclusion of this control does not move the coefficient estimate.

Table 3.6: Robustness of long-run estimates: Controls

	First stage Sleep	Reduced form ln(earnings)	2SLS ln(earnings)
Usual work hours quadratic			
Avg. sunset time	-60.0***(16.3)	-0.053***(0.019)	
Sleep			0.00088** (0.00036)
Only geographic controls			
Avg. sunset time	-70.1***(15.9)	-0.12** (0.051)	
Sleep			0.0017** (0.00079)
No occupation controls			
Avg. sunset time	-62.8***(14.8)	-0.064 (0.039)	
Sleep			0.0010 (0.00068)
Education controls			
Avg. sunset time	-59.2***(16.1)	-0.042***(0.016)	
Sleep			0.00072** (0.00031)
Median age squared control			
Avg. sunset time	-64.0***(16.2)	-0.050***(0.019)	
Sleep			0.00078** (0.00033)
Industry controls			
Avg. sunset time	-61.0***(18.1)	-0.056***(0.019)	
Sleep			0.00092** (0.00037)
Region indicators			
Avg. sunset time	-63.7***(17.8)	-0.046** (0.021)	
Sleep			0.00072** (0.00036)
Longitude control			
Avg. sunset time	-60.5***(16.1)	-0.047** (0.019)	
Sleep			0.00077** (0.00035)

The table shows results from estimating Equation (3.2). Dependent variable is indicated at the top of each column. Unless otherwise noted, controls, number of observations, and standard errors are the same as in Table 3.5. Education controls are shares of observations in 5 attainment levels. Industry controls are shares of observations in 52 industries. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Next, we include only geographic controls, without any other covariates. Although the exclusion restriction for the validity of our instrumental variable estimate is based on the error term for the full model, it is reassuring to see that the coefficient from this minimal specification is just over one standard error away from our preferred estimate (a hypothesis test fails to reject the null hypothesis of zero difference in these two estimates). The additional demographic controls do make the coefficient estimate more precise, however, as can be seen by comparing standard errors between the main result and robustness check. This result implies that sunset time is not highly correlated with the covariates in the main specification, which also provides initial evidence against sorting on sunset time. We then implement a less drastic change in control variables, removing occupation indicators from our preferred specification. These variables are potentially endogenous, so it is important to show that our coefficient estimate does not change when they are excluded. We next add richer sets of individual controls. These include squared median age, a set of industry shares, region indicators, and a longitude control. Adding these additional variables does not change the results.

The second set of robustness checks, presented in Table 3.7, deals with changes to the sample. In our main specification, we control for the share of full-time workers. The first estimates in Table 3.7, however, show that our main results still hold even when we drop part-time employees entirely.

ATUS oversamples weekends so that roughly half of the total observations are from weekend dates (see Table 3.1). We test the sensitivity of our results to this by dropping the weekend diary entries entirely. The estimate is similar to baseline, albeit less precise. While the number of location-level observations is the same, this specification drops roughly half of the underlying ATUS sample. Next we estimate our preferred model excluding counties within four degrees longitude of a time zone border. This drops all counties that might have selected into a time zone based on economic

Table 3.7: Robustness of long-run estimates: Sample

	First stage Sleep		Reduced form ln(earnings)		2SLS ln(earnings)	
Only full time workers						
Avg. sunset time	-44.9**	(19.2)	-0.054***	(0.020)		
Sleep					0.0012*	(0.00063)
Observations	529		529		529	
No weekend diaries						
Avg. sunset time	-54.0**	(21.2)	-0.074***	(0.026)		
Sleep					0.0014**	(0.00069)
Observations	529		529		529	
No time zone border counties						
Avg. sunset time	-97.8***	(37.2)	-0.11***	(0.039)		
Sleep					0.0011**	(0.00050)
Observations	340		340		340	
No Eastern time zone						
Avg. sunset time	-19.6	(26.8)	-0.071**	(0.032)		
Sleep					0.0036	(0.0046)
Observations	244		244		244	
No high-wage cities						
Avg. sunset time	-27.7**	(13.6)	-0.029	(0.021)		
Sleep					0.0011	(0.00087)
Observations	476		476		476	

The table shows results from estimating Equation (3.2). Dependent variable is indicated at the top of each column. Unless otherwise noted, controls, number of observations, and standard errors are the same as in Table 3.5. Results reported under "No high-wage cities" exclude workers in San Francisco, Los Angeles, Chicago, Boston, and New York. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

considerations. Again the results are not statistically distinguishable from our preferred estimates. Exclusion of wealthier, denser Eastern time zone or exclusion of selected high-wage cities similarly has only modest impact on coefficient estimates. In Appendix Table 3.15 we estimate long-run effects for only those workers who report an hourly wage, and recover similar point estimates.

All of these robustness checks support our primary result. The first-stage coefficient on sleep is generally stable between 60 and 70 minutes per week, and typically significant at the 1% level. The reduced-form estimate is more sensitive to specification, but remains in the interval from $-.04$ to $-.07$ in nearly all cases. Its statistical significance varies, but the reduced-form estimate is most commonly significant at the 1%

level. The overall pattern of results, showing stability of the coefficients under different reasonable samples and control sets, suggests our assumption of instrument validity is reasonable.

3.6.3 Instrument validity

The robustness checks in Tables 3.3, 3.4, 3.6 and 3.7 provide evidence against some of the more plausible potential omitted variable or specification failures. In the following subsections, we conduct more direct tests for some of the potential identification failures discussed in Section 3.3.2.

Short-run validity

Daily sunset time follows a seasonal pattern and induces a seasonal pattern in sleep. If other wage determinants follow a causally unrelated, but similar seasonal pattern, we might recover a spurious estimate. First we investigate the possibility that seasonal trends in the composition of employment might bias our results. Appendix Figure 3.6a shows that occupation shares in our sample are constant over the months of the year. Appendix Figure 3.6b shows that the share of ATUS respondents reporting a positive wage is likewise constant over the year.

Long-run validity

One of the primary channels through which pernicious omitted variables might appear is through individuals sorting across locations based on average sunset or its correlates. For sorting to threaten identification, workers would have to sort based on the *timing* of daylight. Sorting on daylight duration would not bias our estimates, as average sunset time is independent of daylight duration. Note that even if workers actually sort on the sunset-induced wage differential, we can still test for the problem by examining sorting on sunset.

Before proceeding with empirical tests, it will be helpful to consider a few theoretical points. First, a worker who decides to sleep more need not move to another city; she can simply sleep more. Only if workers suffer some optimization failure, like an inability to commit to a particular bedtime, will they have an incentive to sort. Second, an optimizing worker responds to real, not nominal, income. If home prices in more productive (higher sleep) locations adjust to offset wage gains, workers will not have a financial incentive to move. This is exactly the prediction of a sorting model like Roback (1982). With perfect worker and firm mobility, the gains from a productive location-specific amenity accrue to owners of land, the fixed factor. Such a model predicts that locations with earlier average sunset times will have higher rents and house prices, even without worker sorting on ability. Using county-level Census data from 2010, Table 3.8 provides evidence that this is indeed so. We regress log median county home value on average sunset time and a rich set of controls.

$$\ln(\text{median home value})_j = \beta\overline{\text{sunset}}_j + \mathbf{x}_i'\gamma + \varepsilon_j$$

A county experiencing sunset one hour earlier than a comparison county will have, on average, a median home value approximately 9% to 13% higher. This result is statistically significant at the 1% level. In levels, the estimated effect on median home value is approximately \$13,000 to \$22,000. Based on the discussion following Table 3.5, a worker's annual income gain from moving to a location where sunset is an hour earlier is approximately \$1340. The present discounted value of this increase, assuming a five percent discount rate, is approximately \$26,800. This result is roughly consistent with the prediction of the Roback model: the wage gains from additional sleep in a location accrue largely to landowners, not workers, and workers have little incentive to sort on sunset time. In Appendix Table 3.19 we show this estimate is robust to additional controls.

Table 3.8: Effects on log median home value

	Log value	Log value
Sunset time	-0.127***	-0.0851***
	(0.0235)	(0.0192)
Geographic controls	Yes	Yes
Demographic controls	No	Yes
Observations	2824	2824
Adjusted R^2	0.342	0.600

White heteroskedasticity-robust standard errors are reported in parentheses. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1. Data are 2010 5-year ACS estimates. Sunset time is the average for a given county. Geographic controls include coastal distance and a ten-piece linear spline in latitude. Demographic controls include percent female, percent in four race categories, occupation shares, and a five-piece linear spline in population density.

Our hedonic results support our interpretation of the findings in Section 3.6.1 and are consistent with a general-equilibrium model in which workers do not sort on ability. They could also, however, be consistent with worker sorting. Therefore we conduct direct sorting tests: first, we examine historical population growth patterns in response to time zone creation in 1883 and 1918. Second, we examine the relationship between current county-level characteristics and sunset time.

Table 3.18 in the Appendix compares present-day county level characteristics by regressing a number of demographic variables on sunset time. Out of nine variables, we find two estimates that are significant at the 5% level. There is a significant negative relationship between average sunset time and population density, which is why we employ a spline in this variable as a control. The results also suggest that unemployment is lower for locations with later sunset time, but this is the reverse of what we would expect if sorting or selection were driving our result. Finally, Table 3.6 also provides present-day sorting evidence by indicating that our estimate is robust to the inclusion or exclusion of demographic characteristics. This indicates that people of different ages, genders, race, and education levels are exposed to roughly equal sunset times, on average, across the United States.

Taken together, Table 3.6, Table 3.18, and Figure 3.8 suggest that sorting does not bias our results. The lack of sorting is perhaps unsurprising given the extremely small wage differences implied by our reduced-form results: even at the extremities of the widest (Central) time zone, the nominal wage differential between two locations at a given latitude is less than five percent.

3.6.4 County average wages

To corroborate the long-run results from ATUS data, we also estimate reduced-form models using data from the BLS Quarterly Census of Employment and Wages (QCEW). Unlike ATUS, QCEW data allow us to observe all US counties. In the following equation, j indexes county and t quarter-year. Data restrictions prevent us from exactly replicating our long-run ATUS specification, but the control set is similar: coastal distance, a linear spline in latitude, share female, share in four racial categories, six occupation shares, and a linear spline in population density. We include dummies for quarter-year, so we are using only the long-run variation in sunset time. The dependent variable is the average weekly wage per worker.

$$\ln(w_{jt}) = \delta_t + \mathbf{x}_j'\gamma + \beta\overline{\text{sunset}}_{jt} + \varepsilon_{jt} \tag{3.4}$$

Table 3.9 presents estimates based on the above equation. The estimate for all workers, -.026, is roughly similar to our estimate of -.049 from ATUS data; we cannot reject a null hypothesis of zero difference between the two. The QCEW allows for greater precision due to the longer time coverage of the data. The larger sample also allows us

Table 3.9: Effects on average wage

	All industries	Goods	Services
Sunset time	-0.0263**	-0.0111	-0.0392***
	(0.0120)	(0.0148)	(0.0107)
Yr-qtr FEs	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes
Observations	291271	287657	288520

Standard errors are reported in parentheses. Clustering is at the county level. Significance indicated by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Data are from the BLS Quarterly Census of Employment and Wages 1990-2013. Sunset time is the quarterly county average. Control set includes: coastal distance, a 10-piece linear spline in latitude, share female, share in four racial categories, six occupation shares, and a 5-piece linear spline in population density.

to begin exploring heterogeneity of the reduced-form effect. The estimate for services, -.039, is considerably larger than the estimate for goods, -.011. In Appendix Table 3.20 we show this result is robust to additional controls.

Other time uses

Our primary analysis demonstrates that workers experiencing an earlier sunset get more sleep. It is natural to ask where the additional sleep time comes from, and the answer to this question informs the interpretation of our estimates. Table 3.10 shows that when faced with an earlier sunset, workers increase sleep by decreasing work and leisure in roughly equal amounts. These estimates are not statistically distinguishable from zero or from each other, so this is at best suggestive evidence. Insofar as these changes in work and leisure impact worker productivity, our sleep estimates also contain those effects. While this might seem undesirable at first glance, it is unavoidable. An agent's time constraint always binds with perfect equality. Even in a laboratory setting, it is not possible to change the time use of interest without also changing at least one other time use.

Table 3.10: Waking time use as a function of sunset time

	Work time	Non-work time
Avg. sunset time	33.9	19.0
	(33.6)	(30.6)
Geographic controls	Yes	Yes
Demographic controls	Yes	Yes
Observations	529	529
Adjusted R^2	0.028	0.090

The table shows results from estimating the first stage of Equation (3.2), replacing sleep time with either work time or waking non-work time as the dependent variable. Dependent variable is indicated at the top of each column. Unless otherwise noted, controls, number of observations, and standard errors are the same as in Table 3.5. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Recall that in Section 3.2.2, we found the sign of the derivative $\partial T_z/\partial\alpha$, the derivative of leisure time with respect to the parameter α, was theoretically ambiguous, depending on wage and hours worked. In a regression of waking non-work time on sunset time (as in column 2 of Table 3.10), this prediction corresponds to smaller, possibly negative, coefficients on sunset time for low-wage workers and high-work, low-sleep workers (in contrast to the overall positive result from Table 3.10). To test this prediction, we estimate separate regressions for these groups. Estimates are reported in Table 3.11. Consistent with our theoretical predictions, estimates for both groups are negative, though not statistically significant.

3.7 Conclusion

Although time use is entangled in a causal web with labor market outcomes, economists have largely ignored these relationships. In particular, the profession has paid scant attention to sleep. Our results demonstrate that sleep has a powerful impact on labor market outcomes and should be considered an integral part of a worker's utility

Table 3.11: Waking non-work hours as a function of sunset time, selected groups

	Non-work time High work hours	Non-work time Low wage earners
Avg. sunset time	-54.2	-62.3
	(68.6)	(84.5)
Geographic controls	Yes	Yes
Demographic controls	Yes	Yes
Observations	495	513
Adjusted R^2	0.098	0.145

The table shows results from estimating the first stage of Equation (3.2), replacing sleep time with waking non-work time as the dependent variable. In column 1 the sample is workers who work more than 8 hrs on the diary date (7th percentile) and sleep less than 6 hrs (10th percentile). In column 2 the sample is workers with log wages below 5.44 (10th percentile). Dependent variable is indicated at the top of each column. Unless otherwise noted, controls, number of observations, and standard error clustering are the same as in Table 3.5. Significance indicated by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

maximization problem. Using individual time-use diaries matched with labor market variables from ATUS, we show that increasing short-run weekly average sleep by one hour produces a 1% higher wage, while increasing long-run weekly average sleep by 1 hour, for all workers in a location, produces a 4.5% higher wage. Our use of instrumental variables techniques addresses the reverse-causality and omitted variable problems that would bias naïve estimates. We buttress this finding with a battery of seasonal and long-run robustness checks, and a hedonic model of home prices showing that long-run wage increases are capitalized into housing.

Sleep is arguably the third most important determinant of productivity, following ability and human capital. Our findings have important implications for individuals, firms, schools, and governments. A worker who desires higher wages might be able to obtain them by increasing sleep. Firms might be able to increase profit by varying start times, providing workers with incentives to sleep more, or with information interventions (e.g. information on how to improve sleep quality or consistency). Governments conducting cost-benefit analyses of policies that change sleep time, for example daylight savings time, should consider the productivity effects to design efficient policies. Countries spanning a wide range of longitudes might benefit from abolishing time zones and adopting a single standard time, preserving ease of coordination while allowing firms and schools to set schedules optimally with respect to local solar cues.

Further attention should be paid to industries characterized by chronic sleep shortages. In addition to wages, optimal sleep plausibly depends on other factors like leisure complimentarities, direct sleep utility, and health optimization. Each of these trade-offs suggests an interesting research question. More broadly, our results demonstrate that non-labor time uses can have first-order effects on labor outcomes—effects that should continue to be investigated in future work.

Acknowledgements: The authors thank Prashant Bharadwaj, Julie Cullen, Gordon Dahl, Joshua Graff Zivin, Daniel Hamermesh, Mark Jacobsen, Sendhil Mullainathan, Dale Squires, Michelle White, and seminar participants at UC San Diego for helpful comments. Shrader gratefully acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1144086. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Chapter 3, in full, is currently being prepared for submission for publication of the material. Gibson, Matthew; Shrader, Jeffrey. The dissertation author was the primary investigator and author of this material.

References

Allen, W. F. (1883). “Open Letters: Standard Railway Time”. In: *The Century* 0026, p. 796.

Anonymous (1883). “The New Standard Time and Its Advantages”. In: *The Manufacturer and Builder* 15.12.

Banks, Siobhan and David F Dinges (Aug. 2007). “Behavioral and physiological consequences of sleep restriction.” In: *Journal of Clinical Sleep Medicine* 3.5, pp. 519–28.

Barattieri, Alessandro, Susanto Basu, and Peter Gottschalk (2010). “Some Evidence on the Importance of Sticky Wages”. In:

Barnes, Christopher M and David T Wagner (2009). “Changing to daylight saving time cuts into sleep and increases workplace injuries.” In: *Journal of applied psychology* 94.5, p. 1305.

Bartky, Ian (1989). “The Adoption of Standard Time”. In: *Technology and Culture* 30.1, pp. 25–56.

Becker, Gary S (1962). “Investment in Human Capital: A Theoretical Analysis”. In: *Journal of Political Economy* 70.5, pp. 9–49.

— (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.

— (1965). “A Theory of the Allocation of Time”. In: *The Economic Journal* 75.299, pp. 493–517.

Belenky, Gregory, Nancy J Wesensten, David R Thorne, Maria L Thomas, Helen C Sing, Daniel P Redmond, Michael B Russo, and Thomas J Balkin (Mar. 2003). “Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study.” In: *Journal of sleep research* 12.1, pp. 1–12.

Biddle, JE and DS Hamermesh (1990). “Sleep and the Allocation of Time”. In: *Journal of Political Economy* 98.5, pp. 922–943.

Black, Sandra E and Lisa M Lynch (2001). “How to Compete: the Impact of Workplace Practices and Information Technology on Productivity”. In: *The Review of Economics and Statistics* 83.August, pp. 434–445.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen (2012). “Americans Do IT Better: US Multinationals and the Productivity Miracle”. In: *The American Economic Review* 102.1, pp. 167–201.

Bonke, Jens (2012). “Do Morning-Type People Earn More than Evening-Type People? How Chronotypes Influence Income”. In: *Annales d’Economie et de Statistique* 105, pp. 55–76.

Borjas, George J (1980). “The relationship between wages and weekly hours of work: The role of division bias”. In: *Journal of Human Resources*, pp. 409–423.

Brochu, Pierre, Catherine Deri Armstrong, and Louis-Philippe Morin (Sept. 2012). “The trendiness of sleep: An empirical investigation into the cyclical nature of sleep time”. In: *Empirical Economics* 43.2, pp. 891–913.

Cappuccio, FP, L D’Elia, P Strazzullo, and MA Miller (2010). “Sleep Duration and All-Cause Mortality: A Systematic Review and Meta-Analysis of Prospective Studies”. In: *Sleep*.

Cohen, Daniel A, Wei Wang, James K Wyatt, Richard E Kronauer, Derk-Jan Dijk, Charles A Czeisler, and Elizabeth B Klerman (Jan. 2010). “Uncovering residual effects of chronic sleep loss on human performance”. In: *Science Translational Medicine* 2.14, 14ra3.

Dinges, D F, F Pack, K Williams, K A Gillen, J W Powell, G E Ott, C Aptowicz, and A L Pack (Apr. 1997). “Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night”. In: *Sleep* 20.4, pp. 267–77.

Frazis, Harley and Jay Stewart (Jan. 2012). “How to Think about Time-Use Data: What Inferences Can We Make about Long- and Short-Run Time Use from Time Diaries?” In: *Annals of Economics and Statistics* 105/106, pp. 231–245.

Grossman, Michael (1972). “On the Concept of Health Capital and the Demand for Health”. In: *Journal of Political Economy* 80.2, pp. 223–255.

Haines, Michael R and Inter-university Consortium for Political and Social Research (2010). *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002*.

Hamermesh, Daniel S., Harley Frazis, and Jay Stewart (2005). “The American Time-use Survey”. In: *Journal of Economic Perspectives* 19.1, pp. 221–232.

Hamermesh, Daniel S, Caitlin Knowles Myers, and Mark L Pocock (2008). “Cues for Timing and Coordination: Latitude, Letterman, and Longitude”. In: *Journal of Labor Economics* 26.2, pp. 223–246.

Kalnay, Eugenia, Masao Kanamitsu, Robert Kistler, William Collins, Dennis Deaven, Lev Gandin, Mark Iredell, Suranjana Saha, Glenn White, John Woollen, et al. (1996). “The NCEP/NCAR 40-year reanalysis project”. In: *Bulletin of the American meteorological Society* 77.3, pp. 437–471.

Krueger, Patrick M and Elliot M Friedman (May 2009). “Sleep duration in the United States: a cross-sectional population-based study.” In: *American Journal of Epidemiology* 169.9, pp. 1052–63.

Landrigan, Christopher P, Jeffrey M Rothschild, John W Cronin, Rainu Kaushal, Elisabeth Burdick, Joel T Katz, Craig M Lilly, Peter H Stone, Steven W Lockley, David W Bates, and Charles A Czeisler (Oct. 2004). “Effect of reducing interns’ work hours on serious medical errors in intensive care units”. In: *The New England Journal of Medicine* 351.18, pp. 1838–48.

LOC (2010). *Time!* Accessed Feb. 4, 2014. Library of Congress.

Lockley, Steven W., John W. Cornin, Erin E. Evans, Brian E. Cade, Clark J. Lee, Christopher P. Landrigan, Jeffrey M. Rothschild, Joel T. Katz, Craig M. Lilly, Peter H. Stone, Daniel Aeschbach, and Charles A. Czeisler (2004). “Effects of Reducing Interns’ Weekly Work Hours on Sleep and Attentional Failures”. In: *The New England Journal of Medicine*, pp. 1829–1837.

Mckenna, Benjamin S, David L Dickinson, Henry J Orff, and Sean Drummond (2007). “The effects of one night of sleep deprivation on known-risk and ambiguous-risk decisions”. In: *Journal of Sleep Research* 16.3, pp. 245–252.

Meeus, Jean H (1991). *Astronomical Algorithms*. Willmann-Bell, Incorporated.

Minkel, Jared, Oo Htaik, Siobhan Banks, and David Dinges (Jan. 2011). “Emotional expressiveness in sleep-deprived healthy adults.” In: *Behavioral sleep medicine* 9.1, pp. 5–14.

Moretti, Enrico (2004). “Workers’ Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions”. In: *The American Economic Review* 94.3, pp. 656–690.

Psacharopoulos, George and Harry Anthony Patrinos (Aug. 2004). “Returns to investment in education: A further update”. In: *Education Economics* 12.2, pp. 111–134.

Roback, Jennifer (1982). “Wages, Rents, and the Quality of Life”. In: *Journal of Political Economy* 90.6, pp. 1257–1278.

Roenneberg, Till, C Jairaj Kumar, and Martha Merrow (2007). “The human circadian clock entrains to sun time”. In: *Current Biology* 17.2, R44–R45.

Smith, Austin C (2014). “Spring Forward at Your Own Risk: Daylight Saving Time and Fatal Vehicle Crashes”.

Solon, Gary, Steven J Haider, and Jeffrey Wooldridge (2013). “What Are We Weighting For?”

Stock, James H and Motohiro Yogo (2002). “Testing for Weak Instruments in Linear IV Regression”.

Szalontai, Gábor (2006). “The demand for sleep: A South African study”. In: *Economic Modelling* 23.5, pp. 854–874.

Taras, Howard and William Potts-Datema (2005). “Sleep and student performance at school”. In: *Journal of School Health* 75.7, pp. 248–254.

Ulmer, Cheryl, Dianne Miller Wolman, Michael ME Johns, et al. (2009). *Resident duty hours: Enhancing sleep, supervision, and safety*. National Academies Press.

USNO (2014). *U.S. Time Zones*. Accessed Feb. 4, 2014. US Naval Observatory.

Van Dongen, Hans P A and David F Dinges (Apr. 2005). “Sleep, circadian rhythms, and psychomotor vigilance”. In: *Clinics in Sports Medicine* 24.2, pp. 237–49, vii–viii.

Van Dongen, Hans P A, Greg Maislin, Janet M Mullington, and David F Dinges (Mar. 2003). “The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation.” In: *Sleep* 26.2, pp. 117–26.

Vgontzas, A N, E Zoumakis, E O Bixler, H-M Lin, H Follett, A Kales, and G P Chrousos (May 2004). “Adverse effects of modest sleep restriction on sleepiness, performance, and inflammatory cytokines”. In: *The Journal of Clinical Endocrinology and Metabolism* 89.5, pp. 2119–26.

Zivin, Joshua Graff and Matthew Neidell (Dec. 2012). “The Impact of Pollution on Worker Productivity”. In: *American Economic Review* 102.7, pp. 3652–3673.

3.8 Supplementary material

3.8.1 Background and summary statistics

Table 3.12: Causal medical studies of sleep and performance

Study	Sleep change (hr/day)	Study duration (days)	Outcome	Elasticities (abs. value)
Belenky et al. (2003)	-4, -2, -1, +1	7	PVT speed	.7, .5, .7, 0
Cohen et al. (2010)	-2.5	21	PVT reaction time	18
Dinges et al. (1997)	-2.4	7	PVT lapses	6
Landrigan et al. (2004)	+.82	21	Serious medical errors	4.5
Lockley et al. (2004)	+.82	21	Attention failures	4
Van Dongen et al. (2003)	-4, -2	14	Memory task	3.3, 2.2
Vgontzas et al. (2004)	-2	7	PVT lapses	2.9
Mean magnitude				3.9

Table includes all studies that experimentally manipulated sleep over at least 7 days, drawing on reviews by Van Dongen and Dinges (2005) and Banks and Dinges (2007). Studies of complete sleep deprivation were excluded. PVT stands for psycho-motor vigilance test, described in Section 3.2.1.

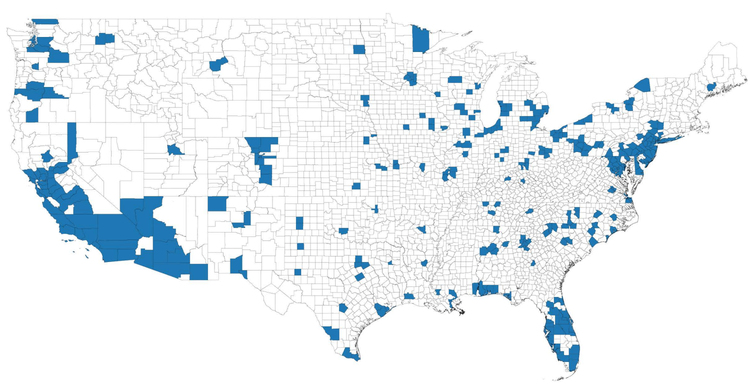


Figure 3.3: ATUS county-level geocoding

The map shows, in blue, locations in the continental United States where we are able to geocode ATUS records at the county level.

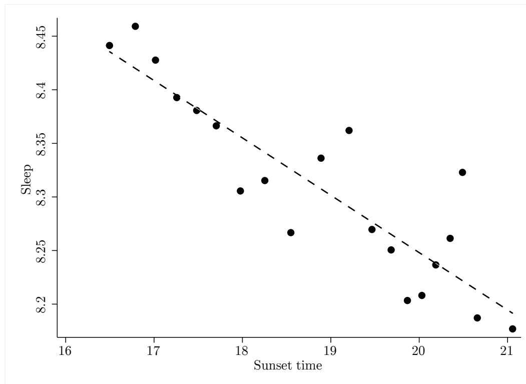


Figure 3.4: ATUS raw correlation: sleep and sunset time

Created using binscatter. Sunset time and sleep are divided into 20 equal-width bins and means are computed within each. These means provide x and y coordinates. Fitted line (dashed) estimated using OLS. Sample is the estimation sample from Table 3.5.

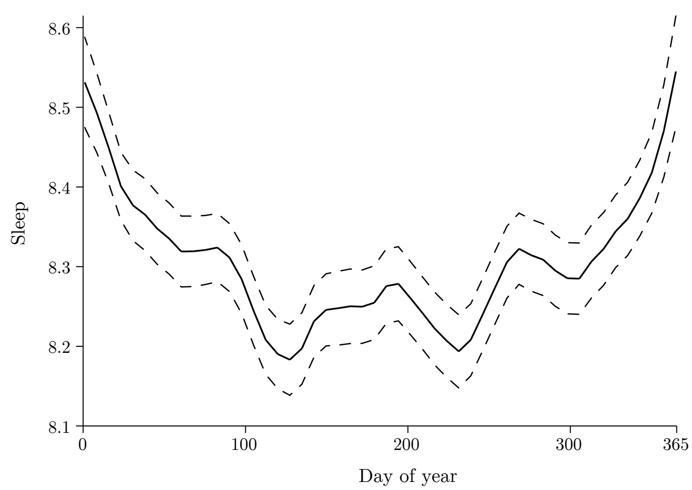
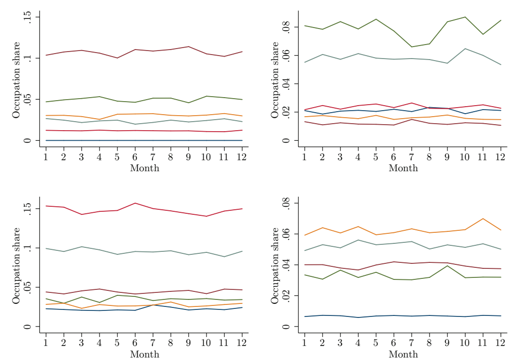
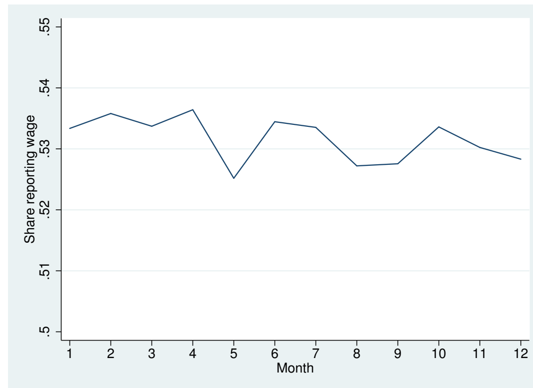


Figure 3.5: Sleep seasonality

The figure shows a local polynomial fit to sleep data from ATUS. Calculations use a bandwidth of 10 days and an Epanechnikov kernel. Note that the range of sleep in the sample is 2 to 16 hours and the standard deviation is 2.03.



(a) ATUS occupation shares by month



(b) ATUS share reporting positive wage by month

Figure 3.6: ATUS occupations do not exhibit seasonality

Panel (a) Lines show average values of our 23 occupation dummies by month, pooled over the period 2003-2013 for our estimation sample. The occupation exhibiting a modest summer dip in the upper-right panel is "Arts, design, entertainment, sports, and media occupations." Excluding this occupation does not change our results. Panel (b) Line shows average value of dummy that equals 1 if the respondent reports a non-zero weekly or hourly wage, by month, pooled over the period 2003-2013 for all ATUS respondents.

Table 3.13: QCEW summary statistics

Variable	**Mean**	**Std. Dev.**
Weekly wage	492.37	171.88
Weekly wage - goods	609.35	240.53
Weekly wage - services	431.84	161.19
Sunset time	18.38	.94
Observations	285,680	

All data are from the Quarterly Census of Employment and Wages at the county level from 1990-2013.

3.8.2 Solar mechanics

Here, we provide a brief summary of how sunset time is calculated and a glossary of terms. We calculate sunset, sunrise, solar declination, and sunlight duration each day using the algorithm of Meeus (1991) as implemented by NOAA's Earth Systems Research Laboratory (ESRL). The calculator takes inputs of the date, time zone offset, latitude, and longitude. The Stata code that we used for calculation is available upon request.

Sunset and sunrise time are both calculated assuming 0.833° of atmospheric refraction, or the bending of the path of light as it passes through the Earth's atmosphere. In practice a refraction correction would need to incorporate information on air pressure and humidity. Also, we calculate sunset assuming an observer with a 0 elevation change view of the horizon. Over a full county, this assumption should introduce minimal error.

Sunlight duration is simply calculated as the difference between sunrise and sunset time for a location on a given day.

Solar declination is the angle of a line segment from the sun to the earth relative to a plane projected from the equator of the Earth. The solar declination is a function only of the day of year and time zone offset (to compute fractional days for high-resolution local time sunset), and changes in solar declination correspond to the seasonal movement of the sun. The highest solar declination, 23.44° occurs on the summer solstice, and the lowest solar declination, -23.44°, occurs on the winter solstice. On the equinox, solar declination is 0°. A rough calculation of solar declination can be made with the following equation:

$$-23.44\cos\left(\frac{360}{365}(d+10)\right)$$

where d is the day of the year. For a detailed glossary, see NOAA's ESRL website.

3.8.3 Measurement error in seasonal estimate

Here we derive expressions for the expected bias in our seasonal estimates reported in Section 3.5. The informal version of this derivation is discussed in Section 3.3.3.

Assume that for a given individual i surveyed on day t, wages are equal to the average of D past sleep observations plus random noise. Thus the true model relating sleep to wages is

$$w_{it,\tau} = \beta\left(D_i^{-1}\sum_{k=\tau-T}^{\tau-1} T_{S,ik}\right) + \varepsilon_{i\tau} \tag{3.5}$$

$$= \beta T^{*}_{S,i\tau} + \varepsilon_{i\tau} \tag{3.6}$$

Thus, we are assuming that earnings change for this individual every D days, and sleep only matters during the earnings determination period. The subscript τ indexes the day that these earnings start to be *observed* in the data. Because of the fixed earnings change frequency for a given individual, these earnings will be observed for days τ through $D+\tau-1$. To be concrete, consider the case of $D=2$. Then we, the researchers, can only sample the individual the day after they received an earnings change or 2 days after, so τ will either be equal to t or $t-1$.

We further assume that τ is uniformly distributed across the year (a person has an equal probability of receiving an earnings change on any given day). This is a strong assumption, but the best available evidence from Barattieri et al. (2010) suggests that it is not broadly incorrect. Of course, for a given year, there will be weekend or holiday effects, but asyptotically, these become less relevant. Moreover, we do not have any information on when a given individual in our sample last experienced an earnings change, so this uniform assumption is a relevant baseline.

Finally, we assume that the researcher has isolated exogenous variation in sleep so that $\mathrm{E}[T_{S,t}\varepsilon_\tau] = 0$ for all t and τ.

If we observed past sleep and knew the earnings change frequency, we could estimate Equation (3.5) and return the correct estimate. Instead, We observe wages and sleep on date $t \geq \tau$, with which we estimate

$$w_{it,\tau} = \beta_1 T_{S,it} + \varepsilon_{i\tau}$$

We wish to know the relationship between β_1 and β.

We will exploit the wage setting structure given above and the functional form for the time series of sunset time from Section 3.8.2 to calculate this relationship. First, given the results from Frazis and Stewart (2012) we can use sunset time to both isolate daily, exogenous variation in sleep and to predict daily sleep for any day of the year, even though we only observe sleep on one day. This individual time series of sleep will have a similar functional form to the instrument, namely

$$T_{S,it} = A\cos(\theta t)$$

where A is the population coefficient on the unconditional version of the first stage of Equation (3.1) and where we drop an ignorable, uncorrelated error term. The value $\theta = 360/365$ scales the wavelength to one year, so we make an additional simplification by assuming that a year is 360 days long so that this term can be ignored. Alternatively, one could, as we do when we analytically calculate the bias, rescale t to incorporate the term. Thus

$$T_{S,it} = A\cos(t) = A\cos(\tau + j) \tag{3.7}$$

where $j = t - \tau$ is the number of days since the latest earnings change for this observation.

We apply Lagrange's identity to rewrite earnings-relevant sleep.

$$T^*_{S,i\tau} = D_i^{-1}\sum_{k=1}^{D_i} A\cos(\tau) = \frac{A}{D_i 2\sin(1/2)}(\cos(\tau - D_i + (\pi-1)/2) - \cos(\tau + (\pi-1)/2))$$

The two cosine functions are simply phase shifts of each other, so we apply phasor addition to reduce this to

$$T^*_{S,i\tau} = AB_1\cos(\tau+\omega) \tag{3.8}$$

where

$$B_1^2 = \frac{(\cos((\pi-1)/2 - D_i) + \cos((\pi-1)/2))^2 + (\sin((\pi-1)/2 - D_i) + \sin((\pi-1)/2))^2}{2D_i\sin(1/2)}$$

$$\omega = \arctan(\cot((D_i+1)/2))$$

This form is convenient because observed sleep can now be written as a phase shift of earnings-relevant sleep and thus suggests that a version of the error-in-variables formula will apply in this setting since by another application of phasor addition, we can linearly relate observed sleep to earnings-relevant sleep plus a correlated error term.

Now note that the only individual heterogeneity is in terms of the frequency of earnings changes, so without loss of generality, we can replace the D_i index with just D. Then, applying the usual variance-covariance formula for the OLS estimator of a single coefficient, we have that our estimator relative to the true coefficient is given by[6]

$$\hat{\beta}_{1,D} = \frac{\mathrm{Cov}(w_D, T_{S,D})}{\mathrm{Var}(T_{S,D})} \tag{3.9}$$

$$= \beta\frac{\mathrm{Cov}(T^*_{S,D}, T_{S,D})}{\mathrm{Var}(T_{S,D})} \tag{3.10}$$

[6] We loosely call this value attenuation even though in practice the estimate can be negative even when the true coefficient is positive.

Where we have dropped the time subscripts based on the calculations below.

To derive closed-form expressions for Equation (3.10) observe that since seasonal sleep is mean zero, and, for all D, we are equally likely to observe sleep on any day of the year, then by the double angle formula and Lagrange's identity, the denominator is

$$\begin{aligned}\mathrm{Var}(T_{S,D}) &= \lim_{T\to\infty} T^{-1}\sum_{t=0}^{T} A^2\cos^2(t) \\ &= \frac{A^2}{2} + \lim_{T\to\infty}\frac{A^2\csc(1)\sin(2T+1)+3}{T} = \frac{A^2}{2}\end{aligned}$$

Where T (not to be confused with T_S) is the total number of time observations and the last equality follows from the boundedness of sine.

The numerator is, by application of the product-to-sum and Lagrange identities

$$\begin{aligned}\mathrm{Cov}(T^*_{S,D}, T_{S,D}) &= D^{-2}\sum_{k=1}^{D}\sum_{j=1}^{D}\lim_{T\to\infty} T^{-1}\sum_{\tau=0}^{T} A^2\cos(\tau-k)\cos(\tau+j) \\ &= D^{-2}\sum_{k=1}^{D}\sum_{j=0}^{D}\lim_{T\to\infty} T^{-1}\sum_{\tau=0}^{T} A^2\cos(\tau-k)\cos(\tau+j) \\ &= \frac{A^2}{2D^2}\sum_{k=1}^{D}\sum_{j=0}^{D}\cos(k+j)\end{aligned}$$

Taking the ratio of these two values gives the relative bias of the seasonal estimate with respect to the true estimate for a given D.

$$\frac{\hat{\beta}_{1,D}}{\beta} \xrightarrow{p} D^{-2}\sum_{k=1}^{D}\sum_{j=0}^{D-1}\cos(k+j) \tag{3.11}$$

Figure 3.7 shows this value for all frequencies of earnings changes less than a year. This shape is the result of two factors. First, for any two of the same sinosoidal functions that are phase shifted from each other by less than a quarter or more than three-quarters of a wavelength, the product will be positive because the two functions are "in phase

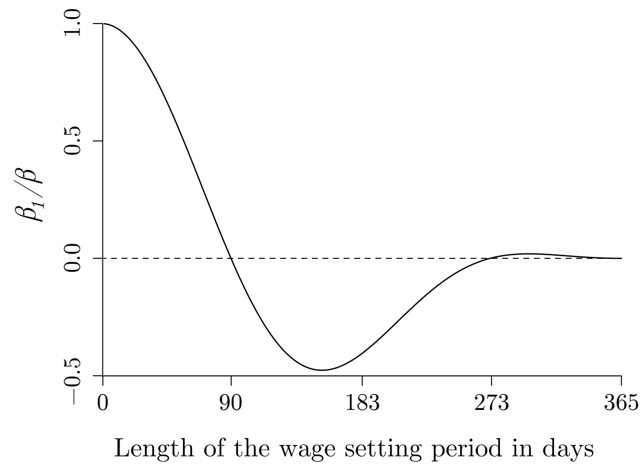


Figure 3.7: Seasonal estimate versus true estimate

The figure shows the ratio of the probability limit of the seasonal estimate to the true estimate on the y-axis for a range of possible frequencies of earnings changes on the x-axis. A value of 1 on the y-axis indicates no bias, while a negative value indicates that the estimated coefficient has the wrong sign.

enough". For a phase shift greater than one-quarter but less than three-quarters of a wavelength, the product will be negative. The attenuation of the estimate is an average of these products, so for frequent earnings changes (small D), we are largely averaging sleep that is less than a quarter wavelength off from the truth. For intermediate values of D, we are averaging in sleep that is phase shifted enough to flip the sign on the estimate. For D near a year, however, we have "crossed the hump" again and are averaging in sleep values that are phase shifted so much that they are back to the beginning of the cosine wave. Beyond $D = 365$, the estimate remains nearly fully attenuated, with slight oscillations around 0.

The expected attenuation for the full population will depend, therefore, only on the distribution of the frequency of earnings changes.

$$\hat{\beta}_1/\beta \xrightarrow{p} \sum_D D^{-2} \sum_{k=1}^{D} \sum_{j=0}^{D-1} \cos(k+j)\,\mathrm{Pr}(D) \tag{3.12}$$

Barattieri et al. (2010) provide estimates of this density function (derived from Figures 12 and 13), with which we can calculate Equation (3.12). We discuss the measures provided by these authors in Section 3.3.3, ultimately concluding that the degree of attenuation will be between 75 to 100%.

Finally, a note on alternative assumptions about the earnings determination process (Equation (3.5)). If earnings are based more on recent earnings rather than historical earnings (for instance if the manager is myopic when writing wage contracts), then our estimate will be closer to the true coefficient because we will be more likely to average together observed sleep that is less than a quarter wavelength phase shifted from the truth. If earnings are based on longer-term sleep or productivity patterns (for instance, the manager is very slow to update the wage contract and needs two earnings change cycles to fully incorporate current productivity changes), then our estimate will either be more biased or will be more likely to be attenuated all the way to zero. In the limit, as the manager or earnings process only takes into account productivity signals from the beginning of the worker's career, our estimate will be zero with probability one.

3.8.4 Auxiliary results and robustness checks

3.8.5 ATUS robustness checks

Tables 3.14 and 3.15 repeat our preferred specification for the sample of workers who report being paid an hourly wage. Estimates are much less precise. The estimates in Table 3.14 are quite close to our primary estimates. In principle the coefficients are not directly comparable, since the change in weekly wage could include both wage and hour effects. Because our instrument induces very small changes in hours worked (roughly half an hour per week), however, the two tables represent roughly the same change.

In Table 3.15, the two-stage estimate of .0013 is within one standard error of our preferred estimate of .00077. The imprecision in these estimates means that size control

Table 3.14: Short-run effects: Hourly workers

	First stage	Reduced form	2SLS	OLS
	Sleep	ln(wage)	ln(wage)	ln(wage)
Seasonal sunset time	-24.6***	-0.0025*		
	(3.15)	(0.0014)		
Sleep			0.00010*	-0.000016***
			(0.000060)	(0.0000020)
Individual controls	Yes	Yes	Yes	Yes
Time controls	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes
FIPS FEs	Yes	Yes	Yes	Yes
Observations	43927	43927	43927	41991
F-stat on IV			62.58	
Elasticity			0.72	

The table shows results from estimating Equation (3.1). The first three columns show the first stage, reduced form, and two-stage least squares estimates. The fourth column reports the uninstrumented version of the second stage of Equation (3.1). The dependent variable is indicated at the top of each column. Earnings refers to "usual weekly earnings". Controls include: location fixed effects; race dummies; age; age squared; a full-time dummy; a gender dummy; dummies for holiday, day of week, and year; and occupation dummies. Standard errors, clustered at the FIPS code (location) level, are reported in parentheses. Significance indicated by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

may be an important problem for the two-stage estimate. According to Stock and Yogo (2002), with a first-stage F of approximately 4, the actual size of the 5% t test on this coefficient could be greater than 25%.

Per the recommendation in Solon et al. (2013), we conduct a modified Breusch-Pagan test for heteroskedasticity of the residuals from the unweighted 2SLS model. The results in Table 3.16 show that location-level observations with smaller underlying counts of ATUS observations exhibit higher variance, as expected, and the relationship is statistically significant at the one percent level. While the constant term is statistically significant, it is an order of magnitude smaller. This suggests that the common error component within location is minimal, so weighting will likely result in an efficiency improvement, and indeed that is what we see in Table 3.17.

Table 3.17 reproduces our long-run results from Table 3.5 above their unweighted counterparts. Weighting does indeed improve efficiency, reducing the standard errors in

Table 3.15: Long-run effects: Hourly workers

	First stage Sleep	Reduced form ln(wage)	2SLS ln(wage)	OLS ln(wage)
Avg. sunset time	-48.6**	-0.064***		
	(24.4)	(0.021)		
Sleep			0.0013*	0.000032
			(0.00074)	(0.000033)
Geographic controls	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Observations	529	529	529	529
Adjusted R^2	0.119	0.576	.	0.564
F-stat on IV	3.98			
Elasticity			4.39	

The table shows results from estimating Equation (3.2), with location-level observations weighted by the count of underlying ATUS respondents. The first three columns show the first stage, reduced form, and two-stage least squares estimates. The fourth column reports the uninstrumented version of the second stage of Equation (3.2). The dependent variable is indicated at the top of each column. Wage refers to hourly wage for those workers who report being paid hourly. Controls include: coastal distance, a 10-piece linear spline in latitude; share full time; median age; race shares; occupation shares; and a 5-piece linear spline in population density. White robust standard errors reported in parentheses. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 3.16: Modified Breusch-Pagan heteroskedasticity test

	Residuals^2
1/Observations	0.15***
	(0.030)
Constant	0.011***
	(0.0014)
Observations	529
Adjusted R^2	0.040

The dependent variable is the squared residual from estimating the unweighted version of (3.2). The variable "1/Observations" is the reciprocal of the number of ATUS interviews underlying a given location-level observation. Because the modified Breusch-Pagan test relies on the assumption of homokurtosis, we compute unmodified OLS standard errors. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

both the first stage and reduced form models. It does not appreciably alter the first-stage coefficient on sleep, but it does increase the magnitude of the reduced-form estimate. This suggests the presence of heterogeneity in the marginal effect of sleep on wages. High-skill urban workers have greater influence on the estimates in the weighted model, so the pattern of results below is consistent with larger marginal effects for such workers.

Table 3.17: Long-run effects, weighted and unweighted

Panel A: Weighted			
	First stage	Reduced form	2SLS
	Sleep	ln(earnings)	ln(earnings)
Avg. sunset time	-64.0***	-0.049***	
	(16.2)	(0.019)	
Sleep			0.00077**
			(0.00033)
Observations	529	529	529
Adjusted R^2	0.137	0.809	0.718
F-stat on IV	15.52		
Panel B: Unweighted			
	First stage	Reduced form	2SLS
	Sleep	ln(earnings)	ln(earnings)
Avg. sunset time	-55.6**	-0.0083	
	(26.0)	(0.022)	
Sleep			0.00015
			(0.00038)
Observations	529	529	529
Adjusted R^2	0.237	0.650	0.652
F-stat on IV	4.57		

The table shows results from estimating Equation (3.2). In Panel A location-level observations are weighted by the count of underlying ATUS respondents, while in Panel B they are unweighted. The dependent variable is indicated at the top of each column. Earnings refers to "usual weekly earnings". Controls are as reported below Table 3.5. White robust standard errors reported in parentheses. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

We have also performed a variety of additional robustness checks with little or no change in estimates. We list them here without full tables, but all results are available upon request. 0.2% of the sample has topcoded wages. A tobit accounting for this does not change the results. Likewise, accounting for the truncation of sleep does not change inference. We have also estimated the models on only the sub-sample that is geocoded at the county or CBSA level. All of these robustness checks do not change inference.

In Table 3.18 we report estimates of county level characteristics as functions of average sunset time. We find a large and statistically significant relationship with population density, which motivates our use of a flexible control for this variable in estimating long-run effects. We also see a statistically significant relationship with unemployment, consistent with our estimated long-run wage effect.

Table 3.18: Robustness: County characteristics

	Log pop. density	Pop. change frac.	Net migration frac.
Sunset time	-0.642***	-0.000931	-0.000980*
	(0.110)	(0.000814)	(0.000533)
Observations	3104	3104	3104
Adjusted R^2	0.012	0.000	0.001

	Log poverty rate	Labor force change	Unemployment rate
Sunset time	0.0157	0.00184	-1.412***
	(0.0221)	(0.00342)	(0.169)
Observations	3103	3103	3103
Adjusted R^2	-0.000	-0.000	0.023

Dependent variable is indicated at the top of each column. All data are from the Census and observations are at the county level. Population, net migration, and unemployment rate are all 2012 values. Poverty is from 2011. Labor force change is from 2000 to 2010. White heteroskedasticity-robust standard errors are reported in parentheses. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Historical sorting

Figure 3.8 shows the county-level growth patterns around the dates of the 1883 and 1918 time zone implementations. For both figures, the 10% of counties that are closest to the eastern or western time zone boundary are considered to be on the eastern or western side, respectively. The dashed lines show median population growth rates (inter-census) for eastern side counties, and the solid lines show the same for western side counties. The composition of these groups differs between the two panels due to changes in the location of the 1883 versus 1918 time zones.

If gross sorting were occurring, one would expect eastern side counties to grow faster than western side counties after time zone implementation. Indeed, one might even expect the incentive to sort with respect to the 1883 time zones to be stronger than in

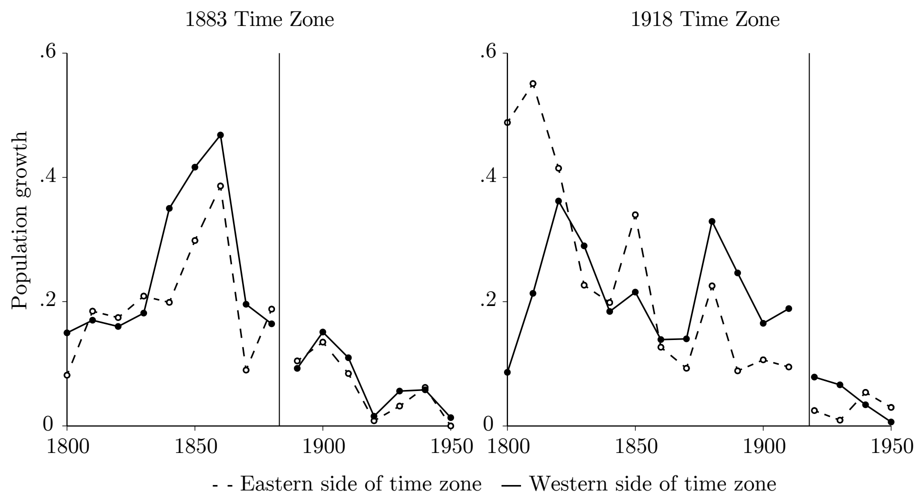


Figure 3.8: Historical time zone sorting

The figure shows median growth rates between censuses in counties on the eastern and western edges of the 1883 (left panel) and 1918 (right panel) time zones. Eastern counties are represented by the dashed line and western counties are the solid line. All data are from Haines and Inter-university Consortium for Political and Social Research (2010).

the present day due to the lack of electrification. Instead, one can see that there is no evidence of gross sorting in response to the 1883 time zone. After implementation, the two regions of the time zones grow at almost identical rates. Growth rates around the 1918 law are more volatile but tell a similar story. Western side counties experience a slightly larger drop in growth rates after 1918 compared to eastern side counties, but the difference in changes between the two groups is not significant.

Hedonic and QCEW robustness checks

In Table 3.19 we show the robustness of our hedonic result to additional controls, including industry shares, a spline in coastal distance, and longitude.

In Table 3.20 we show the robustness of our QCEW reduced-form result (all industries) to additional controls, again including industry shares, a spline in coastal distance, and longitude.

Table 3.19: Hedonic robustness

	Log value	Log value	Log value	Log value
Sunset time	-0.0851***	-0.0811***	-0.0745***	-0.0828***
	(0.0192)	(0.0153)	(0.0190)	(0.0190)
Base controls	Yes	Yes	Yes	Yes
Industry shares	No	Yes	No	No
Coastal distance spline	No	No	Yes	No
Longitude	No	No	No	Yes
Observations	2824	2824	2824	2824
Adjusted R^2	0.600	0.760	0.619	0.603

White heteroskedasticity-robust standard errors are reported in parentheses. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1. Data are 2010 5-year ACS estimates. Sunset time is the average for a given county. Column 1 reproduces the final column of our preferred results from Table 3.8 and “Base controls” denotes the controls from that model. Additional controls employed in this table are a 3-piece linear spline in coastal distance, 13 industry shares, and longitude.

Table 3.20: QCEW robustness

	Log avg. wage	Log avg. wage	Log avg. wage	Log avg. wage
Sunset time	-0.0263**	-0.0317***	-0.0259**	-0.0246**
	(0.0120)	(0.0111)	(0.0120)	(0.0121)
Base controls	Yes	Yes	Yes	Yes
Industry shares	No	Yes	No	No
Coastal dist. spline	No	No	Yes	No
Longitude	No	No	No	Yes
Observations	291271	291271	291271	291271

Standard errors are reported in parentheses. Clustering is at the county level. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1. Data are from the BLS Quarterly Census of Employment and Wages 1990-2013. Sunset time is the quarterly county average. Column 1 reproduces the final column of our preferred results from Table 3.9 and “Base controls” denotes the controls from that model. Additional controls employed in this table are a 3-piece linear spline in coastal distance, 13 industry shares, and longitude.