

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Essays on Economic Incentives in Environmental Policy

Permalink

<https://escholarship.org/uc/item/3b67t6wr>

Author

Globus-Harris, Isla Rose

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays on Economic Incentives in Environmental Policy

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

in

Economics with a Specialization in Interdisciplinary Environmental Research

by

Isla Rose Globus-Harris

Committee in charge:

Professor Joel Watson, Chair
Professor Richard T. Carson
Professor Mark R. Jacobsen
Professor Lisa A. Levin
Professor Craig McIntosh

2017

Copyright
Isla Rose Globus-Harris, 2017
All rights reserved.

The Dissertation of Isla Rose Globus-Harris is approved and is acceptable
in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017

TABLE OF CONTENTS

Signature Page	iii
Table of Contents	iv
List of Figures	vi
List of Tables	viii
Acknowledgements	ix
Vita	x
Abstract of the Dissertation	xi
Chapter 1 An Impossible Goal: When Trade Ratios Can't Achieve No-Net-Loss	1
1.1 Introduction	1
1.1.1 Carbon Offsets	4
1.1.2 Wetland Mitigation	6
1.1.3 Water Quality	7
1.1.4 Additional Examples	7
1.2 Model	9
1.3 Example: Linear Supply and Demand	13
1.4 Existence of Neutral Policies	17
1.5 Efficiency and Trade Ratios	21
1.6 Dynamic Updating of Trade Ratios	23
1.7 Policy Discussion	28
1.7.1 Monitoring	28
1.7.2 Segmenting Supply	29
1.8 Conclusion	30
1.9 Appendices	31
1.9.1 Additional Proofs	31
1.9.2 Efficiency	33
1.9.3 Multiple Trade Ratios	35
1.9.4 Microeconomic Theory Foundation for Model Assumptions ...	41
Chapter 2 Using Waiting Periods to Screen Participants for Voluntary Green Actions	47
2.1 Introduction	48
2.2 Model	52
2.3 Time Delays as a Screening Tool	55
2.4 Optimal Time Delay Mechanisms	62
2.4.1 Case 1: $\delta \geq d$	62

2.4.2	Case 2: $\delta < d$	72
2.5	Policy Discussion	77
2.6	Conclusion	78
2.7	Appendices	79
2.7.1	Proof of Proposition 2.4	79
2.7.2	Optimal Waiting Periods for the Uniform Distribution	83
Chapter 3	Estimating Additionality in Energy-Efficiency Subsidy Programs ...	85
3.1	Introduction	85
3.2	Theoretical Model	88
3.2.1	Limitations and Data Requirements for the Identification Technique	94
3.3	Data	96
3.4	Empirical Model	100
3.4.1	Potential Threats to Identification	105
3.5	Results	107
3.5.1	Additionality Results	110
3.6	Conclusion	113
3.7	Appendix	115
3.7.1	Full regression Results	115
	Bibliography	117

LIST OF FIGURES

Figure 1.1.	How supply shifts as N varies	12
Figure 1.2.	N^* is a neutral trade ratio	13
Figure 1.3.	Linear supply and demand, N^* is the neutral trade ratio	15
Figure 1.4.	Linear supply and demand, no neutral trade ratio exists	16
Figure 1.5.	Linear supply and demand, naïve vs. neutral trade ratios	17
Figure 1.6.	Example where there are two neutral trade ratios, N_1^* and N_2^*	23
Figure 1.7.	Neutral trade ratios exist, but the updating procedure does not converge. $N_0 = 1$	25
Figure 1.8.	A neutral trade ratio exists, but the updating procedure does not converge and $N_k \not\rightarrow \infty$, with $N_0 = 1$	26
Figure 1.9.	The updating procedure converges to N_3^* , the third lowest neutral trade ratio, with $N_0 = 1$	27
Figure 1.10.	A market with three supply sectors	41
Figure 2.1.	Outside option for non-additional agents	57
Figure 2.2.	Outside option for additional agents	58
Figure 2.3.	Optimal waiting period as δ varies	76
Figure 3.1.	Participation over time with a cutoff cost of 2	90
Figure 3.2.	Participation over time with an initial cutoff cost of 2, followed by a cutoff cost of 5 after time period 10	91
Figure 3.3.	Participation over time with an initial cutoff cost of 2, followed by a cutoff cost of 0.5 after time period 10.....	92
Figure 3.4.	Participation over time with an initial cutoff cost of 2, followed by a cutoff cost of 0.5 after time period 10, and a return to a cutoff cost of 2 in time period 15	93
Figure 3.5.	Illustration of the identification technique.....	95
Figure 3.6.	Average amount of incentives for heat pump programs over time .	98

Figure 3.7.	Total participation in all heat pump programs	101
Figure 3.8.	Participation in ducted heat pump programs	102
Figure 3.9.	Total participation in ductless pump programs	103
Figure 3.10.	Projected participation (in the absence of a program gap) and actual participation	111
Figure 3.11.	Confidence bands on projected participation become very large over time	113
Figure 3.12.	Results of the convergence test.....	114

LIST OF TABLES

Table 3.1.	Heat Pump Programs	97
Table 3.2.	Participants in Heat Pump Programs	98
Table 3.3.	Average daily energy usage for participants and non-participants ..	99
Table 3.4.	Regression results for the first estimation equation	108
Table 3.5.	Regression results for the second and third estimation equations ..	109
Table 3.6.	Estimated number of participants in Area C and estimated fraction of non-additional participants as T_{end} varies	112
Table 3.7.	Full regression results for the first estimation equation	115
Table 3.8.	Regression results for the second estimation equation.	116

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Joel Watson for his support as the chair of my committee. His suggestions and guidance were paramount in the writing of this dissertation. I would also like to acknowledge Professor Richard Carson and Professor Mark Jacobsen for their support and guidance.

VITA

2010	Bachelor of Arts, University of Oregon
2017	Doctor of Philosophy, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Essays on Economic Incentives in Environmental Policy

by

Isla Rose Globus-Harris

Doctor of Philosophy in Economics with a Specialization in Interdisciplinary
Environmental Research

University of California, San Diego, 2017

Professor Joel Watson, Chair

This dissertation explores environmental economics, focusing on the interaction between incentives and environmental policies. I use microeconomic theory, mechanism design, and empirical economic techniques to examine the importance of considering incentives and adverse selection when formulating environmental policy. Chapter 1 uses microeconomic theory to demonstrate the importance of considering adverse selection and perverse incentives when formulating environmental policy, with a particular focus on environmental trade ratios. Chapter 2 use mechanism design to demonstrate how heterogeneous outside options can be exploited to enhance screening in an environmental

subsidy program setting. Chapter 3 uses both economic theory and empirics to measure the extent of an adverse selection problem in an environmental subsidy program.

Chapter 1

An Impossible Goal: When Trade Ratios Can't Achieve No-Net-Loss

Abstract: I develop a model of environmental trade ratio policies with adverse selection to explain why current trade ratio policies may fail to achieve their goal of no-net-impact on the environment, as well as demonstrating that it is sometimes impossible to achieve this goal. Environmental trade ratio policies, such as carbon offset discounting and wetland mitigation ratios, typically have the goal of no-net environmental impact; however, these goals are not always met. For example, the US is experiencing a net loss of wetlands despite its national goal of no-net-loss of wetlands and despite the extensive use of wetland mitigation ratios. I provide two possible explanations for these policy failures. First, I demonstrate that naïve methods for setting environmental trade ratios—which do not fully consider supply and demand responses—result in ratios that almost always fail to meet no-net-loss goals. Second, I demonstrate that, contrary to what previous literature has assumed, it is sometimes impossible for trade ratios to achieve no-net-impact on the environment due to an unraveling caused by adverse selection. Finally, I demonstrate how a simple updating procedure can adjust trade ratio policies towards no-net-impact even when policy makers have no information about supply and demand.

1.1 Introduction

Policy makers use trade ratios in environmental markets with adverse selection. The goal is typically environmental neutrality, meaning that the policy has no net impact

on the environment. Examples include carbon offset markets, wetland mitigation, water quality trading, and reforestation programs. However, these programs don't always attain their policy goals.¹ This paper will provide two possible explanations for why these policies may fail. First, I will demonstrate that naïve methods for setting trade ratios—which do not fully consider supply and demand responses—result in ratios that almost always fail to achieve no-net-impact on the environment. Second, I will demonstrate that, contrary to assumptions in previous literature,² it is sometimes impossible to use trade ratios to achieve no-net-impact due to an unraveling caused by adverse selection.

Trade ratio policies are related to cap and trade systems. As in cap and trade, to engage in an environmentally damaging activity—such as deforestation or destruction of a wetland—firms must have a permit. These permits can be obtained from an existing cap and trade system or created by environmental good suppliers. Suppliers get permits by creating an environmental good, e.g., planting trees or reducing carbon emissions.

However, not all suppliers provide equally valuable goods. The policy maker cannot distinguish suppliers of differing quality ex-ante. To compensate for adverse selection, policy makers use trade ratios. Trade ratios require multiple units of the environmental good to create one new permit. For instance, a carbon offset trade ratio of 3:1 requires three tons of carbon offsets to create one ton of carbon permits.

There are many reasons why suppliers might have varying quality. In the case of carbon offsets, some suppliers are additional—engaging in new abatement—while others are non-additional—engaging in abatement they would have done in the absence of the offset program. For wetland mitigation, it could be the quality of the wetland, or whether

¹The best-studied example is the US no-net-loss of wetlands policy. Wetland mitigation ratios are widely used in the US to achieve this policy goal. However, much of the evidence suggests that the US is still experiencing a net loss of wetlands. For example, see Turner et al. [2001] which finds that mitigation is failing to achieve no-net-loss measured both by acres and function, and Dahl and Stedman [2013] which finds that US is losing an average of 80,000 acres of coastal wetlands per year. Empirical evidence on the success or failure of other (non-wetland) trade ratio policies is more limited.

²See, for example, Van Benthem and Kerr [2013] and Bento et al. [2014].

or not the wetland project was completed. In reforestation projects, some suppliers may have lower tree survival rates.

As an example, consider a simplified version of wetland mitigation markets. The United States has a stated policy goal of no-net-loss of wetlands. If a developer wants to build a shopping mall on an existing wetland, they need to first acquire the appropriate permits. To get these permits, they must demonstrate that a new wetland is built in a different location. Shopping mall developers are not experienced at building wetlands, so they instead buy credits from a wetland mitigation bank. However, not all wetland mitigation banks have the same quality. Perhaps one mitigation bank has a project failure rate of 30%, while another has a failure rate of 70%. The policy maker cannot distinguish between the different qualities of mitigation banks. This may be for a variety of reasons, such as high monitoring costs or time lags between project completion and wetland failure. To compensate for both the uncertain wetland quality and the fact that neither supplier has a 100% success rate, the regulator uses a trade ratio. For instance, the regulator might require the wetland mitigation bank to build 2 acres of wetland to create a permit that the developer can use to destroy 1 acre of wetland.

In addition to compensating for varying quality, policy makers also use trade ratios to account for temporal lags (National Research Council [2001]). For instance, it may take many years for trees to reach maturity, or for a wetland to provide certain ecosystem services. Additionally, policy makers sometimes use trade ratios to compensate for risk or uncertainty, such as measurement uncertainty (Shishlov and Bellassen [2016]).

To the best of my knowledge, no other paper attempts to model environmental trade ratios in a general manner. There are a variety of papers that deal with trade ratios in the context of specific policies—primarily carbon offset policies.³ None of the existing literature acknowledges that neutral trade ratios do not always exist. For example,

³See, for example, Bento et al. [2014], Van Benthem and Kerr [2013], and Førsund and Nævdal [1998].

Van Benthem and Kerr [2013] analyze carbon offset trade ratios. While their paper does not focus explicitly on the existence of neutral ratios, they claim that a high enough trade ratio will ensure neutrality. I will demonstrate that this is not necessarily the case, particularly if demand is elastic, as they assume. Bento et al. [2014] also examine carbon offset trade ratios. They assume perfectly elastic demand, leading them to conclude that trade ratios greater than one always crowd out good suppliers. I use a more general model to demonstrate that this is not always true.

In this paper, I develop a generalized model of trade ratios to demonstrate that they cannot always achieve environmental neutrality and to analyze the impacts of market responses on neutral trade ratios. I also analyze the efficiency of neutral trade ratios, and propose an updating procedure for approximating neutral trade ratios.

1.1.1 Carbon Offsets

Some carbon offset markets use trade ratios to compensate for non-additional offsets. An additional offset is one that represents new abatement, while a non-additional offset is for abatement that would have occurred anyways in the absence of the offset.⁴ Non-additional offsets also include abatement of extra emissions that the firm creates solely to sell as an offset.⁵ Because additionality is measured relative to an unobserved counterfactual, it is difficult to establish additionality, even ex-post.

One of the largest offset programs is the Clean Development Mechanism (CDM), tied to the Kyoto Protocol. The CDM implicitly incorporates a 95% discount rate (1.05:1 ratio) to account for measurement uncertainty for some offsets. While not all CDM

⁴For example, consider a deforestation-prevention project that pays a forest owner to not cut down trees. The offset is additional if the owner would have cut down the trees in the absence of the offset. It is non-additional if the forest owner wouldn't have cut down the trees anyways. The policy maker is unable to observe how the forest owner would act in the absence of the offset program.

⁵This was a problem for offsets from the destruction of HFC-23, a refrigerant and potent greenhouse gas. Firms created additional, unneeded HFC-23 so they could destroy it and get an offset (Schneider [2011]).

measurement methodologies incorporate this ratio, the top 10 methodologies—which are used in more than 80% of CDM projects and at least 33% of credits—incorporate it (Shishlov and Bellassen [2016]). It is extremely difficult to get exact estimates for the proportion of CDM projects that are in fact additional. Zhang and Wang [2011] study Chinese sulfur dioxide emissions and find no impact of CDM participation on overall pollution levels, suggesting that many projects are non-additional. Michaelowa and Purohit [2007] study 19 Indian CDM projects in detail and find that five have doubtful additionality. Wara and Victor [2008] examine hydropower projects in China, almost all of which apply for CDM credits, and conclude that many are likely non-additional. Although there isn't enough evidence to conclusively determine how many CDM credits are non-additional, the existing evidence strongly suggests that the rate is high. Because the implicit trade ratio is so small, it is extremely likely that the CDM is not meeting neutrality.

The low number of carbon cap and trade systems is currently a limiting factor in the use of carbon offset trade ratios. However, there is at least one carbon offset program with an explicit trade ratio: the Alberta Offset System in Canada, which uses trade ratios from 0 percent (1:1 ratio) to 30 percent (1.43:1 ratio) (Alberta Agriculture and Forestry [2011]). There are no empirical estimates of the additionality of these offsets, so it is unclear whether or not the program is reaching neutrality.

Policy makers have proposed additional carbon offset trade ratio programs. These include the 2009 Waxman-Markey bill (H.R. 2454) in the United States which called for a 1.25:1 ratio for international offsets, and a call to add explicit trade ratios to the CDM by Korean delegates to the United Nations Framework Convention on Climate Change (Chung [2007]).

1.1.2 Wetland Mitigation

The US has a stated goal of no-net-loss of wetlands. Section 404 of the Clean Water Act regulates filling, dredging, or polluting wetlands, and requires that any impacts be offset by compensatory mitigation. To achieve this, many US wetland mitigation policies use trade ratios. Ratios vary between regions, and are often specific to certain types of wetlands—e.g., tidal marshes use a different ratio than prairie pothole marshes. Ratios range from 1:1 to 24:1 or higher (Washington Department of Ecology [2006]). Government agencies—such as state departments of environmental quality—typically set these ratios.

Despite these trade ratios, the majority of the wetlands literature agrees that the US is not meeting its no-net-loss of wetlands goal. Brown and Lant [1999] examine wetland mitigation ratios and find that while some banks successfully used ratios to achieve no-net-loss by acreage, several other large mitigation banks failed to achieve no-net-loss. They do not examine no-net-loss by function. Turner et al. [2001] conduct a meta-analysis of the wetland mitigation literature, and find that wetland mitigation has failed to achieve no-net-loss overall. They find that the loss comes from a variety of failures consistent with bad suppliers, including failure to initiate permitted projects, failure to complete projects, and projects with low wetland functionality. Kihlslinger [2008] updates Turner et al. [2001], and finds that the goal of no-net-loss of wetlands was still not being met, both by wetland acreage and function.

Many current wetland mitigation ratios are “naïve,” in the sense that they do not account for the selection bias that occurs when trade ratios are changed.⁶ As I will demonstrate in this paper, naïve trade ratios do not typically achieve neutrality.

⁶For an example of how wetland mitigation trade ratios are chosen, see Robb [2002].

1.1.3 Water Quality

The US EPA suggests that states use trading ratios in their water quality trading markets (US EPA [2009]). The EPA categorizes water quality trading ratios into different categories, including: delivery and location ratios, used for trades between pollution at different locations such as an upstream and a downstream source; equivalency ratios, used for trades between different forms of pollutant such as a form of phosphorus that is more bioavailable than another; and uncertainty ratios, used mainly to account for uncertainty in point source to non-point source trades (US EPA [2009]).

The actual trade ratios used vary by state, watershed, pollutant, and other factors. For example, the Virginia Department of Environmental Quality uses a 2:1 trade ratio for trades between point source and non-point source pollution in its Chesapeake Bay Watershed Nutrient Credit Exchange Program (Virginia Department of Environmental Quality [2008]). The Connecticut Department of Environmental Quality uses a varying trade ratio for trades between pollution in different locations in its Nitrogen Credit Exchange program in Long Island Sound (Connecticut Department of Environmental Protection [2010]).

Due to a lack of empirical data, it is unclear if these policies are achieving neutrality.

1.1.4 Additional Examples

The California Department of Fish and Game's policy on western burrowing owl burrows incorporates trade ratios. Western burrowing owls do not build their own burrows, but rather rely on existing burrows left behind by other animals. Development projects, including solar farms, often destroy these burrows. As a mitigation measure, the policy requires that for each owl burrow destroyed, two new artificial burrows must be constructed on protected lands (2:1 ratio) (California Department of Fish and Game

[1995]).

Many US states have regulations on how many trees must be replanted after logging. These numbers are not necessarily the same as the number of trees harvested, and often depend on acreage instead of the number of logged trees (Ellefson et al. [1995]). Although this is not explicitly a trade ratio, it functions in much the same way.

There are many additional policies that are well-situated to use trade ratios. Certain air pollution trading programs currently prohibit trades between geographical regions. For example, Southern California's Regional Clean Air Incentives Market bans trade between downwind and upwind sources (Schmalensee and Stavins [2015]). Instead of banning trade, policy makers could incorporate trade ratios between the differentiated regions.

There are also calls to use compensatory mitigation, a type of trade ratio policy, for fisheries bycatch. Certain species of marine tetrapods—including marine mammals, sea turtles, and seabirds—are caught (and often die) as fisheries bycatch. In recent years, there have been costly fisheries closures when fisheries have reached their bycatch caps and the fishing season was subsequently closed.⁷ Wilcox and Donlan [2007] propose that it would be more cost effective to use compensatory mitigation, where fisheries pay to reduce non-fishing threats to the bycatch species, such as removing invasive rodents on seabird breeding islands. Because the payments are structured such that the payment for one bycatch animal translates into more than one animal saved from invasive rodents, these policies function as a type of trade ratio.

As environmental markets and cap and trade programs become more common, policy makers may turn to trade ratios to achieve neutrality. It is important for them to understand how to properly set neutral trade ratios, what the efficiency implications of

⁷These closures include the Hawaiian longline swordfish fishery from 2001 to 2004, and the New Zealand squid fishery in 2000 and 2002, among others (Gjertsen et al. [2014] and Hodgson [2002]).

neutral trade ratios are, and when neutrality is an impossible goal.

The remainder of this paper is organized as follows. Section 1.2 presents a generalized model of environmental trade ratios. Section 1.3 uses a linear example to demonstrate the main results. Section 1.4 discusses the biases of improperly set ratios and the existence of neutral trade ratio policies. Section 1.5 analyzes the efficiency of trade ratios. Section 1.6 presents an updating procedure for approximating neutral trade ratios. Section 1.7 summarizes key policy implications, and Section 1.8 concludes. Appendix 1.9.1 contains additional proofs, 1.9.2 contains additional efficiency results, and Appendix 1.9.3 examines the case where the policy maker can set different trade ratios for different sectors of supply.

1.2 Model

I model supply and demand for permits in a market where a policy maker fixes a trade ratio for translating units between supply and demand. The trade ratio is the policy maker's only choice variable. The policy goal is environmental neutrality.⁸

The consumers of permits are subject to a cap on the initial quantity of permits. This covers a variety of policies, such as cap and trade policies for carbon dioxide or sulfur dioxide, or policies where the permits allow the consumer to engage in an otherwise forbidden activity like destroying a wetland. I assume the cap is exogenous and weakly positive. Note that if instead the cap is a choice variable, it can be adjusted to compensate for the adverse selection, and trade ratios are not needed (Montero [2000]). The exogenous cap captures the case where the cap is inflexible for political reasons,⁹ the case where a trade ratio is added to an existing cap and trade system, and the case

⁸In 1.9.2 I consider the case where the policy goal is to maximize social welfare.

⁹For example, political lobbying or the voluntary nature of international agreements may make it infeasible to lower the cap.

where the cap is too low to be adjusted downwards.¹⁰ I abstract away from the cap itself, and instead simply consider demand for permits.¹¹ I assume that demand is determined by demand function $D(\cdot) \geq 0$, which is differentiable with $D'(\cdot) < 0$.

The producers voluntarily produce goods in exchange for newly created permits, which they may sell to consumers. There are good and bad suppliers. Good suppliers create the true environmental good, while bad suppliers create a good with no environmental benefits. The policy maker cannot distinguish between the two types of suppliers. Total supply is $S(\cdot) + X(\cdot)$, where $S(\cdot) \geq 0$ is good supply, and $X(\cdot) > 0$ is bad supply.¹² Both S and X are differentiable with $S'(\cdot) \geq 0$, and $X'(\cdot) \geq 0$.¹³ I assume that $S(0) = 0$ to capture the fact that to be a good supplier, they must incur some costs.

The separation of good and bad suppliers does not necessarily mean that some suppliers sell 100% good permits while others sell 100% bad permits. Nothing in this model precludes a continuum of suppliers along a spectrum from all good to all bad. Consider a wetland supplier who builds a 10 acre wetland. They are moderately skilled at creating and maintaining wetlands, but 1 acre of the wetland fails shortly after the completion of the project. In this case, I categorize 9 acres of the wetlands as good supply and 1 acre as bad supply.

I assume that $\epsilon_S(\cdot) > \epsilon_X(\cdot)$ everywhere,¹⁴ where $\epsilon_S(\cdot)$ is the point elasticity of good supply and $\epsilon_X(\cdot)$ is the point elasticity of bad supply. This condition captures the fact that bad suppliers are less price sensitive and that, on average, they are willing to sell their good for lower prices than are good suppliers. For example, in the case of

¹⁰For example, the US policy on wetland loss can be interpreted as a cap of zero, leaving no room to lower it further.

¹¹Although I do not explicitly model the demand function as depending on the cap, changes in the cap should change the demand function. For simplicity I hold the cap constant and therefore do not treat the cap as a variable in the demand function.

¹²This ensures that there is always a strictly positive quantity supplied in the market.

¹³Later assumptions on price elasticities will imply that $S'(\cdot) > 0$.

¹⁴Except where ϵ_S is undefined at $S(0) = 0$.

carbon offsets, bad (non-additional) suppliers are not changing their behavior, and will therefore sell for prices close to zero.¹⁵ Good (additional) suppliers are making costly changes and thus will only sell for a sufficiently high price. In wetland mitigation, there is evidence that lower cost suppliers create wetlands with lower functionality and higher failure rates (King and Bohlen [1994a] and King and Bohlen [1994b]). Additionally, if the true environmental good is less expensive to produce than the fake one, then suppliers would have no incentive to create a fake good.

Let $r(\cdot) = \frac{S(\cdot)}{S(\cdot) + X(\cdot)}$, the fraction of supply that is good. The assumptions on S and X are sufficient to ensure that r exists and is continuous.¹⁶ Lemma 1.1 demonstrates how the elasticity assumption implies that as price increases, the fraction of supply that is good also increases.

Lemma 1.1. $\varepsilon_S(\cdot) > \varepsilon_X(\cdot)$ everywhere if and only if the fraction of the market that is good is everywhere increasing in its argument, i.e., $\frac{dr(\cdot)}{d\cdot} > 0$.

Proof. See Appendix 1.9.1. □

The policy maker selects a trade ratio $N \in [1, \infty)$.¹⁷ The trade ratio transforms N units of supply-side good into 1 demand-side permit. This means that consumers and producers do not face the same price. P denotes the price of permits, i.e., the price faced by consumers. It follows that the price faced by producers is P/N . P is determined by the market equilibrium condition, $D(P) = \frac{1}{N} [S(P/N) + X(P/N)]$. For the remainder of this paper, I treat P as a function of N as determined by the market equilibrium condition.¹⁸

Figure 1.1 demonstrates how supply shifts relative to demand as N changes.

¹⁵ Assuming there are no strategic incentives.

¹⁶ If the model instead started with a function r and attempted to derive the supply functions, they would not be uniquely defined. For this reason my model uses the supply and demand functions as the primitives.

¹⁷ Because $r(\cdot) < 1$, $N < 1$ would never achieve neutrality.

¹⁸ By the implicit function theorem, I must also assume that $D'(P) \neq \frac{1}{N^2} S'(\frac{P}{N}) + \frac{1}{N^2} X'(\frac{P}{N})$. This ensures that the market price as a function of N exists and is differentiable.

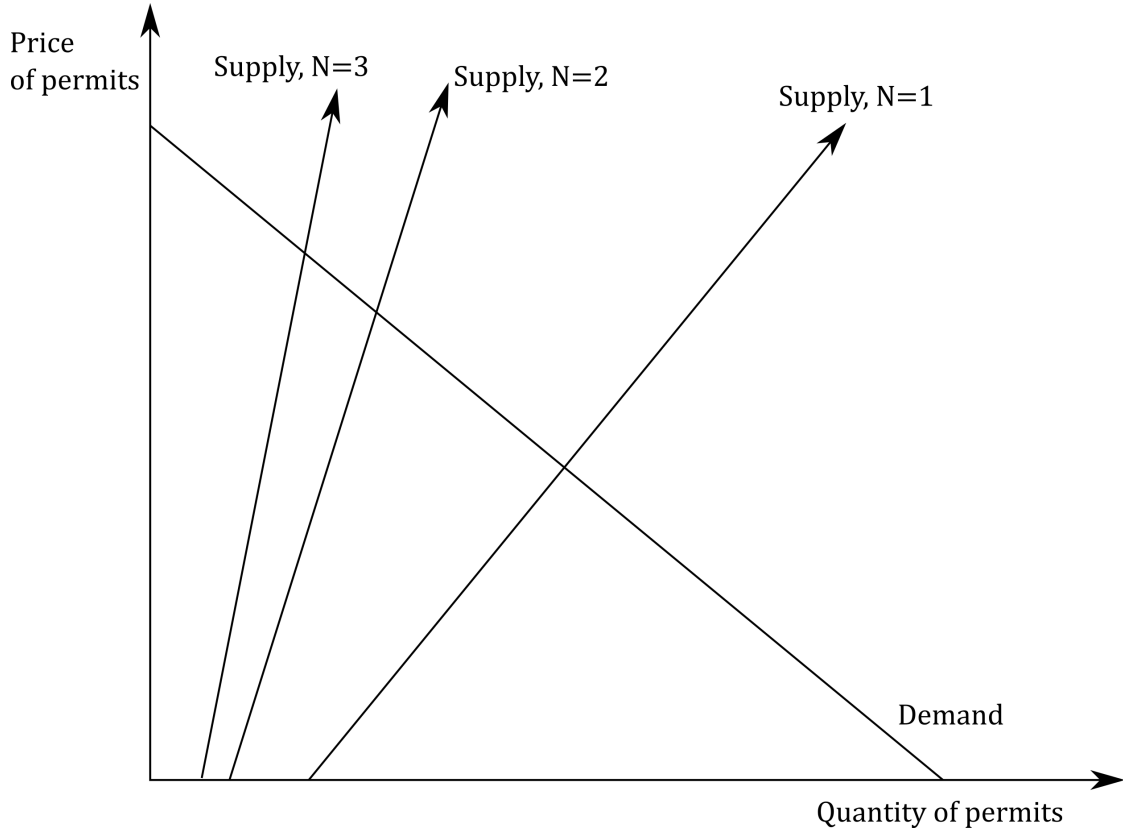


Figure 1.1. How supply shifts as N varies

The policy maker's objective when setting the trade ratio is environmental neutrality. Neutrality is defined as meeting the exogenous cap. This is equivalent to the environmental benefits when no trades are allowed between supply and demand. Mathematically, this means that $D(P(N)) = S(\frac{P(N)}{N})$, i.e., the demand for permits equals the good supply. This is equivalent to $r(\frac{P(N)}{N}) = \frac{1}{N}$. Figure 1.2 demonstrates neutrality by graphing r as a function of the trade ratio and comparing it to the function $\frac{1}{N}$. Throughout this paper, I will use N^* to refer to neutral trade ratios.

Note that Appendix 1.9.4 analyzes a more complex model—with a continuum of supplier types and monitoring costs—and demonstrates how the simplified model presented here can be generated by more nuanced underlying model.

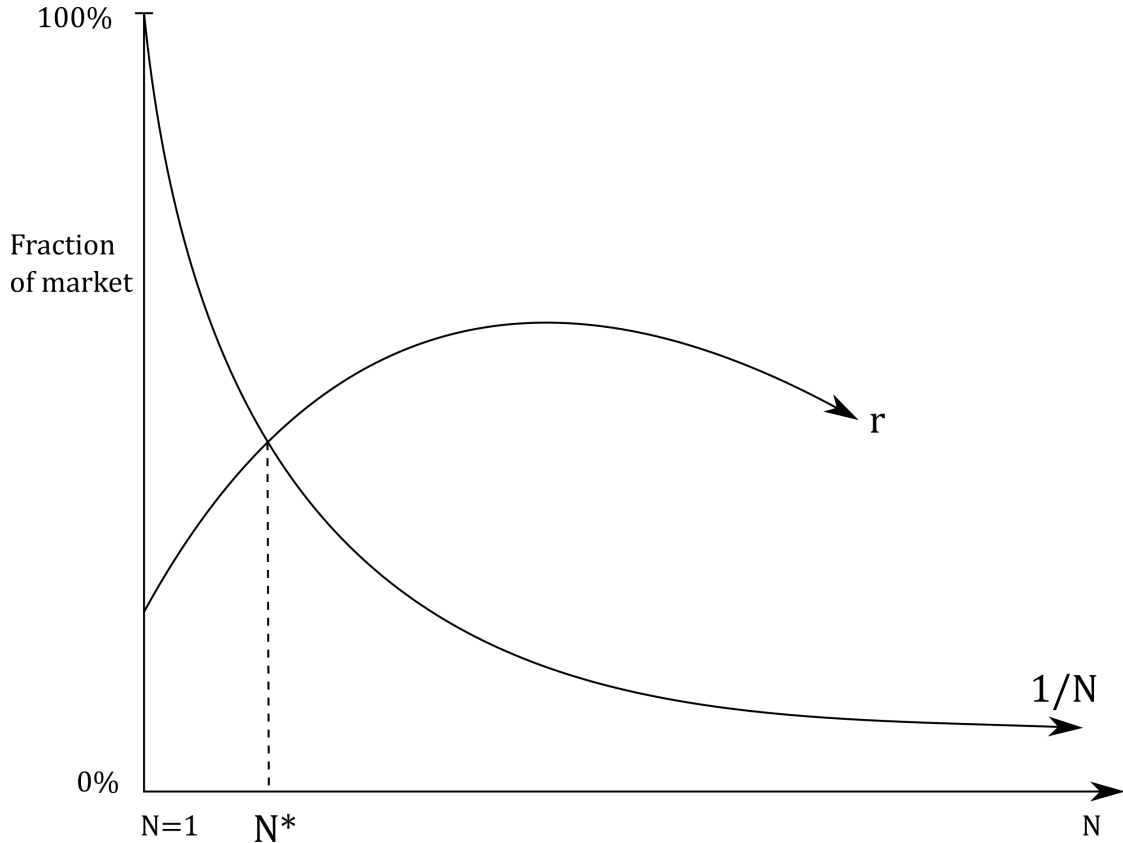


Figure 1.2. N^* is a neutral trade ratio

1.3 Example: Linear Supply and Demand

The following example provides intuition for how trade ratio policies work, and intuition for when they are able to meet neutrality goals. For simplicity, I assume linear supply and demand for the duration of this example.

For this example, demand is $D(P) = a - bP$ where $a, b > 0$. Good supply is $S(\frac{P}{N}) = c\frac{P}{N}$, and bad supply is $X(\frac{P}{N}) = X$ where $c, X > 0$. Note that this satisfies all of the required assumptions, including that the price elasticity of good supply is everywhere strictly greater than the price elasticity of bad supply (which is zero here).

Price is determined by the market equilibrium condition, $D(P) = \frac{1}{N} \left[S(\frac{P}{N}) + X(\frac{P}{N}) \right]$. Using these functional forms, market equilibrium price can be written as a

function of the trade ratio, $P(N) = \frac{aN-X}{bN+c/N}$.¹⁹

A neutral trade ratio is one that achieves no-net-impact on the environment, measured relative to not allowing any trades between supply and demand. Mathematically, this can be written as $D(P(N)) = S(\frac{P(N)}{N})$ or equivalently as $\frac{1}{N} = r(\frac{P(N)}{N})$, where r is the fraction of supply that is good, i.e., $r(\cdot) = \frac{S(\cdot)}{S(\cdot)+X(\cdot)}$. Using the functional forms for this example, solving for the neutral trade ratio, N^* , yields $N^* = \frac{ac+cX}{ac-bX}$. Recall that $N \in [1, \infty)$. However, the formula for N^* above does not always yield a positive trade ratio.²⁰ This means that the neutral trade ratio is not defined for all parameter values. Figures 1.3 and 1.4 graph the function r for two linear examples with different parameter values. Figure 1.3 illustrates an example where the neutral trade ratio exists, and Figure 1.4 illustrates an example where the neutral ratio does not exist.

So what is happening for the parameter values with no neutral trade ratio? For these cases, it is helpful to examine how the market and prices respond as the policy maker increases the trade ratio. Starting at a trade ratio $N = 1$, the market is not at neutrality. Specifically, there is less of the true environmental good than is needed for neutrality. As the policy maker increases the trade ratio, the supply-side price can either decrease or increase.²¹ If the supply-side price is decreasing, good suppliers will drop out of the market at a faster rate than bad suppliers due to the elasticity assumption. If the rate is fast enough, the ratio of good to bad supply can worsen so quickly that the trade ratio can't keep up. Thus in these circumstances, raising the trade ratio causes the fraction of supply that is good to worsen, leaving the market short of neutrality. Section 1.4 discusses the existence of trade ratios in a more general setting.

Policy makers frequently set trade ratios naïvely. This means that instead of fully

¹⁹Technically, $P(N) = \max\left[\frac{aN-X}{bN+c/N}, 0\right]$. However, for simplicity I exclude the parameter values for which the zero lower bound is binding.

²⁰Note that if $N^* > 0$, it will also be greater than 1.

²¹See Proposition 1.1 for detailed results on the direction of the price change.

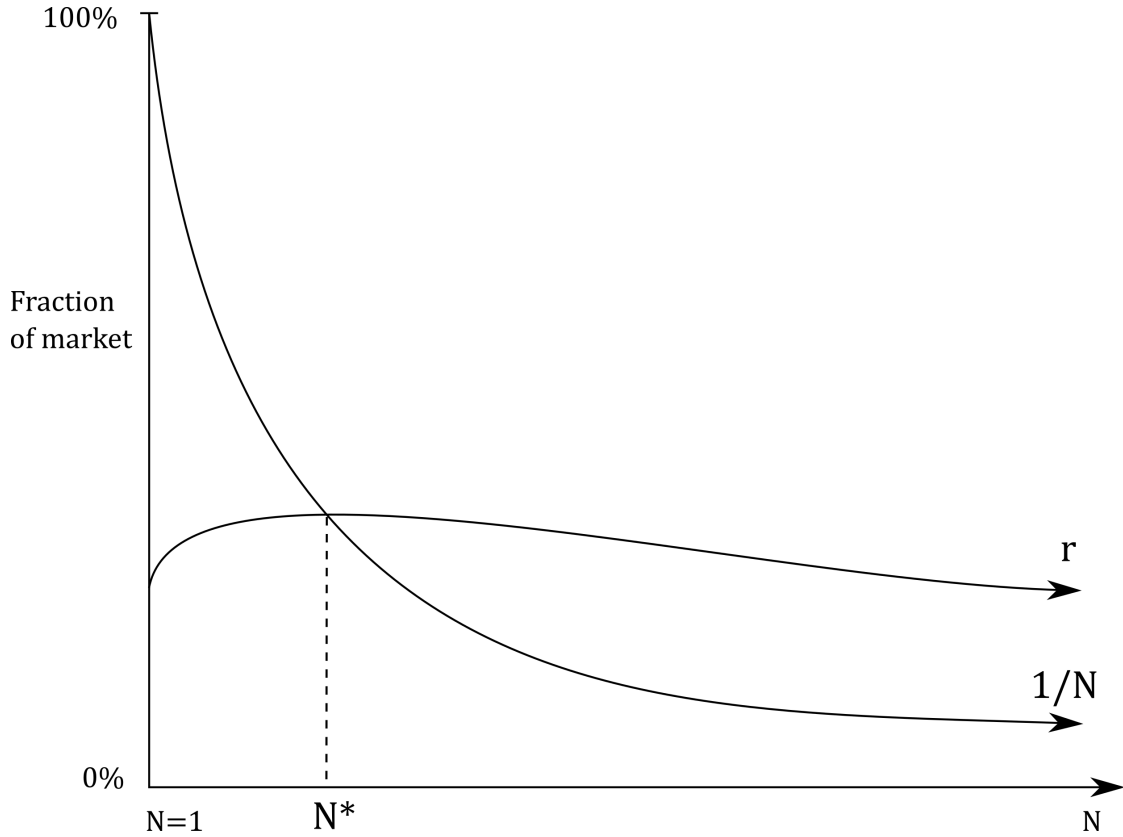


Figure 1.3. Linear supply and demand, N^* is the neutral trade ratio

considering the interactions between the trade ratio, supply, and demand, they simply observe the current fraction of supply that is good and invert it.²² So if they estimate that half of suppliers are good, they would set a trade ratio of two. However, this will rarely result in a neutral trade ratio as it does not consider how the market will react to the new trade ratio. Consider the linear example, starting at a trade ratio $N = 1$.²³ At $N = 1$, r , the fraction of supply that is good, equals $\frac{ac-cX}{ac+bX}$.²⁴ Thus the naïve trade ratio would be the inverse of this, $N^{\text{naïve}} = \frac{ac+bX}{ac-cX}$. This differs slightly from the neutral trade ratio, $N^* = \frac{ac+cX}{ac-bX}$, meaning that the naïve trade ratio will typically not lead to neutrality. Also note that the naïve trade ratio and the neutral trade ratio exist for different parameter

²²See, for example, Robb [2002] for an overview of how many wetland mitigation ratios are chosen.

²³Naïve trade ratios are sensitive to the starting trade ratio.

²⁴Technically it equals $\max\left[\frac{ac-cX}{ac+bX}, 0\right]$ because for some parameter values the entire market may be bad supply. I ignore these parameter values for simplicity.

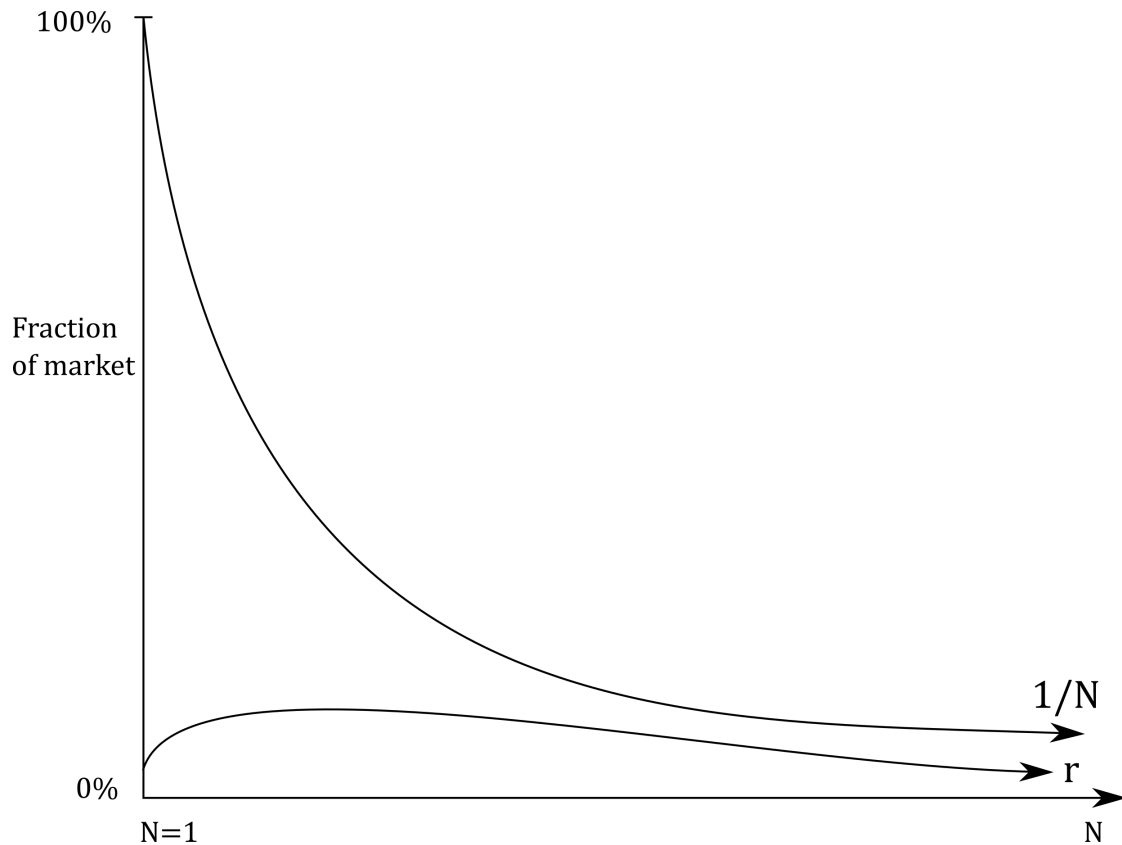


Figure 1.4. Linear supply and demand, no neutral trade ratio exists

values.

Figure 1.5 illustrates an example of how the naïve and neutral trade ratios might differ. In this particular example, $N^{\text{naïve}} > N^*$ and there is an excessive amount of the environmental good relative to neutrality. It is also possible to have a shortage of the environmental good when the trade ratio is too high—if r dips back under $1/N$ between the neutral and naïve ratio. Lastly, if r initially slopes downwards, then the naïve trade ratio will be too low and there will be a shortage of the environmental good relative to neutrality. Section 1.6 discusses naïve trade ratios in more detail.

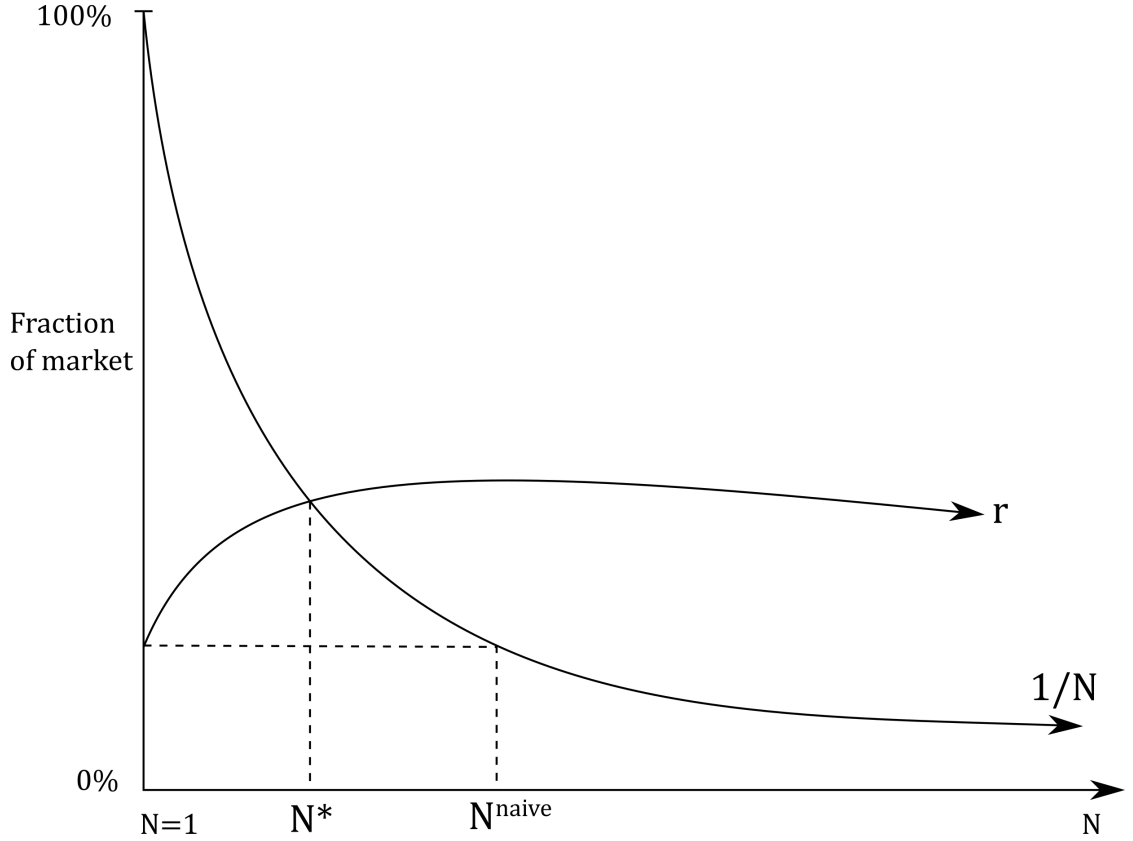


Figure 1.5. Linear supply and demand, naïve vs. neutral trade ratios

1.4 Existence of Neutral Policies

The trade ratio, N , impacts equilibrium price for both consumers and producers. As N increases, the demand-side price $P(N)$ increases.²⁵ However, the direction of the change for the supply-side price $P(N)/N$ is ambiguous. Proposition 1.1 specifies conditions under which $P(N)/N$ is decreasing or increasing in N . For the remainder of this paper, I use the term elastic to refer to demand elasticities less than -1 , and inelastic to refer to demand elasticities between 0 and -1 .

Proposition 1.1. $\frac{dP(N)/N}{dN} < 0$ if and only if demand is elastic at $P(N)$.

Proof. See Appendix 1.9.1. □

²⁵For proof, see Lemma 1.2 in Appendix 1.9.1.

Good and bad suppliers respond to prices changes differently. As demonstrated by Corollary 1.1, Proposition 1.1 has important policy implications. Specifically, a marginal increase in the trade ratio will crowd out good suppliers if demand is elastic. Similarly, a marginal increase in the trade ratio will encourage good suppliers to enter the market if demand is inelastic. Proposition 1.1 and Corollary 1.1 explain how and why naïvely set trade ratios—which do not consider supply and demand responses—typically fail to achieve neutrality. If demand is inelastic, the naïve ratio will be greater than the neutral ratio. If demand is elastic, the naïve ratio will be less than the neutral ratio.²⁶

Corollary 1.1. *The fraction of the market that is good is increasing in N if and only if demand is inelastic at $P(N)$. In other words, $\frac{dr(P(N)/N)}{dN} > 0$ if and only if demand is inelastic at $P(N)$.*

Proof. Lemma 1.1 shows that $\frac{dr(\cdot)}{d\cdot} > 0$. Further, Proposition 1.1 shows that $\frac{dP(N)/N}{dN} > 0$ if and only if demand is inelastic at $P(N)$. It follows that $\frac{dr(P(N)/N)}{dN} > 0$ if and only if demand is inelastic at $P(N)$. $\square$

Corollary 1.1 provides intuition for when a neutral trade ratio can be guaranteed. Recall that neutrality is equivalent to $r(P(N)/N) = \frac{1}{N}$. When demand is inelastic, increasing the trade ratio encourages good suppliers to enter the market, increasing r . Proposition 1.2 formalizes the intuition that if $P(N)/N$ is increasing in N , then a neutral trade ratio must exist.

Proposition 1.2. *If demand is inelastic everywhere, then there exists a trade ratio N such that $r(P(N)/N) = \frac{1}{N}$.*

Proof. Corollary 1.1 demonstrates that if demand is inelastic, then $r(P(N)/N)$ is increasing in N . Note that $r(P(N)/N)$ is a continuous function of N , and that at $N = 1$,

²⁶In this case, the neutral ratio may not even exist.

$r(P(1)) \in (0, 1]$. As N increases, the fraction of the market that is credited decreases, with $\lim_{N \rightarrow \infty} \frac{1}{N} = 0$. Note that this is also continuous. It follows that at some point the two functions must cross. Thus there must exist a neutral trade ratio N . $\square$

Conversely, if demand is elastic everywhere, then $P(N)/N$ is decreasing in N . This does not necessarily mean that a neutral ratio does not exist. As N increases, a smaller and smaller fraction of supply is converted into permits. Non-existence of a neutral trade ratio only occurs when good suppliers drop out of the market fast enough that the trade ratio can't keep up.

Proposition 1.2 is a sufficient—not necessary—condition. Further, it requires that conditions are satisfied everywhere. While this provides helpful intuition, it is not easy to test with empirical data. Policy makers do not know the true supply and demand functions. The following results clarify to what extent existence of a neutral trade ratio can be tested with empirical data and minimal assumptions about the global traits of the supply and demand functions.

Proposition 1.3. *If there exists an N such that $r(P(N)/N) \geq \frac{1}{N}$, then there exists N such that $r(P(N)/N) = \frac{1}{N}$.*

Proof. Recall that at $N = 1$, $r(P(1)) < \frac{1}{N}$, and that r is continuous. If there exists an N such that $r(P(N)/N) \geq \frac{1}{N}$, then it must be the case that at some point $r(P(N)/N)$ crosses $\frac{1}{N}$ and thus an environmentally neutral trade ratio exists. $\square$

Proposition 1.3 demonstrates that we only need to observe a single point to verify the existence of a neutral trade ratio. However, determining when a neutral ratio does not exist requires much more information. Without observing information about r at all trade ratios, we cannot guarantee that a neutral trade ratio does not exist.

I will start by formalizing what I mean by observing data. I assume that the policy maker observes ordered pairs (N, r) . I do not specify how many pairs they observe,

and allow observing an uncountably infinite number of ordered pairs. I assume that the observed ordered pairs are consistent with at least one continuous function $r(\cdot)$.²⁷

Proposition 1.3 specifies under what conditions the policy maker can guarantee existence of a neutral trade ratio: whenever there exists a pair such that $r \geq \frac{1}{N}$. The following result demonstrates that you can never guarantee non-existence without observing pairs of (N, r) almost everywhere.

Proposition 1.4. *Assume the policy maker does not observe any values of r for at least one range (N_1, N_2) where $N_1 < N_2$. If the policy maker does not observe any points such that $r \geq \frac{1}{N}$, then they cannot say whether or not a neutral trade ratio exists.*

Proof. The policy maker observes pairs of (N, r) such that $r < \frac{1}{N}$ for all r and N . I will use the observed points to construct two hypothetical functions, r_1 and r_2 .

The function r_1 is constructed as follows. Let $N' \in (N_1, N_2)$. Now consider the set of ordered pairs that contains all observed ordered pairs, and also the point (N', r') where $r' > \frac{1}{N'}$. Because N' lies in the open set (N_1, N_2) , it must be the case that (N', r') is consistent with some continuous function r_1 . Thus if r_1 is the true function for r , a neutral trade ratio exists.

The function r_2 is constructed as follows. For all values of $N'' \in (N_1, N_2)$, let the corresponding $r'' = \frac{1}{N''} - \varepsilon$ for some $\varepsilon > 0$ such that $\frac{1}{N''} - \varepsilon > 0$ and that is consistent with r_2 being continuous. This is well defined by the separating hyperplane theorem. Thus if r_2 is the true function for r , a neutral trade ratio does not exist.

With the data available to the policy maker, they cannot distinguish between functions r_1 and r_2 . Thus they cannot determine whether or not a neutral trade ratio exists. □

²⁷Further, I assume that if there are certain unobserved points such that only one point is consistent with r being continuous, that the policy maker infers this and considers that extra point observed.

Propositions 1.3 and 3.3 illuminate how much the existence of neutral trade ratios can be examined with empirical data. Natural experiments use shifts in supply or demand to measure the curve between those shifts, but nothing is known about the curve outside of that range. This may provide enough information to establish the existence of a neutral ratio. However, additional assumptions would be needed to find sufficient conditions for non-existence.²⁸

Proposition 3.3 also explains why searching for conditions that are both necessary and sufficient for existence of neutral trade ratios is futile. Any such condition must rely on the entirety of the function r . However, if r is fully known, then the policy maker already knows whether or not neutral trade ratios exists.

1.5 Efficiency and Trade Ratios

Policy makers often strive for environmental neutrality. However, neutrality is not necessarily efficient. Additionally, when neutrality is not possible, policy makers may seek an alternative. Analyzing the efficiency of trade ratios offers some guidance to policy makers.

First, why do policy makers focus on environmental neutrality rather than efficiency? Proposition 1.5 offers a possible explanation: a neutral trade ratio guarantees an efficiency improvement over not allowing any trades. While a neutral trade ratio may not be fully efficient, it is better than not having a trading policy.

Proposition 1.5. *A neutral trade ratio policy is a Pareto improvement upon not allowing any trades between supply and demand.*

Proof. Environmental neutrality ensures that the program has the same environmental

²⁸There also may be empirical difficulties differentiating between good and bad supply. In some cases, this distinction may be clear ex-post, like the ability to observe that a wetland project had high vegetation mortality rates. In other cases, observation may be significantly more difficult, such as additionality concerns in carbon offset markets.

benefits as not allowing any trades between the sectors. All trades are voluntary, thus they must make the participants better off. It follows that the neutral trade ratio is a Pareto improvement upon not allowing any trades. $\square$

Thus a neutral trade ratio makes everyone weakly better off when compared to not allowing trades. The efficiency implications of a non-neutral program, however, are ambiguous. Trade between the two sectors will result in cost savings, but it is unclear how this compares to the change in environmental benefits. For some non-neutral trade ratios, the damage to the environment will outweigh the cost savings.

It is possible that no neutral trade ratio exists. It is also possible that multiple neutral trade ratios exist: Figure 1.6 illustrates such a case. When choosing between multiple neutral trade ratios, which one is most efficient? Proposition 1.6 shows that lowest neutral trade ratio is most efficient.

Proposition 1.6. *If $N_1 < \dots < N_k < \dots < N_K$ are neutral trade ratios, then N_1 is a Pareto improvement upon N_k for $k \in \{2, \dots, K\}$.*

Proof. Let $k \in \{2, \dots, K\}$. N_1 and N_k are both environmentally neutral, thus they have the same environmental benefits.

Recall that $\frac{dP(N)}{dN} > 0$. It follows that $P(N_1) < P(N_k)$. Because consumers face a lower price under N_1 , they must be better off with trade ratio N_1 than with N_k .

If $N_1 < N_k$ and both are neutral, by the definition of neutrality it must be the case that $r(P(N_1)/N_1) > r(P(N_k)/N_k)$. By Lemma 1.1, it follows that $\frac{P(N_1)}{N_1} > \frac{P(N_k)}{N_k}$. Because producers face a higher price under N_1 , they must be better off under trade ratio N_1 than with N_k . Thus N_1 is a Pareto improvement upon N_k . $\square$

This model does not provide any guidance on how the efficiency of non-neutral trade ratios compares to neutral ratios, or on what the most efficient ratio is. Those

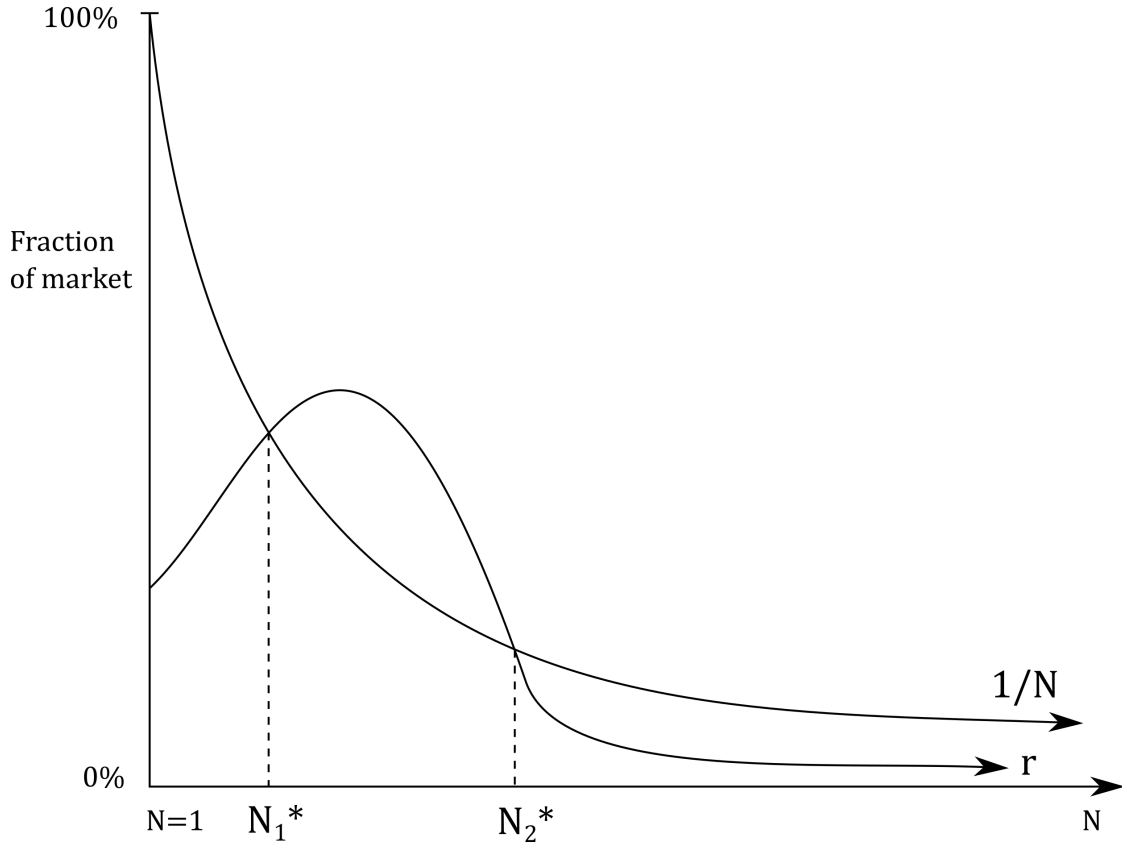


Figure 1.6. Example where there are two neutral trade ratios, N_1^* and N_2^*

comparisons depend on the magnitude of the environmental benefits, which so far we have not considered. I examine this issue in more detail in 1.9.2.

1.6 Dynamic Updating of Trade Ratios

When selecting a trade ratio, not considering the supply and demand responses leads to a non-neutral ratio. I define a naïve trade ratio as one that simply sets $N = \frac{1}{r}$ for the r observed in the market. There are many real world examples of these naïve trade ratios. For example, the procedure for determining wetland mitigation ratios in King and Price [2004] for the US National Oceanic and Atmospheric Administration relies solely on wetland and project characteristics and not on market features. Such ratios are almost surely non-neutral. If demand is inelastic, the neutral trade ratio will be lower than the

naïve trade ratio. If demand is elastic, any neutral trade ratio will be higher than the naïve ratio, and it is possible that a neutral ratio may not exist. However, in this section I demonstrate how these naïve ratios, if properly updated, will approach neutrality.

This section examines dynamic updating of a naïve trade ratios. First, the policy maker observes r in the initial market and sets a naïve trade ratio $N = \frac{1}{r}$. They then observe the new r that is the outcome of the previous choice of N , set a new naïve trade ratio, and continue so forth.

This section assumes that supply (both good and bad) and demand all remain unchanged across time. This is a simplifying assumption, and makes it easier to achieve convergence and neutrality. If the underlying market features are changing over time, it will be more difficult for the policy maker to reach neutrality. Thus these conditions are minimal conditions for achieving neutrality through dynamically updating trade ratios. Additionally, I do not consider any strategic incentives that may arise due to the repeating nature of the interactions between the policy maker, suppliers, and consumers.

Let $N_0 \geq 1$ be the initial trade ratio. Let $N_k = \frac{1}{r_{k-1}}$ and $r_k = r(P(N_k)/N_k)$ for $k \in \mathbb{N}$. The updating procedure converges if $\lim_{k \rightarrow \infty} N_k = c$ for some $c \in [1, \infty)$.²⁹ The updating procedure does not converge if the limit does not exist, or if it is infinite.

The updating procedure does not always converge when a neutral trade ratio exists. Non-convergence can mean $N_k \rightarrow \infty$ (see Figure 1.7), but that is not necessarily the case (see Figure 1.8). As demonstrated in Proposition 1.7, if the procedure converges, then it converges to a neutral trade ratio. It follows that if a neutral trade ratio does not exist, then the updating procedure does not converge, as shown in Corollary 1.2.

Proposition 1.7. *If the updating procedure converges, then it converges to a neutral trade ratio.*

Proof. Because the updating procedure converges, it must be the case that $\lim_{k \rightarrow \infty} N_k = c$

²⁹Note that $c \not\leq 1$ because $r \in (0, 1)$.

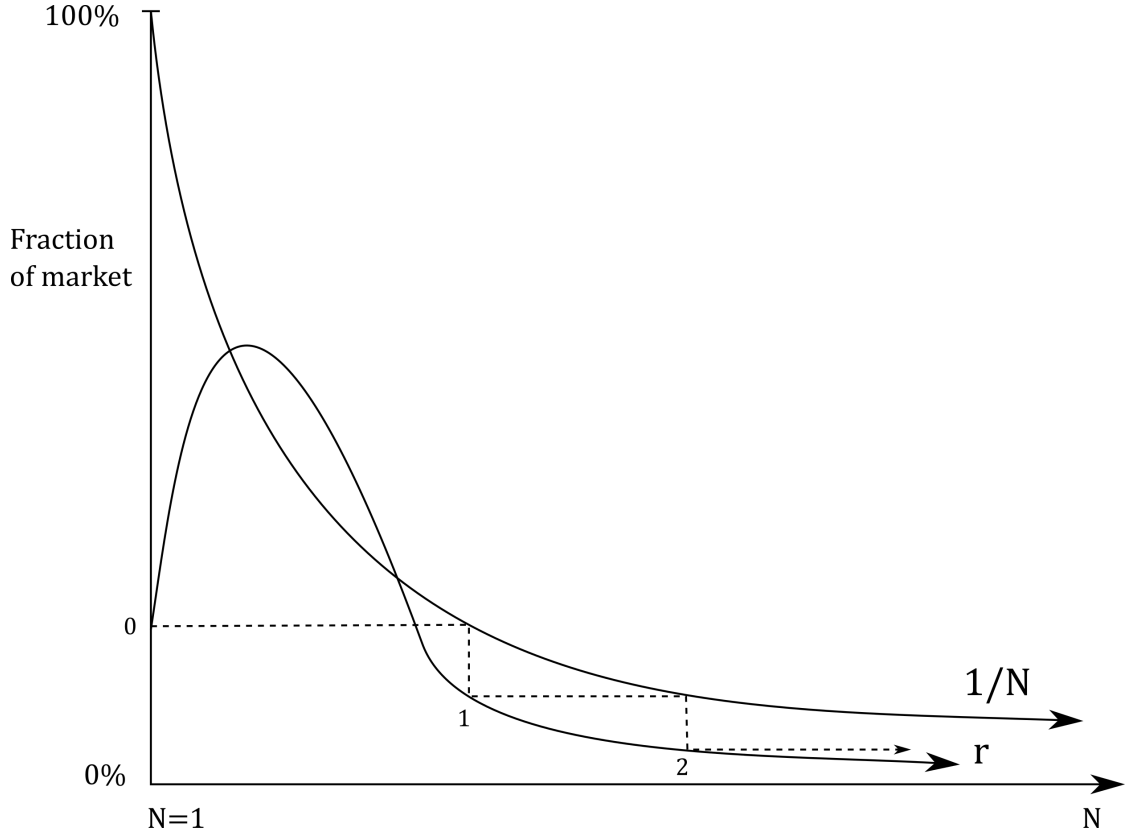


Figure 1.7. Neutral trade ratios exist, but the updating procedure does not converge. $N_0 = 1$

for some $c \in [1, \infty)$. Because $\frac{1}{r_{k-1}} = N_k$, it follows that

$$c = \lim_{k \rightarrow \infty} N_k = \lim_{k \rightarrow \infty} \frac{1}{r_{k-1}} = \lim_{k \rightarrow \infty} \frac{1}{r_k}$$

Because r is continuous, $N = c$ is a neutral trade ratio. □

Corollary 1.2. *If a neutral trade ratio does not exist, then the updating procedure does not converge.*

Proof. Follows directly from Proposition 1.7. □

If there are multiple neutral trade ratios, the updating procedure may converge to one of the higher—and less efficient—trade ratios. Figure 1.9 gives an example of this.

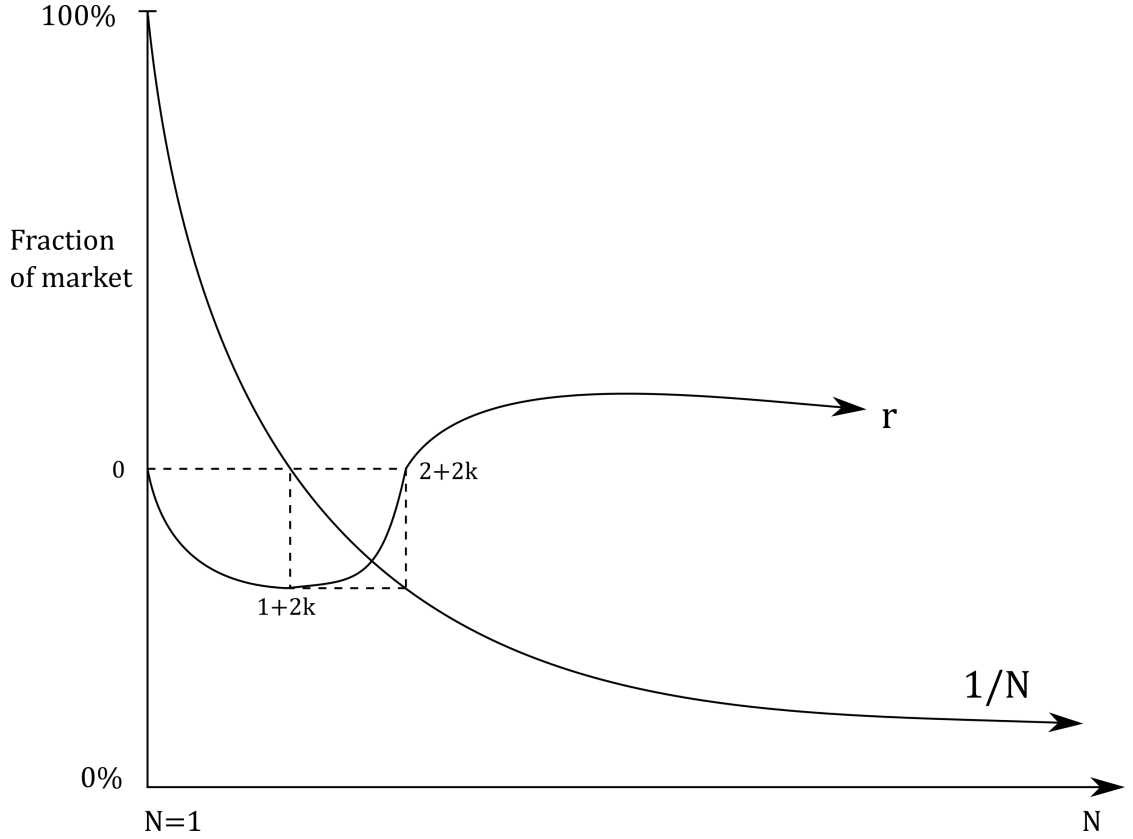


Figure 1.8. A neutral trade ratio exists, but the updating procedure does not converge and $N_k \not\rightarrow \infty$, with $N_0 = 1$

Proposition 1.8 demonstrates that in any instance where the updating procedure does not converge and $N_k \not\rightarrow \infty$, there is sufficient information to bound a neutral trade ratio. The policy maker observes trade ratios N' and N'' such that $r' < \frac{1}{N'}$ and $r'' > \frac{1}{N''}$, and deduces that there exists a neutral N^* between N' and N'' . The policy maker can then select a new N_0 between the bounds and attempt to pin down N^* .

Proposition 1.8. *Let $\lim_{k \rightarrow \infty} N_k$ not exist, and further let $N_k \not\rightarrow \infty$. Then it must be the case that there exists a k' such that $r_{k'} \geq \frac{1}{N_{k'}}$. Further, it is the case that a neutral trade ratio, N^* , exists and that $N^* < N_{k'}$.*

Proof. Assume that there does not exist a k' such that $r_{k'} \geq \frac{1}{N_{k'}}$. Then it must be the case that $N_k < N_{k+1}$ for all k (that is, N_k is strictly increasing in k). It follows that one of the

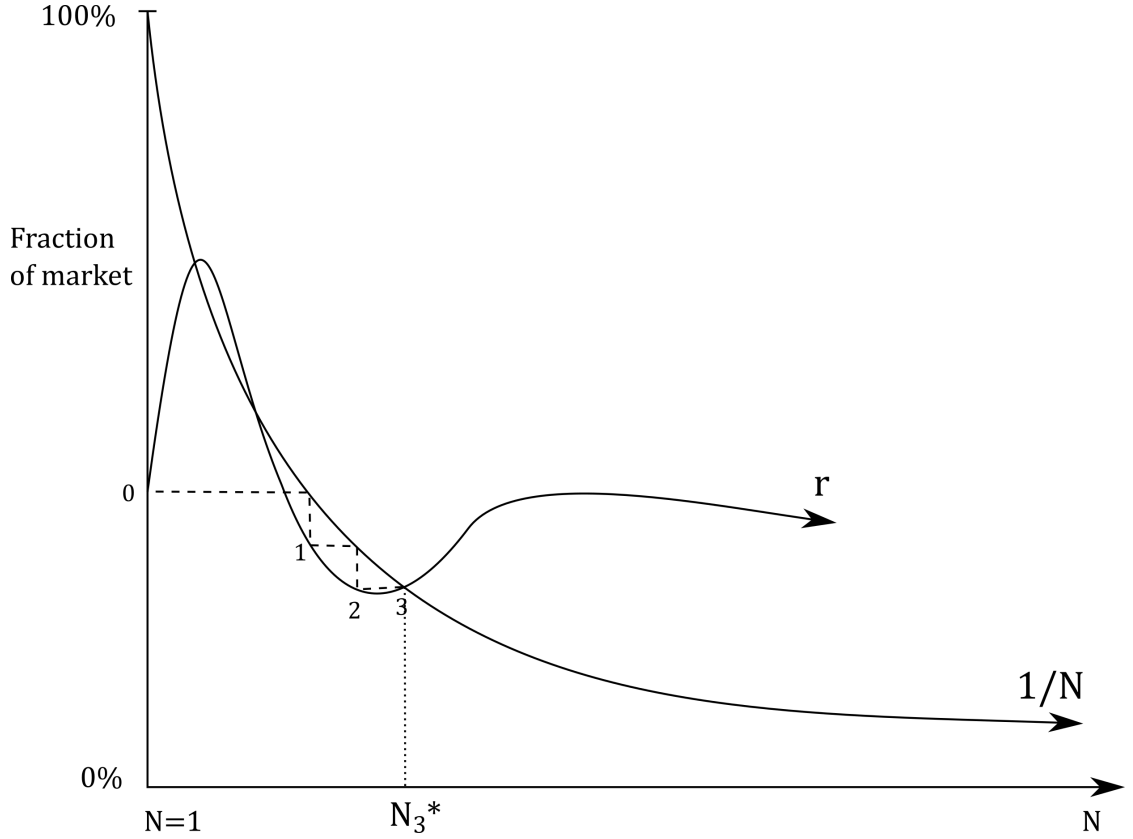


Figure 1.9. The updating procedure converges to N_3^* , the third lowest neutral trade ratio, with $N_0 = 1$

following must be true: either the updating procedure converges to a neutral trade ratio, or $N_k \rightarrow \infty$. Because I have assumed that $\lim_{k \rightarrow \infty} N_k$ does not exist and that $N_k \not\rightarrow \infty$, it must be the case that there exists a k' such that $r_{k'} \geq \frac{1}{N_{k'}}$.

The second statement follows immediately from Proposition 1.3. $\square$

Proposition 1.8 does not cover the situation in which a neutral trade ratio exists but the updating procedure approaches an infinite trade ratio (see Figure 1.7). However, you can consider an infinite trade ratio to be neutral because it essentially bars trade between producers and consumers. Then in all cases, the updating procedure either approaches neutrality or bounds a neutral trade ratio. It may not converge to the most efficient neutral trade ratio. However, the procedure will result in an approximation of

neutrality.

1.7 Policy Discussion

So far, this paper has discussed when neutral trade ratios exist, and how to properly set them. One of the basic assumptions was that the market features were all exogenous. In reality, the policy maker may be able to exert some control over these market features. This section focuses on how a policy maker may be able to influence the makeup of good vs. bad supply, and how this impacts neutral trade ratios.

1.7.1 Monitoring

In some cases, supplier quality can be verified with monitoring. In the model in this paper, I assume that monitoring has occurred prior to the implementation of the policy and that the policy maker can no longer distinguish between the remaining suppliers (additionally, Appendix 1.9.4 demonstrates how monitoring costs could be incorporated into an underlying model of supply). However, in many cases, the policy maker may face a trade off with monitoring. On the one hand, they wish to exclude bad suppliers from the market. On the other hand, the monitoring necessary to determine this could be very costly. While this paper does not address the costs of monitoring, it can give insight into the benefits of additional monitoring.

Consider the linear example in Section 1.3. The neutral trade ratio, when it exists, is $N^* = \frac{ac+cX}{ac-bX}$. The partial derivative $\frac{\partial N^*}{\partial X}$ is strictly positive, meaning that increasing the number of bad suppliers, X , increases the neutral trade ratio. Recall from Proposition 1.6 that lower trade ratios are better for both demanders and suppliers. Thus excluding more bad suppliers from the market is a Pareto improvement for everyone except the excluded bad suppliers.

Thus in situations where monitoring costs are sufficiently low, the policy maker

can use monitoring in addition to the trade ratio policy to achieve neutrality. In fact, many trade ratio policies incorporate monitoring. For example, the Clean Development Mechanism incorporates methodologies such as barrier analysis, which requires that the offset project demonstrates that there are economic barriers preventing the project from going forward in the absence of the offset revenue (Schneider [2009]). For wetlands, ex-post monitoring is often required for five years to ensure that the wetlands are in fact successful (Environmental Law Institute [2002]). When the monitoring technology is imperfect or costly, the policy maker can use a trade ratio to compensate for the remaining bad suppliers.

1.7.2 Segmenting Supply

Sometimes there are observable characteristics that are correlated with supplier type. For example, consider carbon offsets derived from HFC-23 destruction. HFC-23 is a potent greenhouse gas, with 11,700 times the warming potential of carbon dioxide (Schneider [2011]). HFC-23 is produced mainly as a byproduct from HCFC-22 production (a refrigerant), and can either be released into the atmosphere or destroyed at relatively low cost. There is extensive evidence that firms produce an excess of HFC-23 when they can sell its destruction as a carbon offset (Schneider [2011]). Although some HFC-23 derived offsets are additional, overall HFC-23 offsets are more likely to be non-additional than other types of offsets. In 2013, the EU Emissions Trading Scheme stopped accepting offsets from HFC-23 (European Commission [2011]).

Excluding categories of suppliers or specific sources of permits is effective whenever the quality of a supplier is highly correlated with observable characteristics: the policy maker can simply ban suppliers with those characteristics from the permit market. Although some good suppliers may be left out, if the correlation is high enough this impact will be negligible.

In some cases, the policy maker may be able to observe a noisy signal of supplier quality. For example, wetland trade ratios can depend on the expertise of the mitigation agency: the Washington State Department of Ecology bases their ratios on (among other things) “the expertise and experience of the agency or consultant proposing to carry out the project” (Washington Department of Ecology [2005]). This can be an effective strategy when an observable characteristic is correlated with supplier quality.

The policy maker can also assign different trade ratios to different types of suppliers. There is no reason to believe that the makeup of supply is the same across different supplier categories. For instance, the supply curves for HFC offsets might be very different from the supply curves for reforestation offsets. In these situations, the policy maker can gain additional flexibility for achieving neutrality by allowing the trade ratios to vary across supplier types. Appendix 1.9.3 examines and expands the basic trade ratio model to one with several supply sectors, each with its own trade ratio.

1.8 Conclusion

This paper demonstrates that, unlike previously believed, trade ratios cannot always attain environmental neutrality. The model presented here does not incorporate uncertainty, and thus the results should be viewed as minimum conditions for achieving neutrality. In a model with uncertainty, policy makers have less information, making it weakly more difficult to reach neutrality. A full analysis of trade ratios with uncertainty is left for future research.

The model in this paper is static. First, it does not provide any guidance for setting trade ratios in markets where supply and demand are changing over time. Second, none of the participants face any strategic incentives. While these features are outside the scope of this paper, it would be beneficial for future research to examine neutrality in a setting with multiple time periods and strategic incentives.

Failure to consider supply and demand responses to trade ratios results in non-neutrality and potential efficiency losses. If demand is inelastic, the naïve trade ratio will be higher than the neutral trade ratio. If demand is elastic, the naïve trade ratio will be lower than any neutral trade ratios, if they exist. Additionally, policy makers should be aware that neutrality is not always possible with trade ratios. Non-existence of neutral trade ratios occurs when demand is elastic and the ratio of good supply to bad supply is worsening at a sufficiently high rate as the price decreases. If these conditions are present and neutrality is the ultimate goal, it behooves policy makers to consider alternate policies.

1.9 Appendices

1.9.1 Additional Proofs

Proof of Lemma 1.1:

Proof.

$$\epsilon_S(\cdot) = \frac{S'(\cdot)}{S(\cdot)}P > \frac{X'(\cdot)}{X(\cdot)}P = \epsilon_X(\cdot)$$

$\Leftrightarrow$

$$S'(\cdot)X(\cdot) - S(\cdot)X'(\cdot) > 0$$

$\Leftrightarrow$

$$\frac{S'(\cdot)[S(\cdot) + X(\cdot)] - S(\cdot)[S'(\cdot) + X'(\cdot)]}{[S(\cdot) + X(\cdot)]^2} > 0$$

$\Leftrightarrow$

$$\frac{d \frac{S(\cdot)}{S(\cdot) + X(\cdot)}}{d \cdot} > 0$$

□

Proof of Proposition 1.1:

Proof. First note that $\frac{dP(N)/N}{dN} < 0$ if and only if $\varepsilon_{P,N} = \frac{dP(N)}{dN} \cdot \frac{N}{P(N)} < 1$, i.e., the elasticity of the equilibrium price with respect to N is less than 1. This is because

$$\frac{dP(N)/N}{dN} = \frac{dP(N)}{dN} \cdot \frac{1}{N} - \frac{P(N)}{N^2}$$

which implies that $\frac{dP(N)/N}{dN} < 0$ if and only if $\frac{dP(N)}{dN} \cdot \frac{N}{P(N)} < 1$. Recall the equilibrium condition

$$D(P(N)) = \frac{1}{N} \left[S\left(\frac{P(N)}{N}\right) + X\left(\frac{P(N)}{N}\right) \right]$$

Differentiating the equilibrium condition with respect to N yields

$$\begin{aligned} D'(P(N)) \frac{dP(N)}{dN} = \\ -\frac{1}{N^2} \left[S\left(\frac{P(N)}{N}\right) + X\left(\frac{P(N)}{N}\right) \right] + \frac{1}{N} \left[S'\left(\frac{P(N)}{N}\right) + X'\left(\frac{P(N)}{N}\right) \right] \left[\frac{dP(N)}{dN} \cdot \frac{1}{N} - \frac{P(N)}{N^2} \right] \end{aligned}$$

Which holds if and only if

$$\frac{dP(N)}{dN} \cdot \frac{N}{P(N)} = \frac{-\frac{N}{P(N)} \left[S\left(\frac{P(N)}{N}\right) + X\left(\frac{P(N)}{N}\right) \right] - S'\left(\frac{P(N)}{N}\right) - X'\left(\frac{P(N)}{N}\right)}{N^2 D'(P(N)) - S'\left(\frac{P(N)}{N}\right) - X'\left(\frac{P(N)}{N}\right)}$$

Thus $\frac{dP(N)}{dN} \cdot \frac{N}{P(N)} < 1$ if and only if:³⁰

$$\begin{aligned} -\frac{N}{P(N)} \left[S\left(\frac{P(N)}{N}\right) + X\left(\frac{P(N)}{N}\right) \right] - S'\left(\frac{P(N)}{N}\right) - X'\left(\frac{P(N)}{N}\right) > \\ N^2 D'(P(N)) - S'\left(\frac{P(N)}{N}\right) - X'\left(\frac{P(N)}{N}\right) \end{aligned}$$

³⁰Note that both the numerator and denominator are negative.

Which is equivalent to

$$S\left(\frac{P(N)}{N}\right) + X\left(\frac{P(N)}{N}\right) < -NP(N) \cdot D'(P(N))$$

In equilibrium, $\frac{1}{N}[S(\frac{P(N)}{N}) + X(\frac{P(N)}{N})] = D(P(N))$, thus the condition becomes

$$-1 > D'(P(N)) \cdot \frac{P(N)}{D(P(N))} = \varepsilon_D$$

Thus the price per offset is decreasing in N if and only if demand is elastic at $P(N)$. $\square$

Lemma 1.2. *As N increases, $P(N)$ increases, e.g., $\frac{dP(N)}{dN} > 0$.*

Proof.

$$\frac{dP(N)}{dN} = \frac{-N\left[S\left(\frac{P(N)}{N}\right) + X\left(\frac{P(N)}{N}\right)\right] - P(N)\left[S'\left(\frac{P(N)}{N}\right) + X'\left(\frac{P(N)}{N}\right)\right]}{N^3D'(P(N)) - N\left[S'\left(\frac{P(N)}{N}\right) + X'\left(\frac{P(N)}{N}\right)\right]}$$

Note that $S(\cdot) \geq 0$, $X(\cdot) > 0$, $S'(\cdot) > 0$, $X'(\cdot) \geq 0$, $D'(\cdot) \leq 0$, and $N \geq 1$. This implies that $\frac{dP(N)}{dN} > 0$. $\square$

1.9.2 Efficiency

The basic model presented in this paper can only compare the efficiency of different neutral trade ratios. However, these neutral trade ratios are not necessarily the most efficient. How would the policy maker change the trade ratio if they wished to increase efficiency? The basic model does not provide any guidance here, as so far we have not defined the environmental benefits of the policy.

To illustrate the variability of efficient trade ratios, I will work through a simple example with linear additional supply, demand, and environmental benefits, and constant non-additional supply. As in the previous linear example in Section 1.3, demand is

$D(P) = a - bP$, additional supply is $S(P/N) = cP/N$, and non-additional supply is $X(P/N) = X$. Environmental benefits are $d[S(P/N) - D(P)]$. Assume that $a, b, c, d, X \in \mathbb{R}^{++}$. Note that the elasticity condition is satisfied.

The equilibrium price $P(N)$ is given by the market equilibrium condition, which simplifies to $P(N) = \frac{aN-X}{bN+c/N}$.³¹

The neutral trade ratio satisfies $D(P(N)) = S(P(N)/N)$, or equivalently, $a - bP(N) = cP(N)/N$. Substituting in the equilibrium price and simplifying yields a unique solution, $N^* = \frac{c(a+X)}{ac-bX}$. Note that we need $ac > bX$ to ensure that N^* exists and is positive. Note that because $c(a+X) > ac - bX$, N^* will be greater than 1 when it exists.

The efficient trade ratio maximizes total economic surplus, which includes consumer surplus, producer surplus, and the environmental benefits. Consumer surplus is $\frac{1}{2} [a - bP(N)] \cdot [\frac{a}{b} - P(N)]$. Producer surplus is $X \frac{P(N)}{N} + \frac{cP(N)^2}{2N^2}$. Environmental benefits are $d [\frac{cP(N)}{N} - a + bP(N)]$. Thus the maximization problem is:

$$\max_N \frac{1}{2} [a - bP(N)] \cdot [\frac{a}{b} - P(N)] + X \frac{P(N)}{N} + \frac{cP(N)^2}{2N^2} + d [\frac{cP(N)}{N} - a + bP(N)]$$

where $P(N) = \frac{aN-X}{bN+c/N}$ and $N \geq 1$.

The solution to the maximization problem is not tractable. To gain intuition on the efficient trade ratio, I will examine how the parameters d , the magnitude of the environmental benefits, and X , the quantity of bad suppliers, influence it.

First, consider changes in d , the environmental benefits. Note that the value of d has no influence on the neutral trade ratio. If $d = 0$, the efficient trade ratio will be one. In this scenario, there are no environmental benefits, thus the first fundamental welfare theorem of economics states that the market equilibrium (with no policy intervention) is efficient. As d becomes large, the environmental benefits portion of the maximand

³¹For simplicity, I exclude the parameter values for which the price is zero.

overwhelms the market surplus portion. Thus as d approaches infinity, the efficient trade ratio will tend towards either the trade ratio at which environmental benefits are maximized or towards infinity.³² Note that the trade ratio that maximizes environmental benefits is typically not the same as the neutral trade ratio. Also note that depending on the parameter values, the efficient trade ratio can be either higher or lower than the neutral trade ratio.

Next consider variations in the magnitude of X , the quantity of bad suppliers. If X is zero, the neutral ratio will be 1 because all suppliers are good and no market intervention is needed. However, at $X = 0$ the efficient trade ratio could be above 1, particularly if d is high. Recall that as X increases, the neutral trade ratio increases and that for sufficiently high values of X the neutral ratio does not exist. The efficient ratio does not follow such a straightforward pattern. For high values of X , the environmental benefits will always be negative because r never crosses $1/N$. Beyond that, efficiency for high X depends on the value of the other parameters due to the trade off between market surplus and environmental benefits.

1.9.3 Multiple Trade Ratios

In this section I consider the case where the policy maker is able to screen suppliers based on observable characteristics. Each segment of supply has different characteristics, and therefore the policy maker may wish to use different trade ratios for each segment. For example, supply for carbon offsets could be divided into offsets from forestry, offsets from agriculture, and offsets from hydrofluorocarbons, each with a different trade ratio.

These sector-differentiated trade ratio policies are quite common in practice. In wetland mitigation there are frequently different trade ratios for different types of

³²If the maximal environmental benefits are positive, the efficient ratio will tend towards the ratio that achieves that maximum. Otherwise, it will approach infinity.

wetlands. For example, a rare type of wetland may have a higher trade ratio than a more common type of wetland. Trade ratios also vary by the type of wetland restoration: preserving existing wetlands often involves a higher trade ratio than creating a new wetland.³³

There are also cases where some sectors of supply are completely excluded from a trade ratio policy. In 2013, the EU Trading Scheme stopped accepting offsets from HFC-23 destruction. HFC-23 is a byproduct of refrigerant production and a potent greenhouse gas. In banning HFC-23 offsets, the EU cited concerns that many of the offsets did not represent additional carbon abatement (European Commission [2011]).

I model the scenario where there are multiple trade ratios as follows. There are I segments of supply, with labels $\{1, \dots, I\} = \mathcal{I}$. Each segment $i \in \mathcal{I}$ has trade ratio N_i , good supply S_i , and bad supply X_i . Let $\mathbf{N}$ denote the vector $(N_1, \dots, N_i, \dots, N_I)$ of all trade ratios. Demand is still D (consumers don't care what type of offset they buy) and price is still P . I retain all assumptions on D , S_i , and X_i from the single trade ratio model. The market equilibrium condition is:

$$D(P) = \sum_{i \in \mathcal{I}} \frac{1}{N_i} [S_i(P/N_i) + X_i(P/N_i)]$$

For the remainder of this section, let P be a function of $\mathbf{N}$ as determined by the market equilibrium condition.

As in the main model, I will assume that the elasticity of good supply is strictly greater than the elasticity of bad supply for each sector i . Specifically, $\varepsilon_{S_i}(\cdot) > \varepsilon_{X_i}(\cdot)$ for all $i \in \mathcal{I}$.³⁴ Let $r_i(P(\mathbf{N})/N_i) = \frac{S_i(P(\mathbf{N})/N_i)}{S_i(P(\mathbf{N})/N_i) + X_i(P(\mathbf{N})/N_i)}$.

Recall that in the one trade ratio model, neutrality could be defined in two

³³For examples of these types of policies, see the Michigan Department of Environmental Quality's wetland mitigation ratio policies. Most other states have similar policies.

³⁴Except where ε_{S_i} is undefined at $S_i(0) = 0$.

ways: $D(P(N)) = S(P(N)/N)$ or $r(P(N)/N) = \frac{1}{N}$, and that these were equivalent. In the multiple trade ratio model, these two definitions become $D(P(\mathbf{N})) = \sum_{i \in \mathcal{I}} S_i(P(\mathbf{N})/N_i)$ and $r_i(P(\mathbf{N})/N_i) = \frac{1}{N_i}$, which are no longer equivalent. Specifically, the first neutrality condition requires neutrality across the market as a whole, while the second requires neutrality in each sector. Note that if the second neutrality condition is satisfied for all i , the first will also be satisfied. The converse is not true.

Recall the price results from the one trade ratio model: $\frac{dP(N)}{dN} > 0$ always³⁵ and $\frac{dP(N)/N}{dN} > 0$ if and only if demand is inelastic at $P(N)$.³⁶ A variation of the first still holds in this case: $\frac{\partial P(\mathbf{N})}{\partial N_i} > 0$ for all i , meaning that increasing a trade ratio N_i will increase the demand-side price. Note that this implies that $\frac{\partial P(\mathbf{N})/N_j}{\partial N_i} > 0$ for $i \neq j$, meaning that increasing trade ratio N_i weakly increases the supply side price for sector j . It follows that Proposition 1.1 no longer holds for the multiple trade ratio model: increasing trade ratio i increases the supply side price in other sectors, shifting some of the supply to those sectors. This in turn limits the price increase in sector i . Thus it is no longer the case that the supply-side price change can be characterized with only the elasticity of demand. Specifically, it will still be the case that if demand is elastic at P , $\frac{\partial P(\mathbf{N})/N_i}{\partial N_i} < 0$.³⁷ However, if demand is inelastic at $P(\mathbf{N})$, $\frac{\partial P(\mathbf{N})/N_i}{\partial N_i}$ can be positive or negative.

To get a better idea for what is going on in the multiple trade ratio model, I will use a linear example with two supply sectors. The set up is similar to the other linear examples I have examined in this paper. Demand is $D(P) = a - bP$. Good supply in sector $i \in \{1, 2\}$ is $S_i(P/N_i) = c_i \frac{P}{N_i}$, and bad supply is $X_i(P/N_i) = X_i$. The parameters

³⁵See Lemma 1.2.

³⁶See Proposition 1.1.

³⁷The proof of this follows the same technique as the proof for Proposition 1.1, but starts with the market equilibrium condition for the multiple trade ratio model.

a, b, c_i , and X_i are all strictly positive for all i . The market equilibrium condition is:

$$D(P) = \frac{1}{N_1} \left[S_1(P/N_1) + X_1(P/N_1) \right] + \frac{1}{N_2} \left[S_2(P/N_2) + X_2(P/N_2) \right]$$

and the market price is

$$P(N_1, N_2) = \frac{aN_1N_2 - X_1N_2 - X_2N_1}{\frac{c_1N_2}{N_1} + \frac{c_2N_1}{N_2} + bN_1N_2}$$

Recall that there are two different neutrality conditions in this model. The first, $D(P(\mathbf{N})) = \sum_{i \in \mathcal{I}} S_i(P(\mathbf{N})/N_i)$, only requires that the market as a whole reaches neutrality. The second, $r_i(P(\mathbf{N})/N_i) = \frac{1}{N_i}$, requires that each individual sector reaches neutrality, and implies the first neutrality condition. I will start with the first, more general neutrality definition.

Solving for the more general neutrality condition, we get that neutrality is satisfied if

$$\begin{aligned} (ac_2 + c_2X_2)N_1^2 + (ac_1 + c_1X_1)N_2^2 + (bX_2 - ac_2)N_1^2N_2 \\ + (bX_1 - ac_1)N_1N_2^2 + (c_1X_2 + c_2X_1)N_1N_2 = 0 \end{aligned}$$

Thus any set of trade ratios (N_1, N_2) that satisfy the above equation and $N_1, N_2 \geq 1$ will achieve neutrality. In many scenarios there exists a set (or sets) of neutral ratios such that $N_1 = N_2$, meaning that the market does not need to be segmented to achieve neutrality. Of course, forcing these trade ratios to be equal is not necessarily the best or most efficient policy. Specifically, allowing the trade ratios to differ increases the policy maker's choice set and thus the optimal trade ratio policy is weakly better when the market is segmented. Also note that by allowing the trade ratios to differ across sectors, neutrality will be achievable for a larger set of supply and demand functions.

The large set of neutral trade ratios with a segmented market leads to the question of which neutral ratios are best. The following general result offers some guidance.

Lemma 1.3. *Consider two sets of neutral trade ratios, $\mathbf{N} = (N_1, N_2, \dots, N_I)$ and $\mathbf{N}' = (N'_1, N'_2, \dots, N'_I)$ such that $N_i \leq N'_i$ for all $i \in \mathcal{J}$ and $N_i < N'_i$ for at least one $i \in \mathcal{J}$. Then $\mathbf{N}$ is a Pareto improvement upon $\mathbf{N}'$.*

Proof. $\mathbf{N}$ and $\mathbf{N}'$ are both environmentally neutral, thus they have the same environmental benefits.

Recall that $\frac{\partial P(\mathbf{N})}{\partial N_i} > 0$. It follows that $P(\mathbf{N}) < P(\mathbf{N}')$. Because consumers face a lower price under $\mathbf{N}$, they must be better off with trade ratios $\mathbf{N}$ than with $\mathbf{N}'$.

If $N_i \leq N'_i$ for all $i \in \mathcal{J}$ and $N_i < N'_i$ for at least one $i \in \mathcal{J}$, then it must be the case that $r_i(P(\mathbf{N})/N_i) \geq r_i(P(\mathbf{N}')/N'_i)$ for all $i \in \mathcal{J}$ and that the inequality is strict for at least one i . It follows that $\frac{P(\mathbf{N})}{N_i} \geq \frac{P(\mathbf{N}')}{N'_i}$ and that the inequality is strict for at least one i . Because producers face a higher price under $\mathbf{N}$, they must be better off under trade ratio $\mathbf{N}$ than with $\mathbf{N}'$. Thus $\mathbf{N}$ is a Pareto improvement upon $\mathbf{N}'$. $\square$

Thus, sets of trade ratios where both ratios are lower will be more efficient. However, in the two trade ratio linear example, both trade ratios cannot be minimized concurrently. Thus the policy maker must rely on a different criterion to select trade ratios. One possibility is using the more restrictive neutrality definition that requires neutrality in each supply sector individually.³⁸ In addition to limiting the selection of neutral trade ratios, there are several reasons why policy makers may wish for each sector to achieve neutrality. Political reasons may make sector-by-sector neutrality more desirable, so that one sector isn't essentially subsidizing another. It could also be the case

³⁸ Another selection criteria could be based on welfare and efficiency. As we saw in 1.9.2, it is difficult to solve for efficient trade ratios analytically. This, combined with the policy maker's penchant for neutrality, leads me to focus on the more selective neutrality condition.

that different sectors provide slightly different environmental goods—such as different kinds of carbon offsets—and the policy goal is neutrality for each of these types of goods.

By restricting attention to sector-by-sector neutrality, the neutrality condition for the linear case becomes:

$$\frac{1}{N_i} = \frac{c_i P(\mathbf{N})/N_i}{c_i P(\mathbf{N})/N_i + X_i}$$

The set of these equations for each i , combined with the equation for equilibrium price, gives a system of equations that characterizes the neutral trade ratios (if they exist).

To gain further insight on how the sectors' neutral trade ratios relate to each other, I further restrict my attention to the two sector case where $c_1 = c_2$ and $X_1 < X_2$. Under these conditions, the neutral trade ratios satisfy:

$$\frac{1}{N_i} = \frac{cP(\mathbf{N})/N_i}{cP(\mathbf{N})/N_i + X_i}$$

It follows that $\frac{cP(\mathbf{N})}{N_1} + X_1 = \frac{cP(\mathbf{N})}{N_2} + X_2$, which implies that if $X_1 < X_2$ and neutral ratios exist, then $N_1^* < N_2^*$. Thus the sector with fewer bad suppliers has a lower neutral trade ratio.

Figure 1.10 illustrates how the multiple trade ratio model works using an example of a market with three supply sectors and sector-by-sector neutrality. In this example, neutrality can be achieved in sector one with a low trade ratio, in sector two with a higher trade ratio, and cannot be achieved in sector three. Thus the neutral trade ratios are N_1^* , N_2^* , and excluding sector three from the market (essentially an infinite trade ratio for sector three). Because r_i depends on all three trade ratios, the graph holds N_j for $j \neq i$ constant at its neutral rate (with N_3^* being replaced by excluding sector three from the market). So $r_1(P(N_1, N_2, N_3)/N_1)$ is graphed holding N_2 constant at N_2^* and with sector 3 excluded from the market.

This Appendix illustrates why policy makers may choose to segment supply,

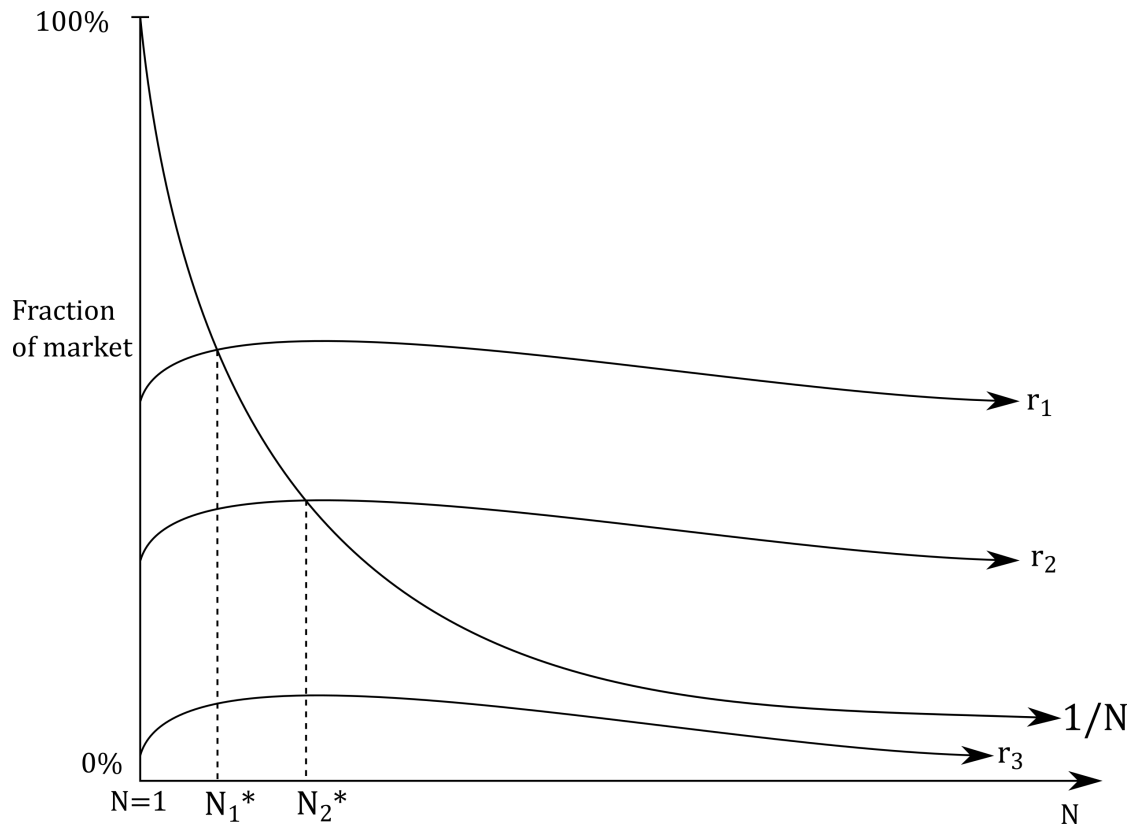


Figure 1.10. A market with three supply sectors

using different trade ratios for observably different types of suppliers. Additionally, it provides an explanation for why some policies exclude certain types of environmental goods, for example certain types of carbon offsets.

1.9.4 Microeconomic Theory Foundation for Model Assumptions

This Appendix provides an additional microeconomic theory foundation for some of the simplifying assumption in the main model. In particular, it demonstrates that the division of suppliers into only two types—good and bad—is not as restrictive as it initially appears, and that a more nuanced underlying model with a continuum of types can generate the simple two-type supply function model. Additionally, I demonstrate how monitoring costs can be incorporated into the model, and that the elasticity assumption

of the main model can be generated by a simple restriction on the relationship between cost and supplier type.

The model in this Appendix consists of a continuum of suppliers, ranging from 100% good to 100% bad (and everything in between). For simplicity, I will refer to the environmental good that the suppliers create as a “good” throughout this section, despite the fact that it is not 100% the true environmental good. I will refer to the supplier’s type as their quality. That is, a high quality supplier provides a higher percentage of the true environmental good than a low quality supplier.

In this model, suppliers’ production costs are correlated with their type. In particular, higher quality suppliers have, on average, higher costs. This can be interpreted in several ways. For wetland mitigation, lower quality suppliers could be skipping steps or hiring less skilled (and less expensive) labor—which lowers both costs and quality. For carbon offsets, it could be the case that lower quality suppliers are not changing their behavior (and therefore not incurring costs) while higher quality suppliers are engaging in costly behavioral changes.

There is an additional cost (referred to as monitoring cost) that is correlated with supplier type in the opposite direction. That is, the monitoring cost is decreasing the higher the quality of the supplier. This cost can be interpreted in several ways. For example, worse supplier types may need to engage in more costly monitoring evasion, extra paperwork to falsify their type, and so forth. It could also be the case that worse supplier types are more likely to get caught producing a sub-par environmental good, and then face fines or legal costs.

Supplier types are denoted by θ , which is distributed on the interval $[0, 1]$ according to some continuous cdf F . Types can be interpreted as the quality of the supplier, that is, as the percentage of true environmental good the supplier creates. Thus a type $\theta = 1$ makes a 100% true environmental good, type $\theta = 0$ creates a fully fake environmental

good, and type $\theta = 0.5$ makes a good that is half and half. Supplier types are private information and are only known to that supplier. For simplicity, I assume that each supplier can provide either zero or one unit of the environmental good.³⁹

Supplier production costs are a function of supplier type. Specifically, the cost is $c(\theta)$. I assume that $c(\theta) \geq 0$, that it is differentiable, and that $c'(\theta) > 0$. That is, higher quality suppliers have higher production costs. I also allow for the possibility that there is an additional, non-deterministic component to production cost. This second component is denoted ε_c and is independently distributed according to some continuous cdf F_c with mean zero. Thus total cost is $c(\theta) + \varepsilon_c$. This allows for the average cost to be increasing in supplier type rather than a strictly deterministic relationship between type and cost. I assume that firms observe their own ε_c prior to their decision to participate in the market.

Monitoring costs are a function of the supplier type, in particular, $m(\theta)$. I assume that $m(\theta) \geq 0$, that it is differentiable, and that $m'(\theta) \leq 0$. That is, higher quality suppliers have lower monitoring costs. The second component of the monitoring cost is an error term, ε_m , which is independently distributed according to some continuous cdf F_m with mean zero. Thus the total monitoring cost is $m(\theta) + \varepsilon_m$. I assume that suppliers do not observe ε_m until after their decision to produce the good and participate in the market.

Individual suppliers have no market power, and treat the market price P/N as given. Suppliers are risk neutral and seek to maximize their expected profits. That is, they maximize the expected value of $P/N - c(\theta) - \varepsilon_c - m(\theta) - \varepsilon_m$. Note that a supplier will choose to produce the good and participate in the market if $P/N \geq c(\theta) + \varepsilon_c + m(\theta)$, i.e., if their expected profits are positive for the given market price. Note that ε_m is not in this expression because it is not observed until after the decision to participate in the

³⁹This is not a restrictive assumption as long as firms who can provide multiple units make their supply decisions independently for each unit.

market.

The total expected cost (prior to the realization of the two error terms) for each supplier is $c(\theta) + m(\theta)$. This expression can be both increasing or decreasing in θ . In particular, it is increasing in θ where $c'(\theta) > -m'(\theta)$, and is decreasing in θ where $c'(\theta) < -m'(\theta)$. This paper focuses primarily on the situation where the first case holds for all θ , where total supplier cost is always increasing in supplier quality. This assumption is supported by empirical evidence. In particular, there is evidence from the wetland mitigation literature that finds that suppliers with lower costs tend to be of lower quality (King and Bohlen [1994a] and King and Bohlen [1994b]). If the reverse were true instead, the many of the results in this paper would simply switch direction or apply to elastic demand rather than inelastic.⁴⁰ Thus for the remainder of this section, I will assume that $c'(\theta) > -m'(\theta)$ for all θ .

Next, consider the expected value of θ , conditional on being in the market, as a function of the price, P/N . Note that this is equivalent to the expected value of r , where r is defined as in the main model, i.e., the fraction of supply that is good.

$$E\left[r\left(\frac{P}{N}\right)\right] = Pr_{\theta}[P/N \geq c(\theta) + \varepsilon_c + m(\theta)]^{-1} \int_0^1 \mathbb{1}\{P/N \geq c(\theta) + \varepsilon_c + m(\theta)\} \theta dF$$

Prior to the realization of ε_c , we can define an expected cutoff value of θ for the supplier who is indifferent between participating and not. I will denote this cutoff as $\bar{\theta}$, defined as the value of θ such that $P/N = c(\bar{\theta}) + m(\bar{\theta})$. Because we are assuming that $c'(\theta) > -m'(\theta)$ for all θ , this cutoff value is unique and well-defined. Additionally, it represents an upper bound on the supplier types who are willing to participate in the market in expectation.⁴¹ Thus supplier types $\theta \in [0, \bar{\theta}]$ will all participate in the market,

⁴⁰For instance, the price change results in Corollary 1.1 would be swapped: simply flip the inequality. Similarly, the existence result in Proposition 1.2 would be for elastic demand, not inelastic.

⁴¹If instead we assumed that $c'(\theta) < -m'(\theta)$ for all θ , it would be a lower bound on the supplier types.

in expectation.

Next, note that $\bar{\theta}$ is increasing in P/N . This is because total cost, $c(\theta) + m(\theta)$ is increasing in θ . Thus as P/N increases, it must be the case that $\bar{\theta}$ increases. It follows that as P/N increases, the expected value of θ for market participants is also increasing. Thus we have demonstrated that the key feature of the simpler, two type model, namely that r is increasing in supply-side price, holds in this more nuanced setting with a continuum of supplier types and monitoring.

Additionally, it is useful to examine the two-type assumption of the paper's main model in more detail. In particular, it can be shown that this model can be generated by a more nuanced underlying model such as the one in this Appendix.

The following demonstrates how to generate a two-type model from the continuum of supplier types model. Recall that $\bar{\theta}$ is defined as the expected cutoff type satisfying $P/N = c(\bar{\theta}) + m(\bar{\theta})$; let this expression define $\bar{\theta}$ as a function of the price, P/N . Now, we need to quantify the total number of expected suppliers in the market at any given price: $\int_0^{\bar{\theta}(P/N)} dF$. This is the total expected supply as a function of P/N . In the two type model, this is equivalent to the sum of $S(P/N)$ and $X(P/N)$. Now we simply need to divide the total supply into the two types. Although suppliers here are on a continuum of good to bad, we can still divide them up using their type to dictate the ratio. It follows that good supply is $S(P/N) = r(P/N) \int_0^{\bar{\theta}(P/N)} dF$, and bad supply is $X(P/N) = [1 - r(P/N)] \int_0^{\bar{\theta}(P/N)} dF$. This will generate two supply functions, good supply S and bad supply X , and a corresponding r that is increasing in P/N .⁴²

⁴²Note that we need additional assumptions to get that S and X are differentiable as in the main model. In particular, we need F and F_c to be atomless. We also need F to have full support on $[0, 1]$, and for F_c to have full support over its domain. The assumption that $X > 0$ is trickier, particularly at a price of zero. In particular, it will be satisfied if monitoring costs are zero. However, the assumption that $X > 0$ is only needed only to ensure that total quantity supplied is strictly positive (to avoid dividing by zero when we compute fractions of the market). If real-world prices are always high enough to ensure some quantity supplied, then this restriction on X is irrelevant. Lastly, the assumption that $S(0) = 0$ follows from assumptions on c .

The assumptions in the main model—in particular, that there are only two types, and that there is no monitoring—may appear restrictive at first glance. However, this Appendix demonstrates that they can easily be generated by a more nuanced model allowing for a continuum of types and monitoring costs.

Chapter 2

Using Waiting Periods to Screen Participants for Voluntary Green Actions

Abstract: I describe a mechanism that uses time delays to screen participants in programs for voluntary actions with both social and private benefits. The setting covers a variety of policy scenarios where consumers are induced to undertake a voluntary “green” (environmentally beneficial) action that may have either positive or negative private net benefits. Examples include subsidies to replace lawns with drought-resistant landscaping, solar panel rebates, or subsidizing energy efficient appliances. The setting is unique in that participants have type-dependent outside options. In other words, some participants would undertake the green action (installing a solar panel or buying an efficient water heater) in the absence of a policy due to its private net benefits (savings on utility bills vs. installation costs), while others would not. This heterogeneity in outside options creates differential impacts of time delays, even when all agents have the same discount factor. The policy designer can exploit the heterogeneous impacts of time delays to screen participants. Using a model with a continuum of agent types, heterogeneous outside options, a homogenous discount factor for all agents, and a budget constraint, I demonstrate three things. First, I characterize how the degree of screening is impacted by the duration of the time delay. Second, I demonstrate that as either the time delay approaches infinity or agents’ discount factor approaches zero, the mechanism approaches full separation. Finally, I describe the optimal waiting period mechanism. These results suggest that waiting periods can be effectively used to screen out non-additional participants in green subsidy programs.

2.1 Introduction

Consumer subsidy programs are commonly used as an incentive to engage in environmentally beneficial activities. However, these programs tend to suffer from an additionality problem: many of the subsidy recipients are not changing their behavior. That is, some so-called “non-additional” participants would have engaged in the environmentally beneficial activity in the absence of a subsidy due to its private benefits. Because these programs also tend to be expensive and have limited budgets, policy makers may desire an effective screening method. If they can remove some of the non-additional participants who are not changing their behavior, they can use their budget to induce more additional participants (who are changing their behavior) to undertake the action, thereby increasing the total environmental benefits. In this paper, I will present and analyze a novel and inexpensive screening mechanism: waiting periods.

The policy setting in this paper applies to a broad set of policies that induce consumers or businesses to take an environmentally beneficial action. Examples include replacing lawns with drought resistant landscaping, weatherizing homes, installing energy efficient appliances, or adding solar panels to a home.¹ For each of these examples, there are private costs, private benefits, and environmental benefits. For example, consider the context of replacing lawns with drought-resistant landscaping. There are private costs, such as paying a landscaper to remove the lawn, and purchasing and planting drought-resistant plants like cacti. There are also private benefits, in the form of reduced water bills. The reduction in water use is an environmental benefit due to the externalities associated with water use.

The key feature of this policy setting is that some consumers have a positive private net benefit of taking the green action, while others have negative net benefits. In

¹There are also examples outside of environmental policies, such as getting vaccinated against the flu.

other words, for some consumers, the private benefits of the action outweigh the private costs, while for others the costs outweigh the benefits. This means that, in the absence of a policy, some consumers would voluntarily take the action on their own. Other agents require a policy incentive to take the action. I will refer to these agents as “additional” and “non-additional.” Additional agents are changing their behavior: they would not have taken the action in the absence of a policy incentive. Non-additional agents do not change their behavior: although they are taking the action, they would have done so in the absence of the policy, and therefore the social benefits they create are not additional. Note that non-additional agents have a positive private net benefit, while additional ones have a negative private net benefit.

The most commonly used policy in this scenario is a subsidy. For example, most utilities in the U.S. offer rebates or subsidized loans for customers who install energy efficient appliances, the federal Weatherization Assistance Program subsidizes weatherization and energy efficiency retrofits for low-income households, and the California Solar Initiative provided rebates for consumers who installed solar panels on their homes. These are just a few examples of a myriad of similar policies.

However, these programs are typically very expensive. For example, the Metropolitan Water District of Southern California went through \$350 million² of rebates in one month for their lawn-replacement program.³ In California, the California Solar Initiative’s programs ran out of money several years early.⁴

There is also evidence that these programs spend a lot of their budget subsidizing non-additional participants. That is, they are paying large volumes of subsidy funds to consumers who are not changing their behavior. In a study of German energy-efficiency

²In 2015 dollars.

³As a comparison, the District’s annual operating budget for that year was \$1.89 billion (2015 dollars).

⁴The programs were intended to run from 2007 through 2016. One ran out of funds in early 2014, another in early 2015, and the third program ran out of funds in mid 2015.

subsidy recipients, Grösche and Vance [2009] estimate that 50% of participants would have engaged in the upgrades without a subsidy. Malm [1996] examined heating system subsidies in the U.S. and estimated that 89% of participants would have done the upgrades without a subsidy.

The combination of limited budgets and non-additional participants leads to a natural question: how can policy makers more effectively screen out non-additional participants? If a policy maker can effectively exclude non-additional participants from the program, they can achieve a higher level of environmental benefits with a given budget. The now excluded non-additional participants will still engage in the environmentally friendly action (because their private net benefits are positive), however they are no longer taking a subsidy payment that could be used to incentivize an additional participant.

Current screening methods focus on observable characteristics. For example, the federal Weatherization Assistance Program only offers subsidies to low-income households. Previous literature on the subject has also focused on screening based on observables. Allcott et al. [2015] recommend limiting participation to households with low incomes, renters, and households who have not previously participated in a related subsidy program. Chidiac et al. [2011] suggest screening for energy efficiency subsidies based on current energy use. DeShazo et al. [2014] recommend screening electric vehicle subsidy recipients based on income.

This paper proposes an alternate screening tool that is not based on observable characteristics: waiting periods. The screening takes advantage of the fact that additional and non-additional participants have different outside options. The policy maker can leverage this heterogeneity to screen participants. This is because non-additional participants are sensitive to time delays, while additional participants are not. Thus a waiting period induces some non-additional participants to not take the subsidy, while retaining all additional participants. Waiting periods are relatively inexpensive to implement, and

can be used in addition to other screening methods (such as screening on income or other observables).

This paper also analyzes optimal waiting periods. I divide this analysis into two cases, one where consumers are more patient than society, and the other where consumers are less patient than society. In the first case, I find that zero-length waiting periods (i.e., no wait) is always optimal. In the second, I find that strictly positive waiting periods may be optimal.⁵ I also present evidence from the literature that the second case—where consumers are less patient than society—is more realistic.

Screening problems are not new in the mechanism design literature.⁶ However, the particular setting here has not been analyzed previously. The most novel feature of this setting that sets it apart from previous research is the type-dependent outside options. For instance, Manelli and Vincent [2007] analyze a similar setting, but without type-dependent outside options.⁷ Moxey et al. [1999] also examine a simplified version of a similar setting to this paper, focusing on screening an agricultural setting. They are able to achieve a separating equilibrium, however their model uses a discrete, two-type distribution and their results do not easily carry over to continuous distributions. Additionally, they do not consider the time aspect used in this paper, nor do they have type-dependent outside options. The mechanism design literature with type-dependent outside options is limited. In particular, because type-dependent outside options can be extremely variable (in the sense that there are many options for the relationship between the type and the outside-options) and depend highly on the specific setting, none of the

⁵The duration of the optimal waiting period depends on the model's parameters, and is often strictly positive.

⁶For early examples of screening problems, see, for example, Vickrey [1961], Myerson [1981], and Mailath [1987].

⁷Additionally, they restrict attention to cases where the social welfare function is linear which is not always the case in this paper. In this paper, I rely on some of their proof techniques, in particular those used in Börgers et al. [2015].

previous literature apply to the setting considered here.⁸ In this paper, I am able to exploit the type-dependent outside options to enable screening. Additionally, the mechanisms described in this paper treat time in a different way than is typically seen. In particular, instead of treating time as a dynamic element of the mechanism design problem, I model it as a delay or waiting period that the mechanism designer can exploit. Combined with the type-dependent outside options, this allows the mechanism designer to use the different sensitivities to delays across types to screen. To the best of my knowledge, no previous papers have used waiting periods or time delays as a screening mechanism.

Section 2.2 presents the model. Section 2.3 demonstrates how a waiting period can screen agents. Section 2.4 examines optimal waiting period mechanisms. Section 2.5 consists of a policy discussion. Section 2.6 concludes. Additional proofs are in the Appendices.

2.2 Model

The model is designed to mimic the setting described previously. That is, a setting where agents have either positive or negative private net benefits for taking a “green” action with external environmental benefits. The policy maker can offer incentives to take this action. Participation in the policy is voluntary. However, conditional on participating, the agent must under take the action prescribed by the policy. The policy assigns actions and transfers as a function of the agent’s reported type. Additionally, the policy maker is subject to a budget constraint.

There is a continuum of agents, each with type $\theta \in \Theta$, where $\Theta = [\underline{\theta}, \bar{\theta}]$. Agents are distributed over Θ according to cdf F and its corresponding pdf f . I assume that

⁸For examples of mechanism design papers with type-dependent outside options in other settings, see Figueroa and Skreta [2007] for results on type-dependent outside options in an auction setting, and Rasul and Sonderegger [2010] for results on type-dependent outside options in principal-agent setting with relationship-specific investment.

F has full support over Θ . Each agent's type, θ , is private information denotes their private net benefit of engaging in the green action (such as installing a solar panel or energy efficient dishwasher). I assume that $\underline{\theta} < 0$ and $\bar{\theta} > 0$, meaning that some agents have positive net benefits, and some have negative net benefits (i.e., costs).⁹ Intuitively, agents with positive net benefits would undertake the green action in the absence of a mechanism, while agents with negative net benefits would not. For example, consider the case where the action is installing an energy efficient water heater. New water heaters are expensive, but an energy efficient one can reduce utility bills. For some agents, it is worthwhile to engage in the upgrade even without an incentive due to the utility bill savings, while for others it is not. The agents' utility function and outside options will model this explicitly.

I define an additional agent as one whose type satisfies $\theta < 0$, and a non-additional agent as one whose type satisfies $\theta \geq 0$. That is, additional agents have negative private net benefits and require an incentive to take the action, while non-additional agents have positive private net benefits and would willingly take the action on their own.¹⁰

The mechanism takes place over continuous time. Let T be the set of all time periods, i.e., $T = \mathbb{R}^+$, and note that it is infinite.¹¹ All agents have discount factor $\delta \in (0, 1)$, which is constant across types. External social benefits are discounted with discount factor $d \in [0, 1]$.

Each agent sends the policy maker a message, denoted $\theta' \in \Theta$. Per the revelation principle, I restrict attention to direct and truthful mechanisms. The mechanism, Γ , assigns an action, $a : \Theta \rightarrow [0, 1]$, a transfer, $m : \Theta \rightarrow \mathbb{R}$, and a time period, $t : \Theta \rightarrow T$ to

⁹If instead $\underline{\theta} \geq 0$, there are no benefits to having a policy. If instead $\bar{\theta} \leq 0$, all agents are additional and a simple subsidy mechanism can be used.

¹⁰The action is continuous on the interval $[0, 1]$, however due to the assumption on the agents' utility functions, they will always choose either 0 or 1 in the absence of a policy.

¹¹ T can easily be restricted to a finite time horizon. This will impact a few results, mainly those where the policy maker wishes to implement a wait longer than the maximal time allowed.

each agent based on the report. The action, a , can be thought of as either a probability of taking a discrete action (like replacing a water heater) or as a percentage of a continuous action (such as replacing part of a lawn). The action can only occur once. The action, a , has an environmental benefit of ωa for some $\omega > 0$. The agent undertakes the action and receives the transfer in time period t .¹² Environmental benefits of the action accrue in time period t as well. If costs or benefits are persistent (e.g., occur in every time period after the action is taken, for instance like utility bill savings), then θ and ω can be thought of as the net present value of the action in the time period in which the action is undertaken.

Agents each have utility $u(\theta'|\theta) = \delta^{t(\theta')}[a(\theta')\theta + m(\theta')]$, where θ is the agent's true type and θ' is the agent's reported type. Note that this means that agents get nonzero utility in the time period for which they take an action, and zero utility in all other time periods. Recall that if the action generates a stream of costs or benefits, these are all incorporated into the model by the net present value at time t . Agents have outside option $\max_{a,t} \delta^t a \theta$. That is, if they do not participate in the mechanism, they can take any action a that they choose, and in any time period t . Thus the mechanism must satisfy an individual rationality constraint (IR), $\delta^{t(\theta)}[a(\theta)\theta + m(\theta)] \geq \max_{a,t} \delta^t a \theta$ for all $\theta \in \Theta$, to ensure that agents voluntarily participate. It must also satisfy an incentive compatibility constraint (IC), $\delta^{t(\theta)}[a(\theta)\theta + m(\theta)] \geq \delta^{t(\theta')}[a(\theta')\theta + m(\theta')]$ for all $\theta, \theta' \in \Theta$, to ensure that agents truthfully report their types.

I incorporate Pareto weights to allow for the possibility that the policy designer does not place equal weight on the agents' utility and environmental benefits. The agents' Pareto weight is ρ , and society's Pareto weight is $1 - \rho$. The social welfare function, W ,

¹²I restrict attention to mechanisms in which the transfer and action occur in the same time period. If the action were to occur before the transfer, that would make the contract more appealing to the non-additional agents and less appealing to the additional agents, which is the opposite of the desired screening. If the action were to occur after the transfer, a real world policy maker may encounter difficulties with enforcement or monitoring.

is:

$$W(\Gamma) = \int_{\Theta} \rho \delta^{t(\theta)} [a(\theta)\theta + m(\theta)] + (1 - \rho) d^{t(\theta)} [\omega a(\theta)] dF$$

The policy maker faces a budget constraint $B \in \mathbb{R}^{++}$, such that $\int_{\Theta} m(\theta) dF \leq B$, which is not sensitive to when funds are spent. Thus the policy maker is facing the problem of maximizing social welfare subject to the budget constraint, the incentive compatibility constraint, and the individual rationality constraint.

2.3 Time Delays as a Screening Tool

The mechanism described in this section separates out agents with positive net benefits from those with negative net benefits. That is, it screens non-additional agents, who are not changing their behavior by participating in the mechanism, from additional agents, who are engaging in costly behavior changes. The key to this separating mechanism is the differential impacts of delays on the two types. Although all agents have identical discount factors, their outside options differ. The policy maker can exploit this to separate the types.

We will start by examining the value of the outside option for the different types of agents. Recall that agents have the outside option of taking whichever action they choose, in whichever time period they choose (the action can only occur once). Thus the outside option is $\max_{a,t} \delta^t a \theta$. For types $\theta < 0$, this is maximized at $a = 0$ and in any time period. For types $\theta > 0$, this is maximized at $a = 1$ and $t = 0$. Thus, in the absence of a mechanism (or if they choose to not participate), agents with types $\theta < 0$ take the non-action of zero at any time period, and agents with types $\theta > 0$ take the full action of one in time period zero (the type zero agent is indifferent between everything).

Because the value of the outside option varies across types, time delays can have differential impacts. Consider a mechanism that offers the following three contracts for

agents to choose between.¹³ The first contract has $a = 1$, $t = 0$, and $m = 0$. That is, it offers agents an action of 1 in time period 0, and no transfer. The second contract offers $a = 1$, $t = \tau > 0$ for some $\tau \in T$, and $m = \mu > 0$. That is, it offers agents an action of 1 in time period τ , and a transfer of μ . The third contract offers $a = 0$, $t = 0$, and $m = 0$, which is action 0 in time period 0 and no transfer. Which agents will choose which contracts? As τ approaches infinity, the $\theta > 0$ types will all prefer contract 1 to contract 2. Because contract 1 is equivalent to their outside option, their IR constraint is satisfied. Agents with the lowest types, $\theta \in [\underline{\theta}, -\mu]$, will select contract 3, which is equivalent to their outside option. Finally, the middle agents $\theta \in [-\mu, 0]$ will select contract 2. Although contract 2 may have a limited present value due to τ being large, it will be their preferred option as their next best option has a value of zero. For finite τ , some of the lower positive types will select into contract 2 instead of contract 1.

Figure 2.1 illustrates the choice faced by non-additonal participants. The figure graphs their indifference curve on the τ - μ plane, where the indifference curve has the same utility as their outside option. Above this indifference curve, non-additonal agents prefer participating in the mechanism by receiving transfer μ and taking the action after delay τ . Below the indifference curve, they would rather take their outside option. Figure 2.2 illustrates the same choice for an additional agent. These agents will participate in the mechanism as long as the transfer is greater than the absolute value of their type, and with no regard to the waiting period. We can now begin to see how the policy maker can go about separating the agents: there are regions where additional agents will choose to participate but where non-additonal agents will not.

The following is a formal definition of a waiting period mechanism. For now, I restrict attention to a limited set of mechanisms that offer a single waiting period and

¹³ Although the model requires that agents are assigned a contract based on their type, because they can report any type this is equivalent to present them with a menu of contracts and then allowing them to choose.

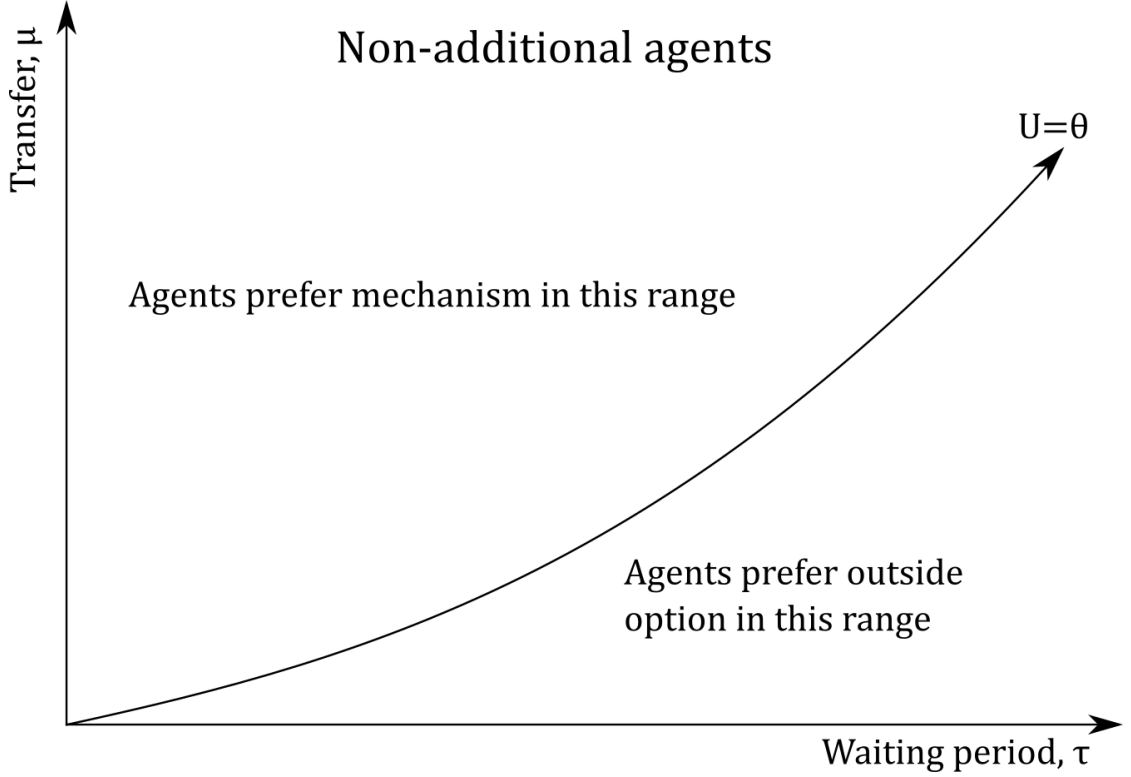


Figure 2.1. Outside option for non-additional agents

transfer, and demonstrate how it can screen agents.

Definition 2.1. A waiting period mechanism is one such that:

- i) for types $\theta \in [\underline{\theta}, \theta_1)$, the mechanism specifies $a(\theta) = 0$, $t(\theta) = 0$, and $m(\theta) = 0$
- ii) for types $\theta \in [\theta_1, \theta_2]$, the mechanism specifies $a(\theta) = 1$, $t(\theta) = \tau$ for some $\tau \in T$, and $m(\theta) = \mu$ for some $\mu > 0$
- iii) for types $\theta \in (\theta_2, \bar{\theta}]$, the mechanism specifies $a(\theta) = 1$, $t(\theta) = 0$, and $m(\theta) = 0$.

For some $\theta_1, \theta_2 \in \Theta$ such that $\theta_1 < \theta_2$.

Intuitively, this mechanism “excludes” participants with very low types (below θ_1), giving them their outside option and no transfer, which involves taking the non-action $a = 0$.¹⁴ For intermediate types between θ_1 and θ_2 , the agents receive a transfer

¹⁴The mechanism specifies $t = 0$ for these types. However, it would be equally effective to specify any other time period.

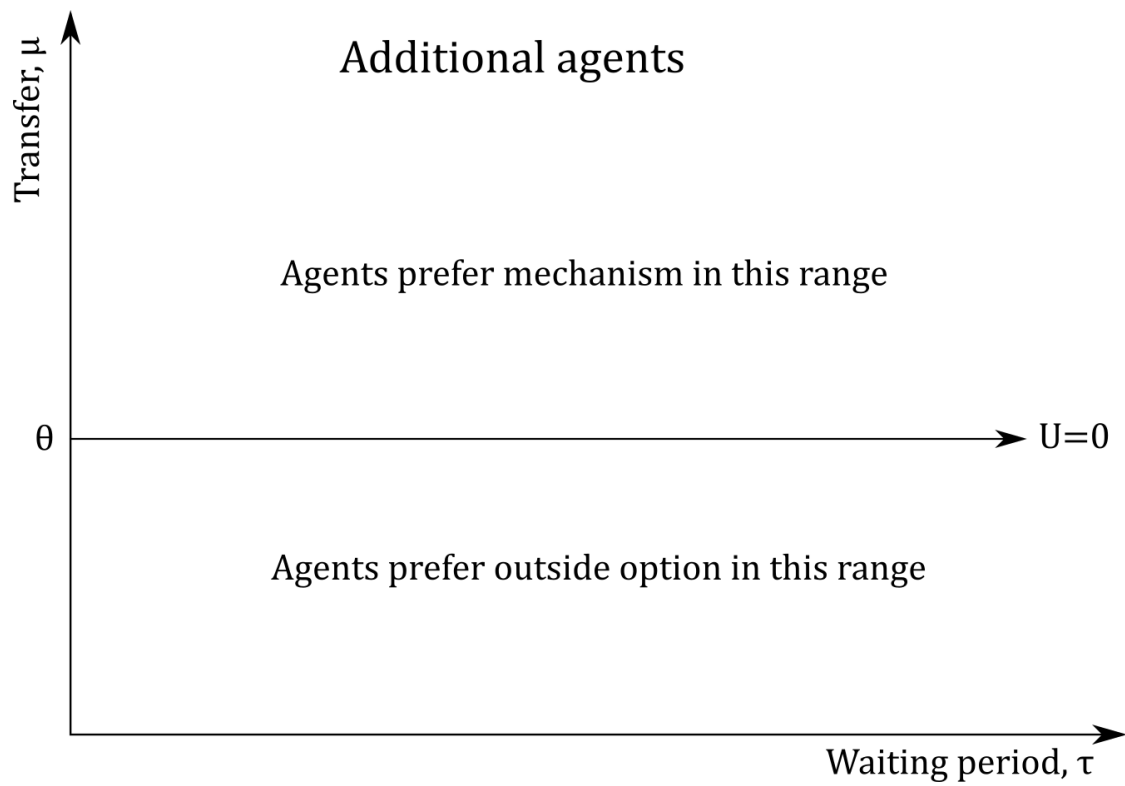


Figure 2.2. Outside option for additional agents

and take the action $a = 1$, but after a waiting period τ . High types (above θ_2) are also essentially “excluded,” receiving their outside option and no transfer. Note that although the mechanism is defined in a way that all agents participate, the same outcome could be achieved by a policy maker who offers a single, voluntary contract to agents.

Lemma 2.1. *A waiting period mechanism satisfies IC and IR for all types if and only if $\theta_1 = -\mu$ and $\theta_2 = \min[\frac{\delta^\tau}{1-\delta^\tau}\mu, \bar{\theta}]$.*

Proof. Consider a mechanism that offers agents a single, voluntary contract consisting of action 1, time period τ , and transfer μ . Agents can either choose this contract or take their outside option. I will demonstrate that under these conditions, only agents with types $[-\mu, \min[\frac{\delta^\tau}{1-\delta^\tau}\mu, \bar{\theta}]]$ will choose to participate, and all other agents will take their outside option. This can be easily seen by calculating the two agents who are indifferent to taking the contract or leaving it. On the negative side of the type space, the indifferent agent will have a type that satisfies $0 = \delta^{t(\theta)}[a(\theta)\theta + m(\theta)] = \delta^\tau[\theta + \mu]$. It follows that the lower cutoff agent will have type $-\mu$, and therefore $\theta_1 = -\mu$. Similarly, the cutoff on the positive side will satisfy $\theta = \delta^{t(\theta)}[a(\theta)\theta + m(\theta)] = \delta^\tau[\theta + \mu]$. It follows that the upper cutoff agent will have either type $\frac{\delta^\tau}{1-\delta^\tau}\mu$ or $\bar{\theta}$, and thus $\theta_2 = \min[\frac{\delta^\tau}{1-\delta^\tau}\mu, \bar{\theta}]$. Therefore any waiting period mechanism that specifies τ and μ will have the cutoffs described above. Furthermore, such a mechanism satisfies both the IC and IR constraints. $\square$

Thus any waiting period mechanisms that satisfies IC and IR can be characterized by two variables: μ and τ . Note that for $\tau = 0$, the bounds will be $\theta_1 = -\mu$ and $\theta_2 = \bar{\theta}$. Next, I will consider the degree of separation achieved by a waiting period mechanism.

Definition 2.2. *A waiting period mechanism achieves full separation if $\theta_2 = 0$.*

In other words, full separation involves only giving transfers to agents who are not

taking the same action as their outside option. This may be desired for various political reasons, distributional concerns, or budget concerns.

Proposition 2.1. *1. As $\tau \rightarrow \infty$, the incentive compatible and individually rational waiting period mechanism approaches full separation.*

2. As $\delta \rightarrow 0$, the incentive compatible and individually rational waiting period mechanism approaches full separation.

Proof. 1 follows immediately from $\lim_{\tau \rightarrow \infty} \frac{\delta^\tau}{1-\delta^\tau} \mu = 0$. 2 follows immediately from $\lim_{\delta \rightarrow 0} \frac{\delta^\tau}{1-\delta^\tau} \mu = 0$. $\square$

For a given waiting period mechanism that satisfies IC and IR, we can characterize the degree of separation of participants as a function of τ and μ . Recall that the cutoffs will be $\theta_1 = -\mu$ and $\theta_2 = \min[\frac{\delta^\tau}{1-\delta^\tau} \mu, \bar{\theta}]$ for a waiting period mechanism satisfying IC and IR. Thus agents with types $\theta \in [-\mu, 0]$ can be considered “additional” participants, meaning that they are taking action $a = 1$ which is different from their outside option. Similarly, agents with types $\theta \in [0, \min[\frac{\delta^\tau}{1-\delta^\tau} \mu, \bar{\theta}]]$ can be considered “non-additional” participants, because they are receiving a transfer μ but are not changing their behavior (relative to their outside option). It follows that the fraction of participants who are additional is:

$$r(\mu, \tau) = \frac{\int_{-\mu}^0 dF}{\int_{-\mu}^{\min[\frac{\delta^\tau}{1-\delta^\tau} \mu, \bar{\theta}]} dF} = \frac{P(-\mu \leq \theta \leq 0)}{P(-\mu \leq \theta \leq \min[\frac{\delta^\tau}{1-\delta^\tau} \mu, \bar{\theta}])}$$

As τ increases, $\frac{\delta^\tau}{1-\delta^\tau} \mu$ moves towards zero, decreasing the number of non-additional participants and increasing r . As δ increases, $\frac{\delta^\tau}{1-\delta^\tau} \mu$ increases, meaning that the number of non-additional participants increases and thus r decreases. Changes in μ have an ambiguous sign: increasing μ increases both the number of non-additional participants and the number of additional participants, and thus the effect will depend on F .

What if the policy maker has a goal for the quantity of environmental benefits? For example, the policy maker could have a goal of “converting 1000 acres of lawn into drought-resistant landscaping.” The following demonstrates how, as long as the policy maker has sufficient budget to pay for the additional agents, this goal can always be achieved, albeit with a potentially long waiting period. Although the proposition is framed in terms of total environmental benefits, the same idea applies to goals for additional benefits.

Proposition 2.2. *Suppose the policy maker has a goal of achieving at least $\bar{\omega}$ environmental benefits. That is, they have the goal of $\int_{\Theta} \omega a(\theta) \geq \bar{\omega}$. Define $\hat{\theta} \leq 0$ as either the value that satisfies $\int_{\hat{\theta}}^{\bar{\theta}} \omega dF = \bar{\omega}$ or zero, whichever is lower. Then the policy goal can be achieved using an incentive compatible and individually rational waiting period mechanism if and only if:*

- i) The budget satisfies $B > \int_{\hat{\theta}}^0 \theta dF$ if $\hat{\theta} < 0$ and*
- ii) The time delay, τ , satisfies $B \geq \int_{\hat{\theta}}^{\min[\frac{-\delta\tau}{1-\delta\tau}\hat{\theta}, \bar{\theta}]} -\hat{\theta} dF$.*

Proof. To show sufficiency, consider a waiting period mechanism such that *i)* and *ii)* hold. This does not fully describe the mechanism, so in addition, let $\mu = -\hat{\theta}$, and let τ satisfy *ii)* with equality. Note that the budget constraint, IC, and IR are all satisfied. This mechanism induces agents $\theta \in [\hat{\theta}, 1]$ to take the action $a = 1$, and by the definition of $\hat{\theta}$ the policy goal is therefore met.

To show necessity, consider first a budget that violates *i)*. That is, $B \leq \int_{\hat{\theta}}^0 \theta dF$ (and where $\bar{\omega}$ is such that $\hat{\theta} < 0$). For any τ chosen, the budget is insufficient to induce all agents on the interval $[\hat{\theta}, 0]$ to take action 1. In particular, any mechanism that satisfies the budget constraint cannot also satisfy the IR constraint for the types $\theta \in [\hat{\theta}, 0]$. Thus *i)* is necessary.

Next, consider a mechanism where the time delay, τ , does not satisfy *ii)*. It follows that

the mechanism must violate one of the following: the policy goal, the IR constraint, or the budget constraint. Thus *ii*) is necessary. $\square$

Intuitively, Proposition 2.2 is saying that in order to get meet a fixed environmental benefits goal, $\bar{\omega}$, two conditions must be met. First, there must be at least enough budget to pay all the additional agents and induce them to take an action of 1. If this is not the case, then achieving the desired environmental benefits goal would either violate the budget constraint or the IR constraint. Second, the time delay must be sufficiently large that enough non-additional agents are pushed out of the subsidy category, leaving sufficient budget for the additional agents. Note that this characterizes a trade off between time delays and budget: the higher the budget, the shorter the needed delay to accomplish a certain environmental benefits goal. Also note that to minimize the time delay needed to achieve the policy goal, the inequality in *ii*) should be set as an equality.

2.4 Optimal Time Delay Mechanisms

The previous section describes how a policy maker can use waiting periods to screen out non-additional agents. However, it says nothing about whether or not this is desirable from a social-welfare standpoint. This section examines socially optimal waiting period mechanisms. It is divided into two cases: one where agents are more patient than society, and the other where agents are less patient than society. I will also provide empirical evidence from the literature that agents are likely less patient than society.

2.4.1 Case 1: $\delta \geq d$

In this section, I consider the case where $\delta \geq d$. Although so far this paper has focused on a single type of mechanism—waiting period mechanisms with a single waiting period and transfer—I now extend my analysis to a more general set of mechanisms for

this case. In particular, this set of mechanisms places no restrictions on the functions t , a , and m .

I will start by considering the case where $\delta = d$. Because the action a is continuous on $[0, 1]$, and because $\delta' = d'$ is also on $(0, 1)$, it is without loss of generality to simply combine the two terms, which I will call $\hat{a}$. Note that the assumption that $d = \delta$ is critical here, as it must be the case that changes in the waiting period have the same impact on $\hat{a}$ for both agents and the rest of society (via the environmental benefits).

Thus a mechanism, $\hat{\Gamma}$, simply specifies $\hat{m}$ (similarly transformed by the discount rate) and $\hat{a}$, and no t . Letting $\hat{\Gamma}$ represent this new form of mechanism with $\hat{a}$ and $\hat{m}$, social welfare becomes:

$$W(\hat{\Gamma}) = \int_{\Theta} \rho [\hat{a}(\theta)\theta + \hat{m}(\theta)] + (1 - \rho)[\omega\hat{a}(\theta)]dF$$

We can already begin to see the intuition for why waiting periods will never be optimal in the case where $\delta = d$. For any $\hat{a}$, there is a continuum of possible a and t such that $\delta'a = \hat{a}$.¹⁵ In particular, for any $\hat{a}$ there exists a a and t that satisfy it such that $t = 0$. However, the budget constraint is on m , not $\hat{m}$, and longer time delays drive down $\hat{m}$ relative to m . Thus we will find in this section that a zero-length waiting period will be optimal. The following proofs formalize this idea, as well as demonstrating that the optimal mechanism is a subsidy.¹⁶

Note that the proofs in this section roughly follow the proofs in Section 2.2 of Börger et al. [2015], which in turn presents a standard set of tools and techniques used in mechanism design and screening.¹⁷ The proofs here have been adapted to fit the

¹⁵Technically, $\hat{a} = 1$ has only one possible a and t . All other $\hat{a}$ have a continuum of possible a and t . Note that for large $\hat{a}$, the maximum t is short, while small $\hat{a}$ allow a wider range of t .

¹⁶Alternate possible mechanisms would involve menus of contracts for agents and/or actions other than $a = 0$ and $a = 1$.

¹⁷Section 2.2 of Börger et al. [2015] covers optimal pricing for a single indivisible good. However, the primary proof techniques apply to other settings, including the one presented here.

particular policy scenario of this paper. In particular, they have been adapted to allow for type-dependent outside options and a budget constraint.¹⁸ Thus although the primary proof techniques may not be novel, the results for this policy setting are new.

I will start by characterizing the set of mechanisms that satisfy the incentive-compatibility constraints.

Lemma 2.2. *For any mechanism, $\hat{\Gamma}$, that satisfies IC, $\hat{a}(\theta)$ is weakly increasing in θ . I.e., if $\theta > \theta'$, then $\hat{a}(\theta) \geq \hat{a}(\theta')$.*

Proof. Let $\theta > \theta'$. IC implies that the following two inequalities must be true:

$$\hat{a}(\theta)\theta + \hat{m}(\theta) \geq \hat{a}(\theta')\theta + \hat{m}(\theta')$$

$$\hat{a}(\theta)\theta' + \hat{m}(\theta) \leq \hat{a}(\theta')\theta' + \hat{m}(\theta')$$

Subtracting the second equation from the first yields:

$$\hat{a}(\theta)\theta - \hat{a}(\theta)\theta' \geq \hat{a}(\theta')\theta - \hat{a}(\theta')\theta'$$

Which implies that:

$$\hat{a}(\theta)(\theta - \theta') \geq \hat{a}(\theta')(\theta - \theta')$$

Because $\theta - \theta' > 0$, it follows that $\hat{a}(\theta) \geq \hat{a}(\theta')$. Thus $\hat{a}$ is increasing in θ . $\square$

Let $U(\theta)$ be the maximum utility an agent can get by participating in the mechanism. That is, $U(\theta) = \max_{\theta' \in \Theta} \hat{a}(\theta')\theta + \hat{m}(\theta')$. For any given θ' , U is an increasing, affine, and convex function of θ . This implies that U is increasing and convex, which in turn implies that it is absolutely continuous. Additionally, by the envelope theorem $U'(\theta) = a(\theta)$.

¹⁸The standard screening problem has neither of these features. Although budget constraints are often seen in mechanism design problems, type-dependent outside options are less common.

Lemma 2.3. *For any mechanism, $\hat{\Gamma}$, satisfying IC, $U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \hat{a}(x)dx$.*

Proof. U is absolutely continuous, therefore by the fundamental theorem of calculus it is the integral of its derivative. $\square$

Lemma 2.3 now allow us to pin down the transfer function for any incentive compatible mechanism, as demonstrated in Lemma 2.4.

Lemma 2.4. *For any mechanism, $\hat{\Gamma}$, satisfying IC, it must be the case that:*

$$\hat{m}(\theta) = \hat{m}(\underline{\theta}) + \hat{a}(\underline{\theta})\underline{\theta} - \hat{a}(\theta)\theta + \int_{\underline{\theta}}^{\theta} \hat{a}(x)dx$$

Proof. We know that:

$$U(\theta) = \hat{a}(\theta)\theta + \hat{m}(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \hat{a}(x)dx$$

Solving for $\hat{m}(\theta)$ and substituting in for $U(\underline{\theta})$ yields:

$$\hat{m}(\theta) = \hat{m}(\underline{\theta}) + \hat{a}(\underline{\theta})\underline{\theta} - \hat{a}(\theta)\theta + \int_{\underline{\theta}}^{\theta} \hat{a}(x)dx$$

$\square$

We can now fully characterize the set of mechanisms that satisfy IC.

Proposition 2.3. *A direct mechanism, $\hat{\Gamma}$, satisfies IC if and only if:*

i) $\hat{a}$ is increasing in θ

ii) $\hat{m}(\theta) = \hat{m}(\underline{\theta}) + \hat{a}(\underline{\theta})\underline{\theta} - \hat{a}(\theta)\theta + \int_{\underline{\theta}}^{\theta} \hat{a}(x)dx$

Proof. We demonstrated necessity in Lemmas 2.2 and 2.4, so we only need to demonstrate sufficiency.

Consider type θ . We need to demonstrate that $U(\theta) \geq \hat{a}(\theta')\theta + \hat{m}(\theta')$ for any $\theta' \in \Theta$.

Start with:

$$U(\theta) \geq \hat{a}(\theta')\theta + \hat{m}(\theta')$$

This holds if and only if:

$$U(\theta) \geq \hat{a}(\theta')\theta - \hat{a}(\theta')\theta' + \hat{a}(\theta')\theta' + \hat{m}(\theta')$$

Which is true if and only if:

$$U(\theta) \geq \hat{a}(\theta')\theta - \hat{a}(\theta')\theta' + U(\theta')$$

Which holds if and only if:

$$U(\theta) - U(\theta') \geq (\theta - \theta')\hat{a}(\theta')$$

Substituting in for U , we get that this holds if and only if:

$$\int_{\theta'}^{\theta} \hat{a}(x)dx \geq \int_{\theta'}^{\theta} \hat{a}(\theta')dx$$

Which must be true because a is increasing in θ . □

Now that I have characterized the set of mechanisms, $\hat{\Gamma}$, that satisfy the IC constraint, I can turn attention to the IR constraint. The IR constraint is slightly trickier than the IC constraint, because we are straying from a standard mechanism design type problem by allowing the outside options to vary by type. My proof deals with the type-dependent outside options in the following way. First, I will characterize the set of IR-compatible mechanisms for types $\theta \leq 0$. These types all have the same outside

option, valued at 0. Thus it will be simple to show that as long as the lowest type, $\underline{\theta}$ is willing to participate, the types $\theta \in (\underline{\theta}, 0]$ will also be willing to participate. The other half of the type space, $\theta \in (0, \bar{\theta}]$ is trickier, as value of their outside options depends on θ . For these types, I will proceed by simply ignoring the IR constraint. After I have characterized the optimal mechanism, I will return to this constraint and verify that it is indeed satisfied and was not binding.

I start by demonstrating that transfers must always be positive.

Lemma 2.5. *For any mechanism, $\hat{\Gamma}$, satisfying IR, it must be the case that $\hat{m}(\theta) \geq 0$ for all $\theta \in \Theta$.*

Proof. The IR constraint states that $\hat{a}(\theta)\theta + \hat{m}(\theta) \geq \max_{\hat{a}} \hat{a}\theta$ for all $\theta \in \Theta$. Because $\hat{a}(\theta)\theta \leq \max_{\hat{a}} \hat{a}\theta$, it follows that $\hat{m}(\theta) \geq 0$. $\square$

Next, I must quantify the value of the outside option for all agents. Recall that the value of the outside option is $\max_{\hat{a}} \hat{a}\theta$. That is, agents may choose any combination of actions and time delays, which is equivalent to selecting any $\hat{a}$. For types $\theta < 0$, this involves taking $\hat{a} = 0$, yielding an outside option value of zero. For types $\theta > 0$, this involves taking $\hat{a} = 1$, yielding an outside option value of θ . Note that type $\theta = 0$ is indifferent between all actions $\hat{a} \in [0, 1]$, and has an outside option value of zero.

Next, I characterize the set of IC-compatible mechanisms that also satisfies IR for types $\theta \in [\underline{\theta}, 0]$.

Lemma 2.6. *If a mechanism, $\hat{\Gamma}$, that satisfies IC has $U(\underline{\theta}) \geq 0$, then it satisfies IR for types $\theta \in [\underline{\theta}, 0]$. Note that $U(\underline{\theta}) \geq 0$ is equivalent to $\hat{m}(\underline{\theta}) \geq \hat{a}(\underline{\theta})\underline{\theta}$.*

Proof. The value of the outside option for all types $\theta \in [\underline{\theta}, 0]$ is zero. Recall that U is increasing in θ for all mechanisms satisfying IC. It follows that if IR is satisfied for the lowest type, $\underline{\theta}$, then it will be satisfied for all higher types with the same outside option value. $\square$

Although I have not fully characterized the set of IC and IR compatible mechanisms (I have ignored the IR constraint for part of the distribution of types), I now proceed to describing the optimal mechanism in this set of IC and partially-IR compatible mechanisms. Once I have characterized the set of optimal mechanisms, I will return to the remaining IR constraint and verify that it is indeed satisfied.

Recall that the policy maker is subject to a budget constraint, and that social welfare is:

$$W(\hat{\Gamma}) = \int_{\Theta} \rho[\hat{a}(\theta)\theta + \hat{m}(\theta)] + (1 - \rho)[\omega\hat{a}(\theta)]dF$$

Thus the policy maker's problem is to maximize social welfare subject to the IC, IR, and budget constraints. I will assume that the budget constraint is binding. If the budget constraint does not bind, then the design problem is less interesting as the policy designer is not facing any trade offs between budget and social welfare. Additionally, if the budget constraint does not bind, it can be made to bind by simply offering unconditional lump sum transfers to the agents.

Because the budget constraint is binding, it must be the case that for any optimal mechanism, the lowest type has utility zero. In particular, recall that to satisfy the IR constraint, $U(\underline{\theta}) \geq 0$. Also, recall that all other transfers depend on the utility level of the lowest agent for any mechanism satisfying IC. Thus, if $U(\underline{\theta})$ is strictly greater than zero, then the policy maker can save money by simply reducing everyone's transfers (subject to the condition on $\hat{m}$ given in Proposition 2.3) until the lowest type reaches zero utility. Because the budget constraint is binding, the policy maker will wish to do this. Note that the transfers cannot be lowered past this point because that would violate the IR constraint. Thus it must be the case that $U(\underline{\theta}) = 0$, or equivalently, that $\hat{m}(\underline{\theta}) = \hat{a}(\underline{\theta})\underline{\theta}$.

Definition 2.3. A subsidy mechanism is one such that, for some $\hat{\theta} \in [\underline{\theta}, 0]$:

i) for $\theta < \hat{\theta}$, $\hat{a}(\theta) = 0$ and $\hat{m}(\theta) = 0$

ii) for $\theta \geq \hat{\theta}$, $\hat{a}(\theta) = 1$ and $\hat{m}(\theta) = -\hat{\theta}$.

In other words, a subsidy mechanism is one where the mechanism designer offers a single type of contract specifying the action $\hat{a} = 1$ and a subsidy amount. Agents are then free to take this contract or their outside option.

Proposition 2.4. *A subsidy mechanism maximizes social welfare subject to the IC and IR constraints.*

Proof. See 2.7.1. □

The general idea behind this proof is that social welfare is linear in the actions, $\hat{a}$. This follows from the fact that the sum of the transfers can be replaced with the budget cap, eliminating them from the social welfare function.¹⁹ Because social welfare is linear in the actions and the set of budget constrained, IC and partial IR compatible mechanisms is compact and convex, an extreme value theorem can be applied to demonstrate that social welfare must be maximized by extreme values of $\hat{a}$.²⁰ It can then be shown that an action $\hat{a}(\theta)$ is an extreme point if and only if $\hat{a}(\theta) \in \{0, 1\}$ for all θ . It then follows that the optimal mechanism must be a subsidy mechanism, as those are the only incentive compatible mechanisms with $\hat{a} \in \{0, 1\}$ for all agents.

We now need to go back and verify that the IR constraint does not bind for the agents with types $\theta \in (0, \bar{\theta}]$. From the mechanism description, we can see that each of these agents is taking action $\hat{a}(\theta) = 1$ and getting a weakly positive transfer. Because their outside option is taking that same action, but getting no transfer, their IR constraint must be satisfied.

We also need to revisit the budget constraint. So far, we have determined that the optimal mechanism is a subsidy. However, we have not specified the level of that

¹⁹This follows from the fact that the optimal wait must be zero.

²⁰This is the same general idea as maximizing a simple linear function over a closed and bounded interval on $\mathbb{R}$: the maximum will always be one of the end points. In this case, we are simply maximizing over a compact and convex set of functions instead.

subsidy. Adding in the budget constraint will pin down the subsidy level. Recall that we assumed that the budget constraint was binding. Thus the sum of transfers, where the transfer must equal $-\hat{\theta}$, must satisfy the budget constraint with equality: $\int_{\hat{\theta}}^1 -\hat{\theta} dF = B$. In other words, simply set the subsidy as high as the budget allows. Note that the transfer level also pins down the cutoff type: that is, $\hat{\theta}$, the lowest type who will take the subsidy and action $\hat{a} = 1$.

Lemma 2.7. *The following mechanism, $\hat{\Gamma}$, maximizes social welfare subject to the budget constraint:*

- i) Let $\hat{\theta}$ be such that $\int_{\hat{\theta}}^1 -\hat{\theta} dF = B$.
- ii) For $\theta < \hat{\theta}$, $\hat{a}(\theta) = 0$ and $\hat{m}(\theta) = 0$.
- iii) For $\theta > \hat{\theta}$, $\hat{a}(\theta) = 1$ and $\hat{m}(\theta) = -\hat{\theta}$.

Proof. Follows immediately from Proposition 2.4 and the budget constraint. $\square$

We now need to relate this back to the original scenario, where instead of $\hat{a}(\theta)$ and $\hat{m}(\theta)$ we have $\delta^{t(\theta)}a(\theta)$ and $\delta^{t(\theta)}m(\theta)$. In the optimal mechanism, $\hat{a}$ only takes on the values zero and one. Because $\delta < 1$, this means that the waiting period for agents taking action $\hat{a} = 1$ must always be zero. The waiting period for agents taking action $\hat{a} = 0$ is irrelevant, and can be anything. For simplicity, we will also assign these agents $t = 0$.

Proposition 2.5. *The following mechanism, Γ , maximizes social welfare subject to the budget constraint:*

- i) Let $\hat{\theta}$ be such that $\int_{\hat{\theta}}^1 -\hat{\theta} dF = B$.
- ii) For $\theta < \hat{\theta}$, $a(\theta) = 0$, $t(\theta) = 0$, and $m(\theta) = 0$.
- iii) For $\theta > \hat{\theta}$, $a(\theta) = 1$, $t(\theta) = 0$, and $m(\theta) = -\hat{\theta}$.

Proof. Follows immediately from Proposition 2.7 and the definition of $\hat{a}$. $\square$

Thus for the case where $d = \delta$, social welfare will always be maximized with no time delay and all actions occurring immediately. Furthermore, this outcome can be achieved by a basic subsidy mechanism with no time delays.

I now turn my attention to the case where $\delta > d$. We saw previously that 1) no time delay is optimal if $\delta = d$, and 2) that it is more difficult to screen agents with time delays when they are more patient. Add to this the fact that agents prefer having their payments and actions (assuming they have positive net value) sooner rather than later, and we can immediately see that no time delay is optimal if $\delta > d$.

Proposition 2.6. *If $\delta > d$, the following mechanism maximizes social welfare:*

- i) Let $\hat{\theta}$ be such that $\int_{\hat{\theta}}^1 -\hat{\theta} dF = B$.
- ii) For $\theta < \hat{\theta}$, $a(\theta) = 0$, $t(\theta) = 0$, and $m(\theta) = 0$.
- iii) For $\theta > \hat{\theta}$, $a(\theta) = 1$, $t(\theta) = 0$, and $m(\theta) = -\hat{\theta}$.

Proof. First, consider the optimal t to maximize environmental benefits. At $\delta = d$, it is optimal to not screen. If we now move to $\hat{\delta} > d$, for any given $t(\theta) > 0$ environmental benefits are strictly lower than they were before (because delays are less effective at screening). Note that environmental benefits are the same for $t(\theta) = 0$ across δ and $\hat{\delta}$. Therefore because $t(\theta) = 0$ was better with δ , and we have made all other t relatively worse by moving to $\hat{\delta}$, $t(\theta) = 0$ must still be best for environmental benefits.

Next, consider maximizing agent utility. Agents always prefer not waiting (and the net present value of the transfers lowers as the delay increases), therefore $t = 0$ maximizes agent utility.

Thus $t = 0$ maximizes both agent utility and environmental benefits when $\delta > d$. Therefore it must maximize social welfare.

Once we know that $t = 0$ is optimal, we can use the exact same proof techniques we used for the $\hat{\Gamma}$ mechanism earlier in this section and prove that a subsidy mechanism is

optimal. □

It is useful to consider the intuition behind this result. In particular, it is not immediately obvious that no delay is optimal in this case, particularly if the Pareto weight on the agents is very low or zero (meaning most weight is put on the environmental benefits). The intuition for the result comes from consideration of the incentive compatibility constraint: as δ grows, waiting periods are less effective at screening out non-additional agents. In particular, whenever $\delta > d$, the cost incurred by society due to delayed environmental benefits never outweighs the increased quantity of environmental benefits. Thus when $\delta > d$, it is never optimal to delay.

2.4.2 Case 2: $\delta < d$

In this next section, I consider the case where the discount factors satisfy $\delta < d$. This is the case where the agents are less patient than society. This is the more likely case, as consumers tend to have higher discount rates than government agencies.²¹

Of particular relevance is a large literature using revealed preferences to estimate consumer discount rates based energy efficient appliance purchase data. One of the original papers in this literature, Hausman [1979] estimates a 20% discount rate for consumers using data on energy-efficient appliance purchases. Corum and O'Neal [1982] estimate average discount rates between 30% and 41% using data on energy-saving home construction practices. Meier and Whittier [1983] analyze consumer purchases of energy efficient refrigerators and find that 2 out of every 5 of consumers have discount rates above 60%. More recent work by Wasi and Carson [2013] analyzes water heater purchases in Australia and finds a high degree of heterogeneity in consumer discount rates: although they find a median discount rate of 9.4%, individual consumers have

²¹Note that δ and d are discount factors. Thus $\delta < d$ implies that agents have a higher discount rate than society.

rates ranging from below 2% to above 40%. It is widely accepted that the high consumer discount rates reflect consumers' liquidity constraints, limited access to credit, and the high interest rates they face when borrowing.²²

On the other hand, government agencies tend to use relatively low discount rates. For example, the Office of Management and Budget recommends discount rates of between 1% and 7% for cost-benefit analyses depending on the context. The Congressional Budget Office typically uses a 2% discount rate (Zerbe et al. [2002]). Even the highest of these rates is lower than the estimated median discount rates for consumers.

For this analysis, I once again return to the limited class of waiting period mechanisms that offer a single waiting period and a single transfer to all agents, who may either take it or leave it. This simpler mechanism also restricts a to the set $\{0, 1\}$. That is, a is binary rather than continuous. Future extensions of this paper will include an appendix with a full analysis of optimal mechanisms across the set of all mechanisms. However, these more complex mechanisms can sometimes be overly complicated for implementation as real-world policies.²³ The simpler mechanisms presented in this section represent an improvement in social welfare relative to a mechanism without a waiting period, while remaining easy for policy makers to implement.

A waiting period involves trade offs: although the budget may be stretched (via money saved by excluding non-additional participants) to induce more agents to undertake the green action, the environmental (as well as private) benefits of these actions are delayed. This leads to the natural question of how long of a waiting period is optimal? Recall that I am currently restricting attention to mechanisms with a single option for agents. That is, agents all have the option between waiting period τ and transfer μ , or

²²For example, Bankrate.com, a website that tracks credit card interest rates, reports that the average consumer credit card APR was 16.28% in November 2016. For more information on the drivers of consumer discount rates, see Carson and Roth Tran [2009].

²³Alternate mechanisms could be difficult or costly to implement. For instance, they could involve an infinite set of contracts that agents select from.

taking their outside option. In addition to being more realistic from a policy perspective, this limited class of mechanisms is more tractable, and the optimal mechanism can be easily computed.

The optimal waiting period mechanism will maximize social welfare subject to the IC constraint, IR constraint, and budget constraint. Using the definition of a waiting period mechanism and substituting in for the IC and IR compatible cutoffs, this is equivalent to maximizing the following:

$$\max_{\tau, \mu} \int_{-\mu}^{\min[\frac{\delta^\tau}{1-\delta^\tau}\mu, \bar{\theta}]} \rho[\delta^\tau(\theta + \mu)] + (1 - \rho)d^\tau \omega dF + \int_{\min[\frac{\delta^\tau}{1-\delta^\tau}\mu, \bar{\theta}]}^{\bar{\theta}} \rho\theta + (1 - \rho)\omega dF$$

subject to the budget constraint. If we assume that the budget constraint binds (holds with equality), then this becomes a maximization problem in one variable, τ . If the budget constraint does not bind, then we can artificially induce it to bind by offering lump sum transfers to all agents, not contingent on any actions.

This maximization problem is not analytically tractable. However, it can be easily solved numerically for any set of parameters and any distribution F . I examine one particular case and its optimal waiting periods in more detail—the case where F is the uniform distribution on $[-1, 1]$ —in 2.7.2.

To understand the intuition behind a strictly positive optimal waiting period, it is useful to consider an extreme example. In particular, consider the case where $\rho = 0$ (all Pareto weight is on the environmental benefits, and none on the agents) and $d = 1$ (society is perfectly patient). What is the optimal policy in this case? Because $d = 1$, society is not harmed at all by waiting. Thus the only objective is to maximize the number of agents taking action $a = 1$. Under this extreme scenario, the optimal wait will either be at least the minimum wait required to get all additional participants to participate (e.g., the minimum wait needed for the budget to subsidize all additional agents) or infinity,

depending on the size of the budget. Of course, this is an extreme example and not particularly realistic. However, it can shed light on the intuition behind strictly positive optimal waiting periods.

Additionally, the policy maker's maximization problem can be used to analyze comparative statics. In particular, how does the optimal waiting period change as the underlying parameters change?

First, consider changes in the budget. As B increases, the same set of actions can be achieved with a shorter wait time and higher total transfers (that is, less screening needs to occur). Because waiting periods serve only to separate agent types (and are otherwise detrimental to social welfare), increasing the budget cannot increase the optimal waiting period. It follows that as the budget, B , increases, the optimal waiting period weakly decreases.

Next, consider changes in d , the discount factor for society. Changes in d —all else equal—have no impact on agent utility or on the ability to screen agents. However, increases in d do make waiting less costly from the standpoint of environmental benefits. That is, as d increases, waiting becomes less costly for society. It follows that an increase in d will produce a weakly longer optimal waiting period.²⁴

Comparative statics on δ are less straightforward. I already demonstrated that if $\delta \geq d$, the optimal wait is always zero. However, what happens to the optimal wait as δ varies in the region where $\delta < d$? Several things happen here. First, as δ increases, screening becomes more difficult. The more patient agents are, the less effective waiting periods are. This has an impact on environmental benefits via changes in the trade off between the costs and benefits of waiting. Second, as δ increases, agent utility is less severely impacted by waiting periods. In particular, for a given waiting period, the budget

²⁴Note that changes in d can shift us from case 1—where $\delta \geq d$ —to case 2—where $\delta < d$. This is consistent with the comparative static above.

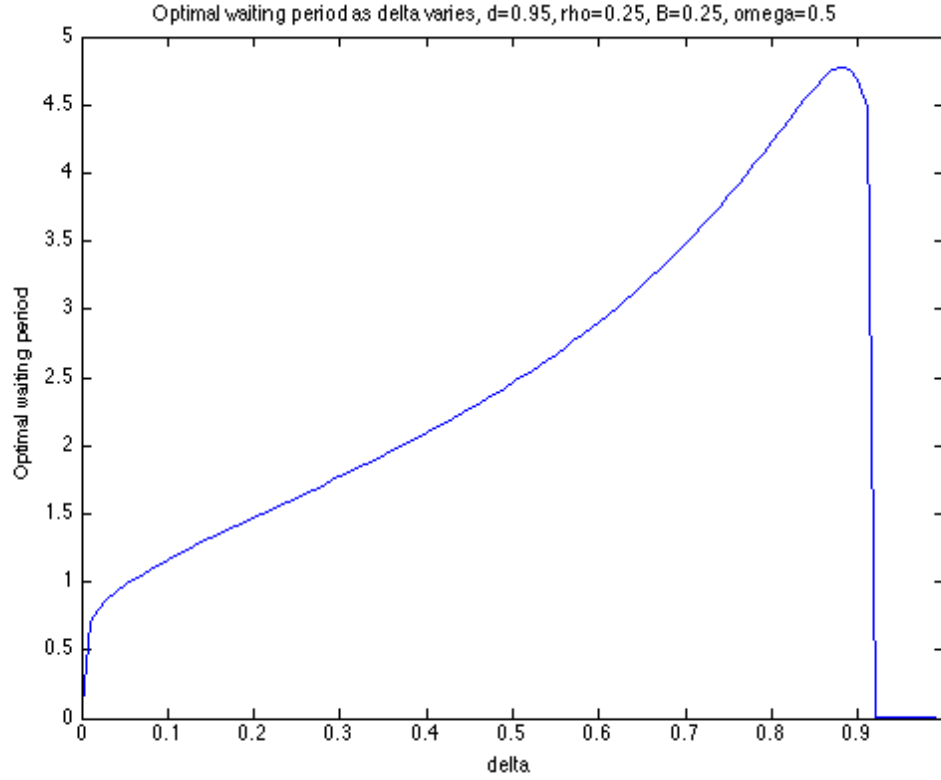


Figure 2.3. Optimal waiting period as δ varies

becomes more valuable, thereby increasing agent utility all else equal. Thus the costs of waiting (on the agents) are reduced as δ increases. Results of numerical simulations indicate that either of these impacts can outweigh the other. This means that changes in δ can move the optimal waiting period in either direction. Figure 2.3 demonstrates how the optimal waiting period changes as δ varies from 0 to 1 for the case where $F = U[-1, 1]$. The graph assumes $d = 0.95$, $\rho = 0.25$, $B = 0.25$,²⁵ and $\omega = 0.5$. Note that it is not monotone, and that the optimal wait drops to zero before δ surpasses d .

Finally, consider changes in ρ , the Pareto weights. As ρ increases, weight is shifted from the social benefits to the agents' utility. Agents always prefer shorter waiting periods. In particular, the shorter the waiting period, the higher the net present value of

²⁵Note that the budget can be interpreted as the average payment to agents.

the budget constraint. Thus cumulative agent utility is higher with shorter waits. Thus as ρ increases, the optimal waiting period weakly decreases.

2.5 Policy Discussion

Waiting periods can be a cost-effective screening tool. Section 2.3 demonstrated how waiting periods can be used to separate additional agents from non-additional agents. In general, adding a waiting period to a policy is fairly inexpensive for the policy maker. Although it may slightly increase administrative costs, it will often be less expensive than more traditional screening methods that rely on observable characteristics.²⁶

Waiting periods are most effective at screening when agents are impatient. That is, the more impatient agents are, the shorter the required waiting period to achieve a given level of separation between agent types. If agents are very patient, in particular if they are more patient than society, it is never optimal to implement a non-zero waiting period. However, this scenario is unlikely due to the fact that consumers tend to have higher discount rates than society.

Waiting periods may be less desirable if the environmental benefits of the action are urgent (low d). For instance, consider a city attempting to decrease water use by its citizens. If the city well is about to run dry, they will not wish to implement a six month waiting period for a low-flow shower head subsidy program. However, shorter waiting periods will require higher budgets to achieve a given number of agents taking the green action. Conversely, if the environmental benefits are not urgent (high d), waiting periods can be used to screen agents and achieve more people taking the green action for a given budget.

Waiting periods may also be desirable if a subsidy program has distributional

²⁶Costs for other screening methods include processing large volumes of paperwork, verifying agents' characteristics, and potentially monitoring characteristics.

concerns associated with it. For example, if the majority of non-additional agents are thought to be high income, it may be desirable to use a waiting period to screen out those agents so that they do not receive a subsidy. This would leave more subsidies for lower-income, additional agents. This will be true whenever non-additionality is correlated with another feature that makes the policy maker wish to not subsidize these types. The same reasoning may apply to political pressures facing the policy maker.

Policy makers may be concerned that not all agents have the same discount factor, as assumed by the model in this paper. In particular, if the action has costs as soon as it is taken, and a stream of benefits afterwards,²⁷ they may worry that non-additional agents have higher discount factors than additional agents. Recall that the screening relies on the discount factor, but that in particular, it relies on the discount factor of the non-additional agents. Thus the policy maker will wish to consider the non-additional agents' discount factor when setting a waiting period screening policy. Future extensions of this paper will include an analysis of the scenario where agents have heterogenous discount factors.

2.6 Conclusion

I have shown how waiting periods can be used to screen in programs that induce consumers to take an environmentally beneficial action that also has private benefits and costs. I have also demonstrated that if agents are more patient than society (an unlikely scenario), waiting periods are never optimal. If agents are less patient than society (which is supported by empirical evidence on discount rates), waiting periods may be optimal, and the optimal wait can be easily calculated numerically for the given setting. Future work on this topic will expand my analysis of the case where agents are less patient than society to include a broader set of mechanisms. However, the mechanisms presented in

²⁷For instance, many efficient appliances have a high up front cost followed by a stream of utility bill savings.

this version of the paper are simple and easy for a real-world policy maker to implement. Additionally, they represent an increase in social welfare relative to a policy with no waiting period.

Bureaucratic delays and long waits are typically viewed as an inefficiency, and as a sign that the agency in question is failing its constituents. This paper presents a different view: that delays can in fact be used to increase social welfare by acting as a screening mechanism.

2.7 Appendices

2.7.1 Proof of Proposition 2.4

The proof in this appendix follows the techniques in Section 2.2 of Börger et al. [2015].²⁸ The proofs are adapted to allow for several peculiarities of this setting, with the largest adaptations in this appendix being for the budget constraint.

Let $\mathcal{F}$ be the set of all bounded functions from Θ into $\mathbb{R}$. Give $\mathcal{F}$ a linear structure and the L^1 norm and the metric induced by this norm. Note that $\mathcal{F}$ is a vector space. Let $\mathcal{A}$ denote the set of weakly increasing functions such that $\hat{a} : \Theta \rightarrow [0, 1]$, e.g., the set of functions that are incentive compatible (here, we are assuming that $\hat{a}$ determines $\hat{m}$ as required by the IC constraint). Note that $\mathcal{A} \subset \mathcal{F}$. Lemma 2.6 in Börger et al. [2015] demonstrates that $\mathcal{A}$ is compact and convex.²⁹

I will use the following definition of linearity.³⁰

Definition 2.4. *Let $f : X \rightarrow Y$ be a function between two vector spaces X and Y . The function f is linear if the following two statements are true:*

²⁸Section 2.2 of Börger et al. [2015] covers optimal pricing for a single indivisible good. However, the primary proof techniques apply to other settings, including the one presented here.

²⁹ $\mathcal{A}$ is convex because the convex combination of two increasing functions is increasing. Compactness is implied by Helly's selection theorem (see Rudin [1976], p.167) and the bounded convergence theorem of Lebesgue integration (see Rudin [1976], p.322).

³⁰For similar definitions, see, for example, Definition 9.4 in Rudin [1976].

- i) $f(bx) = bf(x)$ for all $b \in \mathbb{R}$ and all $x \in X$
 ii) $f(x + x') = f(x) + f(x')$ for all $x, x' \in X$

The maximization problem is:

$$\max_{\hat{a}} \int_{\Theta} \rho \hat{a}(\theta) \theta + \hat{m}(\theta) + (1 - \rho) \omega \hat{a}(\theta) dF$$

subject to IC, IR, and the budget constraint.

Recall that in the set of IC and IR compatible mechanisms, $\hat{m}(\theta) = \hat{m}(\underline{\theta}) + \hat{a}(\underline{\theta})\underline{\theta} - \hat{a}(\theta)\theta + \int_{\underline{\theta}}^{\theta} \hat{a}(x)dx$, and that because the budget constraint binds, we want to set $U(\underline{\theta}) = 0$. Thus $\hat{m}(\theta) = -\hat{a}(\theta)\theta + \int_{\underline{\theta}}^{\theta} \hat{a}(x)dx$.

Now, we must demonstrate that the maximand is linear in $\hat{a}$. To do this, first note that $\hat{m}$ is linear in $\hat{a}$. In particular, if you multiply $\hat{a}$ by a constant b , you get a transfer $b\hat{m}$. If you add $\hat{a}'$ to $\hat{a}$, the resulting transfer is $\hat{m} + \hat{m}'$. It follows that the entire maximand is linear in $\hat{a}$. Note that the maximand is also continuous in $\hat{a}$.

Although the maximand is linear and continuous in $\hat{a}$, it is still subject to a budget constraint. Before we can apply an extreme point theorem, we need to demonstrate that the maximization problem as a whole is linear in $\hat{a}$ by substituting in for the budget constraint. This is not as simple as it initially may appear, as the budget constraint is on m , not $\hat{m}$, and $\hat{m}$ depends on t which we have been ignoring.

First, note that $\hat{a}$ does two things: it pins down $\hat{m}$, and it pins down a relationship between a and t . However, for any $\hat{a}$, there exists a (a, t) such that $t = 0$. In particular, this set of a and t will maximize the net present value of $\hat{m}$ for a given m . Because the budget constraint is on m , not $\hat{m}$, and the budget constraint binds, it is optimal to select the t that maximizes $\hat{m}$ for a given m . This means that the optimal mechanism will select $t = 0$. It follows that $\hat{m} = m$, and that $\int_{\Theta} \hat{m}(\theta) dF = \int_{\Theta} m(\theta) dF = B$. We can now substitute into the maximand and eliminate the budget constraint. Thus the maximization

problem becomes:

$$\max_{\hat{a}} \int_{\Theta} \rho \hat{a}(\theta) \theta + (1 - \rho) \omega \hat{a}(\theta) dF + B$$

Because B is a constant, we can drop it from the maximand without impacting the solution. The resulting maximand is both linear and continuous in $\hat{a}$ and is not subject to a budget constraint. We can now apply an extreme point theorem.

The following definition is taken from Definition 2.4 in Börgers et al. [2015].

Definition 2.5. *If C is a convex subset of a vector space X , then $x \in C$ is an extreme point of C if for every $y \in X$ such that $y \neq 0$, it is the case that either $x + y \notin C$ or $x - y \notin C$.*

The extreme point theorem³¹ states that if X is a compact, convex, subset of a normed vector space, and if $f : X \rightarrow \mathbb{R}$ is a continuous linear function, then the set E of extreme points of X is nonempty and there exists an $e \in E$ such that $f(e) \geq f(x)$ for all $x \in X$.

This implies that any extreme point $\hat{a}$ that maximizes social welfare among all extreme points in $\mathcal{A}$ will also maximize social welfare among all functions in $\mathcal{A}$. Thus we only need to consider the set of mechanisms where $\hat{a}$ is an extreme point.

Lemma 2.8. *A function $\hat{a} \in \mathcal{A}$ is an extreme point of $\mathcal{A}$ if and only if $\hat{a}(\theta) \in \{0, 1\}$ for almost all $\theta \in \Theta$.*

Proof. This proof follows the proof of Lemma 2.7 in Börgers et al. [2015].

Consider an $\hat{a}$ such that $\hat{a}(\theta) \in \{0, 1\}$ for almost all $\theta \in \Theta$. Let $\hat{a}'(\theta)$ be such that $\hat{a}'(\theta) \neq 0$ for some θ . Now consider the case where $\hat{a}'(\theta) > 0$. If $\hat{a}(\theta) = 0$, then $\hat{a}(\theta) - \hat{a}'(\theta) < 0$ which implies that $\hat{a} - \hat{a}' \notin \mathcal{A}$. If $\hat{a}(\theta) = 1$, then $\hat{a}(\theta) + \hat{a}'(\theta) > 1$ which implies that $\hat{a} + \hat{a}' \notin \mathcal{A}$. The case of $\hat{a}'(\theta) < 0$ is similar. Thus $\hat{a}$ is an extreme

³¹See Proposition 2.4 in Börgers et al. [2015] and Ok [2007] p.658.

point of $\mathcal{A}$.

Next consider a function $\hat{a}$ such that $\hat{a}(\theta) \notin \{0, 1\}$ for some θ . Let θ' be such that $\hat{a}(\theta') \in (0, 1)$. Let $\hat{a}'$ be such that $\hat{a}'(\theta) = \hat{a}(\theta)$ if $\hat{a}(\theta) \leq 0.5$, and let $\hat{a}'(\theta) = 1 - \hat{a}(\theta)$ if $\hat{a}(\theta) > 0.5$. Note that $\hat{a}' \neq 0$. Now consider $\hat{a} + \hat{a}'$. $\hat{a}(\theta) + \hat{a}'(\theta) = 2\hat{a}(\theta) \in (0, 1)$ for $\hat{a}(\theta) \leq 0.5$ and $\hat{a}(\theta) + \hat{a}'(\theta) = 1$ for $\hat{a}(\theta) > 0.5$. Now consider $\hat{a} - \hat{a}'$. $\hat{a}(\theta) - \hat{a}'(\theta) = 0$ for $\hat{a}(\theta) \leq 0.5$, and $\hat{a}(\theta) - \hat{a}'(\theta) = 2\hat{a}(\theta) - 1 \in (0, 1)$ for $\hat{a}(\theta) > 0.5$. Thus $\hat{a}$ is not an extreme point of $\mathcal{A}$. $\square$

Definition 2.6. Let $\hat{\theta} \in [\underline{\theta}, 0]$. A mechanism is a subsidy mechanism if:

- i) for $\theta < \hat{\theta}$, $\hat{a}(\theta) = 0$ and $\hat{m}(\theta) = 0$
- ii) for $\theta \geq \hat{\theta}$, $\hat{a}(\theta) = 1$ and $\hat{m}(\theta) = -\hat{\theta}$

In other words, a subsidy mechanism sets a cutoff type, $\hat{\theta}$. The cutoff type is weakly below zero, meaning they are incurring a cost to undertake the green action (it would not be optimal to set the cutoff type above zero, as those types would willingly take the green action on their own). All types with benefits higher than the cutoff type get a subsidy equal to the cutoff type's cost, while all types below the cutoff do not take the action and do not receive a transfer. Note that as defined, a subsidy mechanism satisfies IC and IR. Also note that the cutoff type, $\hat{\theta}$, is indifferent between the two options and can be assigned to either group.

Lemma 2.9. If $\hat{a}$ is an extreme point of $\mathcal{A}$ that satisfies IC, then the mechanism is a subsidy mechanism.

Proof. If $\hat{a}$ is an extreme point of $\mathcal{A}$, then $\hat{a} \in \{0, 1\}$. Because the mechanism satisfies IC, it must be the case that $\hat{a}$ is increasing in θ . Thus there must be some cutoff type, $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$ (including possibly one of the end points), where types above $\hat{\theta}$ take action $\hat{a} = 1$ and types below $\hat{\theta}$ take action $\hat{a} = 0$.

Next, note that it would never maximize social welfare to set $\hat{\theta} > 0$. In particular, we

could get higher environmental benefits with $\hat{\theta} = 0$ and zero transfers to all agents. Thus any mechanism that sets $\hat{\theta} > 0$ can be improved upon by lowering $\hat{\theta}$ to zero.

Recall that the actions pin down the transfers. Thus now that we have established that there is a cutoff type for the actions, the transfers are pinned down and the mechanism must be a subsidy mechanism as described above. $\square$

2.7.2 Optimal Waiting Periods for the Uniform Distribution

This section expands upon the analysis in Section 2.4.2. In particular, it analyzes the optimal waiting period mechanism when F is the uniform distribution on $[-1, 1]$.

When $F = U[-1, 1]$, the maximization problem can be simplified to the following:³²

$$\max_{\tau} \int_{-\mu}^{\min[\frac{\delta^{\tau}}{1-\delta^{\tau}}\mu, 1]} \rho[\delta^{\tau}(\theta + \mu)] + d^{\tau}(1 - \rho)\omega d\theta + \int_{\min[\frac{\delta^{\tau}}{1-\delta^{\tau}}\mu, 1]}^1 \rho\theta + (1 - \rho)\omega d\theta$$

where μ is determined by the budget constraint and τ . Let $c = \min[\frac{\delta^{\tau}}{1-\delta^{\tau}}\mu, 1]$. The maximization problem simplifies down to:

$$\max_{\tau} \left[\rho\delta^{\tau} \left(\frac{\theta^2}{2} + \mu\theta \right) + d^{\tau}(1 - \rho)\omega\theta \right]_{-\mu}^c + \left[\rho\frac{\theta^2}{2} + (1 - \rho)\omega\theta \right]_c^1$$

Or equivalently:³³

$$\max_{\tau} \rho\delta^{\tau} \left(\frac{c^2}{2} + \mu c + \frac{\mu^2}{2} \right) + d^{\tau}(1 - \rho)\omega(c + \mu) - \rho\frac{c^2}{2} - (1 - \rho)c\omega$$

We can now solve the maximization problem numerically, as it is simply an unconstrained maximization problem in one variable. Depending on the parameter

³²For simplicity, I drop the $\frac{1}{2}$ from the cdf. This simply scales the entirety of social welfare, and therefore has no impact on the optimal τ .

³³At this stage, I drop constants, as they do not impact the optimization problem.

values, the optimal τ can be either strictly positive or zero. Figure 2.3 in Section 2.4.2 illustrates how the optimal waiting period varies as δ varies.

Chapter 3

Estimating Additionality in Energy-Efficiency Subsidy Programs

Abstract: I develop a novel identification technique for estimating the fraction of non-additional participants in subsidy programs. I then apply the technique to a unique data set for a heat pump subsidy program at a medium-sized utility in Oregon. Many utilities in the US offer subsidies for households who engage in energy efficiency upgrades in their homes. The environmental and social benefits of these programs depend on the fraction of participants who are non-additional—that is, participants who would have undergone the energy efficient upgrade in the absence of the subsidy (and are therefore not changing their behavior in response to the subsidy). Using a theoretical model of household behavior, I develop an identification technique that exploits program interruptions to estimate the fraction of participants who are non-additional. Applying this to the heat pump program in Oregon, I estimate that approximately one in three participants (31.2%) are non-additional.

3.1 Introduction

Many electricity utilities in the United States offer incentives to their customers to undergo energy efficiency upgrades in their homes or businesses. However, due to the cost-savings that come with an efficiency upgrade, some households would be willing to undergo the upgrade in the absence of an incentive. These so-called “non-additional” households receive subsidies (using up program budget) while not providing

any additional environmental benefits.¹ Any cost-benefit analysis of energy efficiency subsidy programs therefore must include an estimate of what proportion of participants are non-additional.

Non-additionality, however, is not easily observed. By definition, it is measured relative to an unobserved counterfactual: what the household would have done in a world with no subsidy program. Previous studies of additionality in energy efficiency subsidy programs have relied on two approaches, described below. I develop a third approach.

The first of the approaches for estimating additionality uses engineering and other estimates to attempt to predict which customers would voluntarily undergo the upgrade. For example, Grösche and Vance [2009] use willingness-to-pay estimates for households in Germany for their estimate that 50% of energy efficiency participants would have undergone upgrades in the absence of a subsidy. Similarly, Malm [1996] finds that 89% of participants in an energy efficient heating system subsidy program would have undergone the upgrade without an incentive. There is evidence that some engineering estimates overestimate the cost-savings to the household, meaning this may not be a reliable estimate and may overestimate non-additionality (Allcott and Greenstone [2012]).

Additionally, there is a literature demonstrating that households often fail to undergo upgrades that appear to be in their financial best-interest, which introduces a source of error into this first approach to estimating additionality. There are a variety of possible explanations for why households may fail to undergo upgrades that appear to be in their best interest, ranging from liquidity constraints to informational barriers to inaccurate cost estimates. For example, Fowlie et al. [2015b] find evidence that households have large non-monetary costs to participate in a weatherization program, explaining a low uptake

¹Here, additional refers to the fact that in the absence of a subsidy, the household would still provide the environmental benefits. Thus the subsidy itself is not causing any environmental benefits.

rate. Allcott [2011] collects evidence for households' misperceptions of energy costs. As another example, discount rates estimated from energy efficiency upgrade data tend to be quite high, perhaps reflecting liquidity constraints or other factors impacting households' decisions. For instance, Corum and O'Neal [1982] estimate discount rates between 30% and 41% using data from energy-saving home construction practices. Similarly, Meier and Whittier [1983] estimate discount rates above 60% for 40% of the households in their data on energy efficient refrigerators. Taken together, these factors all suggest that additional measures relying on engineering estimates and other revealed preference approaches may tend to over-estimate the number of non-additional households.

The second main approach for estimating additionality relies on self-reported data, such as surveys. This involves asking the program participants what they would have done in the absence of the subsidy program. These surveys are not incentive-compatible, and therefore customers may not respond truthfully (Carson and Groves [2007]). This will be especially true if non-additional households are concerned about backlash if they reveal their true type. Therefore it is likely that this approach under-estimates the number of non-additional households.

In this paper, I develop a third approach for estimating the fraction of non-additional participants. Using a model of household behavior for program participation, I develop an identification technique that exploits an unanticipated gap in a subsidy program. I then apply this approach to a unique dataset from a heat pump subsidy program at a utility in Oregon. This identification technique avoids the problems inherent in using both engineering estimates and self-reported data. However, the technique requires particular features in the dataset. For the heat pump program in Oregon, I estimate that 31.2% of program participants are non-additional.

There is a broader related literature on energy efficiency programs. This includes papers on the rebound effect (see Davis et al. [2012] and Fowlie et al. [2015a]), which

estimate actual changes in energy use after participation. This complements the additional literature: both additional and rebound effects significantly impact the actual energy savings realized by these programs. Another related literature examines household willingness-to-pay for energy efficiency upgrades (see Banfi et al. [2008]) and the impact of incentives on demand for energy efficient appliances (see Wasi and Carson [2013]).

The remainder of the paper is organized as follows. Section 3.2 presents a theoretical model of household behavior and describes the identification technique. Section 3.3 describes the novel dataset and the key features of the utility's energy efficiency subsidy programs. Section 3.4 describes the empirical model. Section 3.5 presents the regression results as well as the corresponding estimates of additional. Section 3.6 concludes.

3.2 Theoretical Model

I start by modeling the household's decision to participate in the program or not. For a given premise, the household can only participate in the program once.² For simplicity, for now I assume that households do not move. This means that once the household has undergone the upgrade they are no longer eligible to undertake the upgrade in the future and thus do not have a decision to make.

If the household chooses to undergo the upgrade in period t , they incur a one time installation cost, $c_{i,t}$, receive a one-time rebate,³ $r_t \geq 0$, and begin to receive a stream of benefits, $\{b_{i,t}\}$, where $b_{i,t} > 0$ for all i and all t .

The installation cost is subject to a stream of shocks. Household i 's installation cost in period t is a function of both the previous period's cost, $c_{i,t-1}$ and of a shock, $\varepsilon_{i,t}$. Specifically, $c_{it} = c_{i,(t-1)} + \varepsilon_{it}$. I assume that $c_{i,0} > 0$. The shocks are distributed

²For the majority of my analysis, I combine the two heat pump programs, as a customer will not install both a ducted and a ductless heat pump.

³For simplicity, I only consider the case of rebates and ignore the subsidized loans.

identically and independently across both time and individuals, and have mean zero.

Household i who undergoes the upgrade in time period τ will receive the following utility in each time period t :

$$u_i(\tau, t) = \mathbb{I}\{t = \tau\}[r_\tau - c_{i,\tau}] + \mathbb{I}\{t \geq \tau\}b_{i,t}$$

Households have discount factors $\delta_i \in [0, 1]$. For simplicity, let $B_i = \sum_{t=0}^{\infty} \delta^t b_{i,t}$. Let $U_i(t)$ be the net present value of utility for household i in time period t , given that they behave optimally. For simplicity, assume that $r_t = r$ for all t . Thus in time period t , a household who has not yet undergone the upgrade faces the decision described by:

$$U_i(t) = \max\{r - c_{i,t} + B_i, \delta E_\varepsilon[U_i(t+1)]\}$$

U_i is monotone in c . This will imply that there is a threshold value $\bar{c}_i$ for participation. It follows that the household will undergo the upgrade in the first time period for which $c_{i,t} \leq \bar{c}_i$ for some threshold $\bar{c}_i$. At the threshold, the household will be indifferent between both choices, i.e., $r - \bar{c}_i + B_i = \delta U_i(t+1)$. A closed-form characterization would depend on the distribution of ε and would not provide any useful information for the identification technique. Thus for simplicity, I simply note that the cutoff value exists, and that it is less than $\hat{c}$, where $\hat{c} = r + B_i$ (this is due to the fact that $U_i(t) > 0$ for all t).

Let $\bar{c}_i(r)$ be the cutoff value for household i facing rebate r . A household is said to be non-additional in time period t and if they would undergo the upgrade even without a rebate. That is, they are non-additional if $c_{i,t} \leq \bar{c}_i(0)$. They are additional in time period t and at rebate level r if they will only undergo the upgrade with the rebate, that is, $c_{i,t} \in (\bar{c}_i(0), \bar{c}_i(r)]$. If their cost is higher, then they are a non-participant in time period t and at rebate r and will not undergo the upgrade even with the rebate. Note that $\bar{c}_i(r)$ is

increasing in r .

Continue assuming a constant rebate amount, and assume that all households are present in time period zero (that is, there is no entry of new households). The model predicts that there will be a spike in participation in the first time period, followed by a declining trend in participation.⁴ See Figure 3.1 for an illustration. This figure and the following figures all assume 50,000 households and 40 time periods. Their initial costs are uniformly distributed on the interval $[0, 10]$, and their shocks are distributed i.i.d. standard normal. In Figure 3.1, the cutoff cost for participation is 2. To aid in comparisons across figures, they all use the same draws of initial costs and shocks.

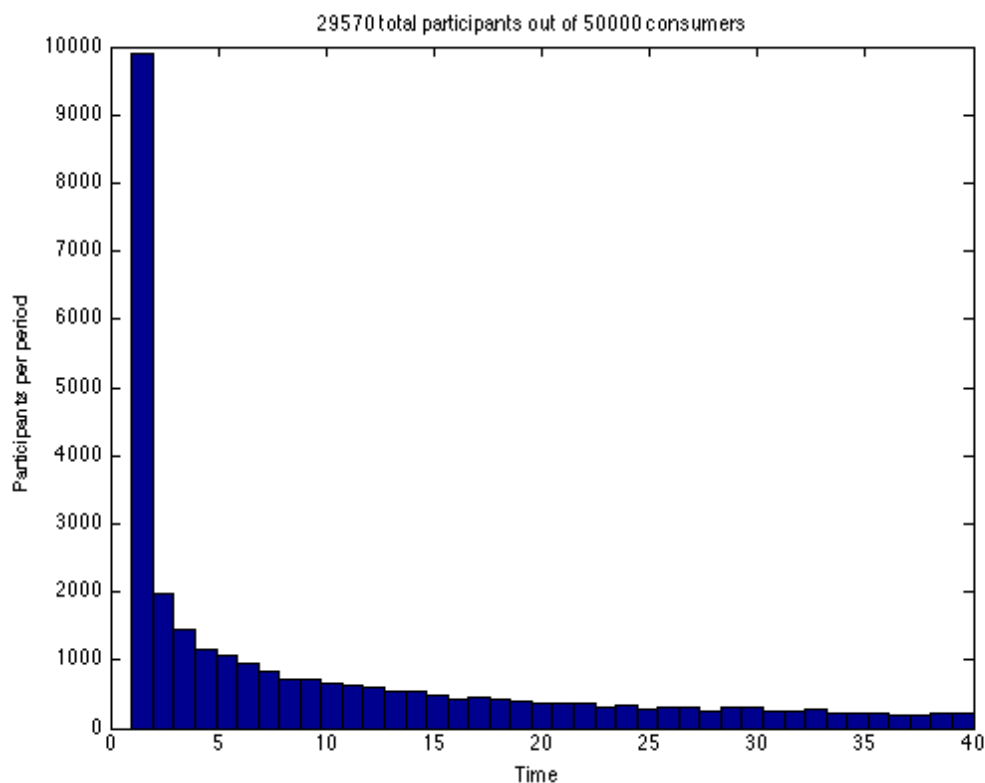


Figure 3.1. Participation over time with a cutoff cost of 2

⁴If new households can enter, then the model still starts with a large spike, however the ongoing trend can be either increasing or decreasing depending on the rate of new arrivals.

It is also useful to consider what happens when there are unanticipated changes in the rebate amount. First, consider an unanticipated increase in the rebate amount. This will cause an immediate spike in participation. Participation will then decline, although initially at both a higher level and more rapid rate of decrease than before. See Figure 3.2 for an illustration. As with Figure 3.1, the cutoff costs starts out at 2. However, following an unanticipated increase in the rebate amount in time period 10, the cutoff cost increases to 5.

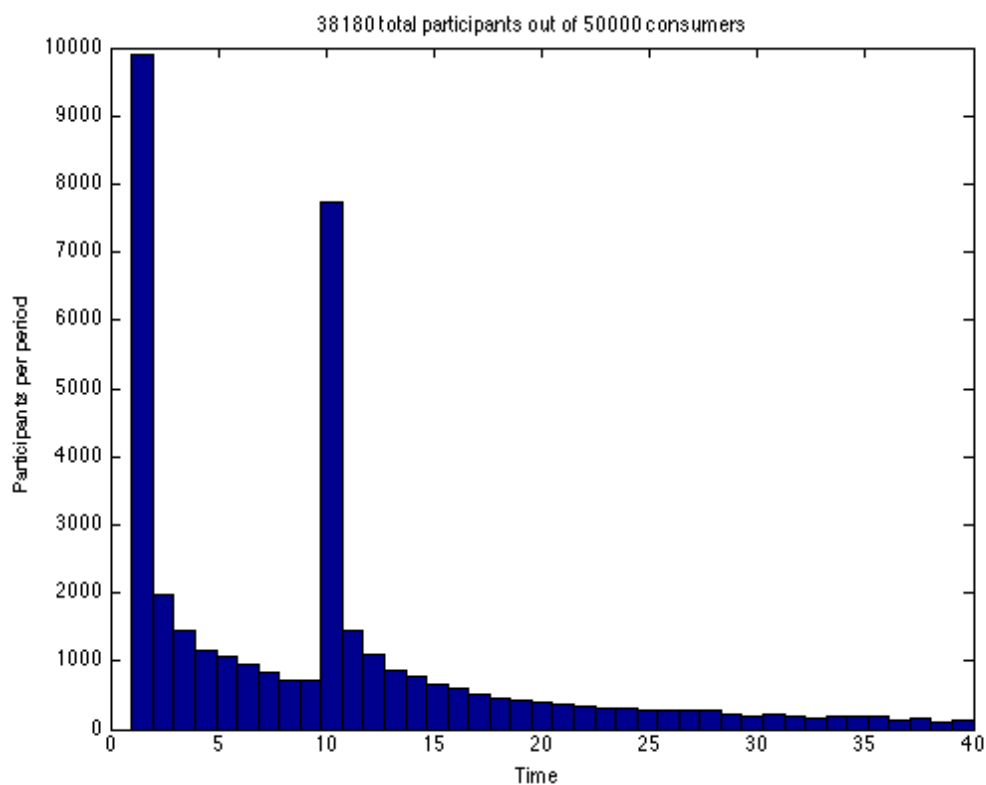


Figure 3.2. Participation over time with an initial cutoff cost of 2, followed by a cutoff cost of 5 after time period 10

Next, consider an unanticipated decrease in the rebate amount. This will cause an immediate drop in participation. This is followed by a brief increasing trend in participation before the participation trend decreases once again. Figure 3.3 illustrates

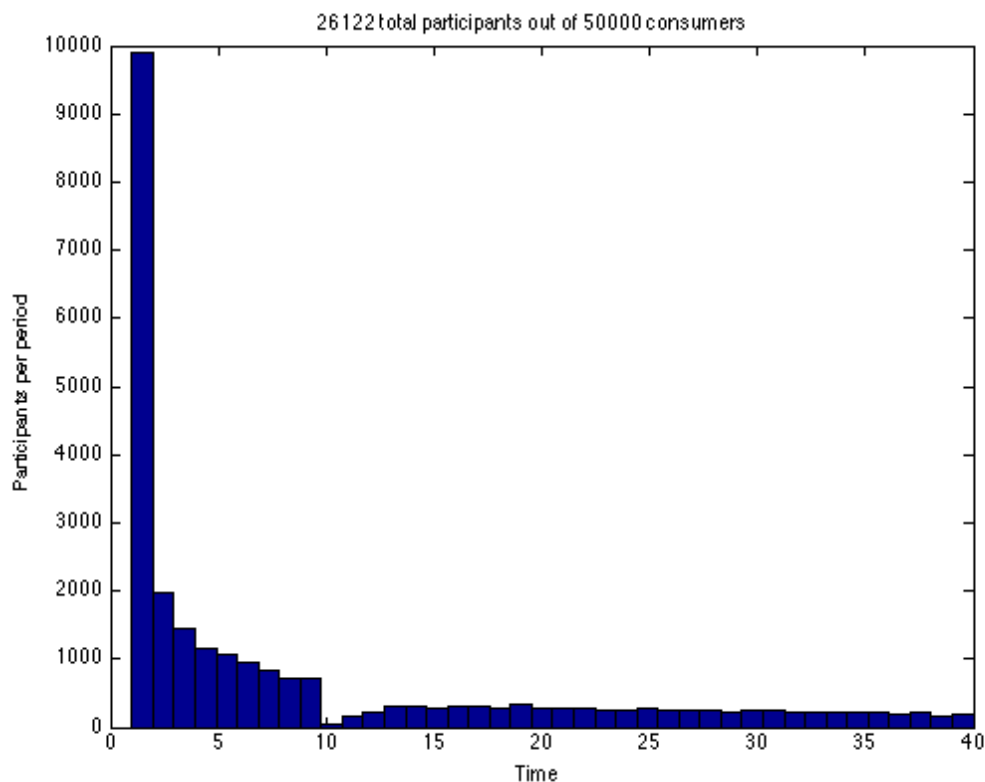


Figure 3.3. Participation over time with an initial cutoff cost of 2, followed by a cutoff cost of 0.5 after time period 10

this case. Once again, the initial cutoff cost is 2, and then it drops to 0.5 in time period 10.

Finally, consider an unanticipated but temporary disappearance of the rebate. To simulate this, I once again start with a cutoff cost of 2. Then in time period 10, the cutoff cost decreases to 0.5 (to simulate an unanticipated disappearance of the rebate).⁵ Then in time period 15, the cutoff value returns to 2, simulating an unanticipated restart of the rebate program. Figure 3.4 illustrates this case.

This last sequence of rebate changes is what underlies the identification technique used in this paper. Figure 3.5 illustrates how an unanticipated gap in a rebate program

⁵Note that the cutoff cost will still be positive because there are private benefits due to savings on energy bills.

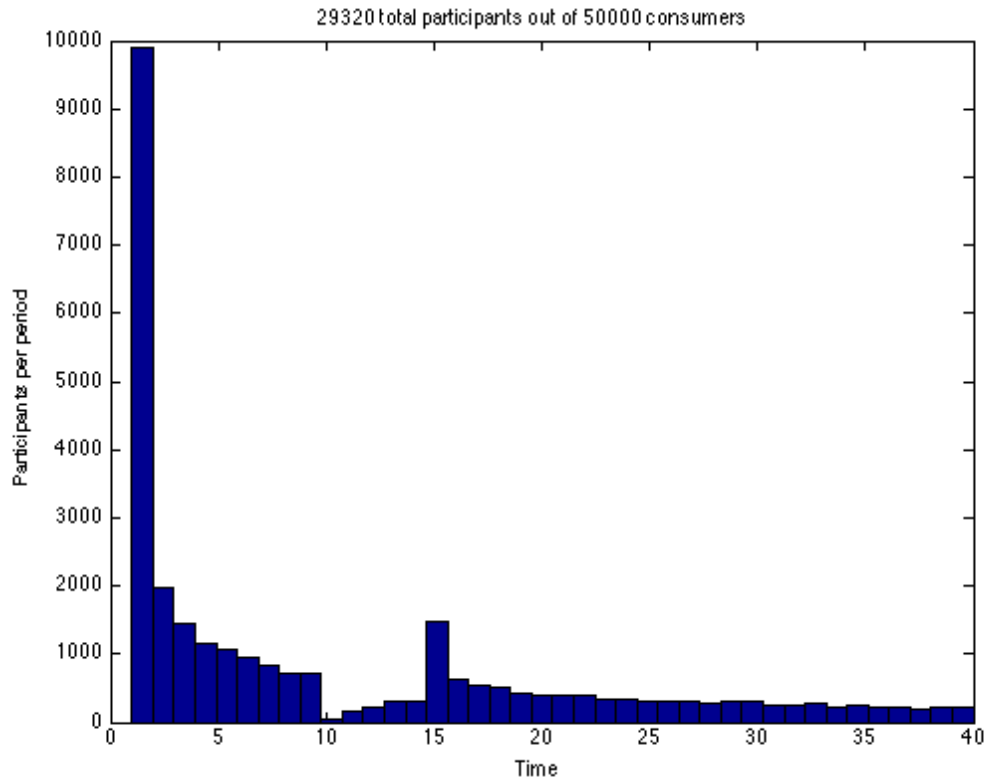


Figure 3.4. Participation over time with an initial cutoff cost of 2, followed by a cutoff cost of 0.5 after time period 10, and a return to a cutoff cost of 2 in time period 15

can be used to identify the fraction of participants who are non-additional. At time T_0 , the program has been up and running for a sufficient number of time periods that we do not see the spike that occurs when the program and time begin. At time T_1 , the rebate program unexpectedly ends. At time T_2 , the program unexpectedly restarts. T_{end} represents the end of the data period. Line x illustrates rebate program participation prior to the gap. Lines y and z represent what participation would have been if the program had not had a gap. Line w represents participation in the rebate program after the gap ends. Note that line w asymptotes towards line z . Line u represents households who undergo the efficiency upgrade during the program gap (and do not receive a rebate), while line v represents households who would have undergone the upgrade even if the rebate program

did not restart.

The fraction of participants who are non-additional can be expressed as $\frac{B}{A+B}$ using the areas labeled in Figure 3.5. However, data are not collected on line u during the program gap, and thus B cannot be directly inferred.⁶ Lines x and w can be observed, and lines y and z can be estimated using extrapolation from line x .⁷ This means that area $A + B$ can be estimated, as can area C . However, the model predicts that for a sufficiently high T_{end} , it will be the case that $C = A$. Thus the program gap can be exploited to get an estimate of $\frac{A+B-C}{A+B}$ which equals the fraction of program participants who are non-additional.

3.2.1 Limitations and Data Requirements for the Identification Technique

There are several threats to identification using this technique. Returning to Figure 3.5, anything that makes the estimate of area C not equal to area A will cause a threat to identification, as will anything that alters areas A and B .

The identification technique assumes that all program changes are unanticipated. If this is not the case, the estimate will be biased. First, if households anticipate the program ending, households may shift from areas A and B to before the program gap. This will lead to an underestimate of area $A+B$. Second, if households anticipate the restarting of the program, that will shift households from area B into area C . This will bias the estimate of non-additional households downwards.

The next concern comes from the choice of T_{end} . If T_{end} is not sufficiently far out, this will result in an underestimate of areas A and C , meaning that the estimate of additional households will be biased downwards. Although the basic theory model

⁶These households do not formally participate in the program and do not receive a rebate, thus there is no record of their efficiency upgrades.

⁷Note that this relies on the fact that the program has been running long enough so that x does not include the spike seen when the program initially starts.

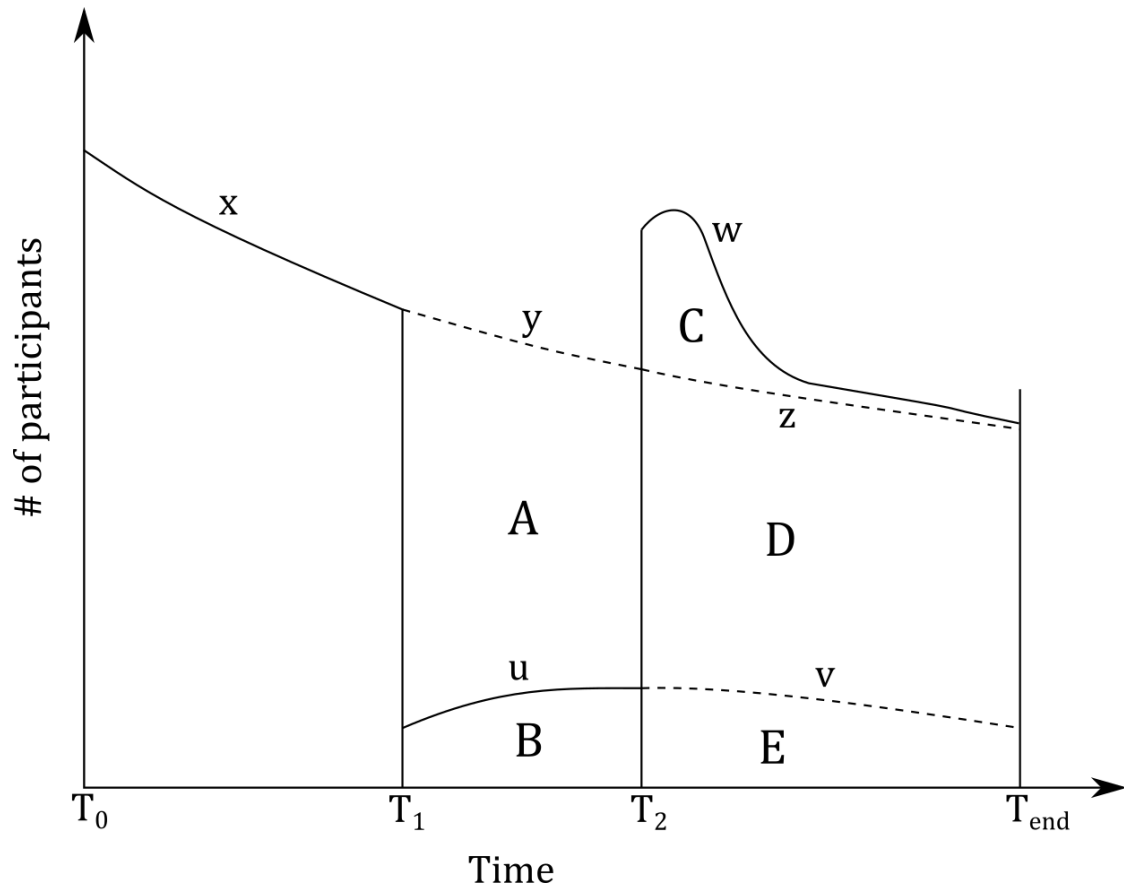


Figure 3.5. Illustration of the identification technique

suggests using a T_{end} as far out into the future as possible, practical matters suggest otherwise. This is because the model assumes no changes in underlying demand for the energy efficiency upgrade. In particular, the farther out T_{end} is, the more likely it is that there is an unobserved shift in underlying demand for the efficiency upgrade. These changes could be due to a variety of factors, such as changes in the price of energy, installation costs, or economic factors. I attempt to control for potential changes in demand in my empirical model. However, it may not be possible to control for all factors impacting demand. This suggests that T_{end} should be chosen to balance these two, competing factors. That is, it should be far enough out that the two trends have effectively converged, but close enough in that it is not picking up extraneous variations in demand.

The last potential concern is changes in the incentive structure of the program, in particular changes co-occurring with the program gap. This would include changes in the loan or rebate amount, changes in the difficulty of the application process, or changes in federal or state tax incentives. Another possible concern is if local energy efficiency installation prices respond to a program gap.⁸ All of these can cause shifts in demand and in program participation rates and invalidate the identification technique.

Although this identification technique requires a very particular scenario, it is nevertheless useful to have another potential tool for estimating additionally. Due to its unobservability, additionality estimates are traditionally difficult to obtain, and this methodology adds another technique to the toolbox for additionality estimates.

3.3 Data

The dataset is from a medium-sized utility in Oregon, with approximately 84,000 residential customers as of 2016. The utility runs several energy efficiency programs that offer rebates and subsidized zero-interest loans to residential customers who engage in specific energy efficiency upgrades in their homes. All residential customers may be eligible for the programs, with some exceptions for homes that already have the relevant upgrades. Both rental units and owner-occupied units are eligible. The loan programs are subject to approval based on utility payment history and a credit report, while rebates are not subject to any restrictions based on credit. Rebate amounts vary depending on the program and total project cost, and loan amounts vary based on the program, project cost, and credit of the applicant. Households may participate in multiple programs.

Table 3.1 summarizes the maximum loan and rebate amounts available for the programs considered in this analysis—the heat pump programs.⁹ A heat pump pumps

⁸For instance, a local installer might drop prices to encourage customers during a subsidy gap.

⁹Amounts are for single-head systems in non-manufactured homes as of April 2017. The utility did not share records of historical incentive amounts, however I do include average incentive amounts over time in

thermal energy between a cool and a warm location (typically between outside a house and inside). They can be used either by themselves, or in conjunction with a traditional heating system. According to the Department of Energy, a heat pump can decrease energy use by up to 50% relative to a traditional heating system. They work best in mild and temperate climates. Exact energy savings depend on a variety of factors, including the type of heating system, type of heat pump, and house characteristics. In addition to the rebate or loan from the utility, some heat pump upgrades are eligible for state and federal tax credits.¹⁰

The utility has historically offered several other programs, including multiple weatherization programs, a water heater program, and an appliance program. These other programs are not a focus of this analysis because their features do not meet the criteria needed for the identification technique (mainly due to inconsistencies in these programs over time).

Table 3.1. Heat Pump Programs

Program	Max rebate amount	Max loan amount
Ducted heat pump	\$1,000	\$12,000
Ductless heat pump	\$650	\$4,000

One potentially confounding factor with this dataset is that the average incentives has varied somewhat over time. Figure 3.6 shows the average rebate and loan amount for the ducted and ductless heat pump programs over time. For the purposes of identification, ideally the incentives would remain perfectly constant over time. This topic is addressed further in Section 3.4.1.

The dataset contains information on energy efficiency program participation and

Figure 3.6.

¹⁰From 2009 to 2013, certain heat pumps were eligible for up to a \$300 federal tax credit. As of 2014 they were no longer eligible. I discuss the implications of this change in Section 3.4.1. The state of Oregon also offers a tax credit for certain heat pumps, however this program remained unchanged during the relevant time period.

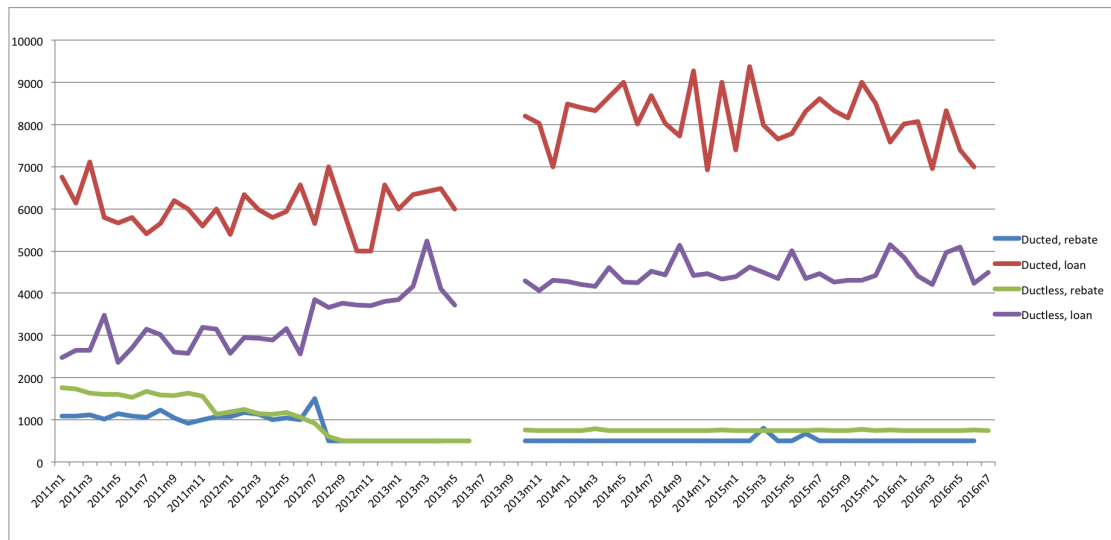


Figure 3.6. Average amount of incentives for heat pump programs over time

a subset of residential customer utility bills from January 2011 through September 2016. The dataset focuses on the utility’s programs for ducted and ductless heat pumps. Table 3.2 lists how many customers participated in each of these programs during the study period.¹¹

Table 3.2. Participants in Heat Pump Programs

Program name	Participants
Ducted heat pump	493
Ductless heat pump	2,630
Total unique participants	3,123

The energy efficiency program dataset includes ID numbers for each customer and premise, the project start date and completion date, information on the premise (including zip code, square footage, type of dwelling, age of house, heat source, and foundation type), estimated kWh savings from the project, project cost, and rebate or loan amount.

¹¹The utility has run a variety of other programs over the years, and although the dataset does not include comprehensive information on these other programs, it does include information on if the customers in the programs listed in Table 3.2 also participated in another program during the study period.

In addition to the energy efficiency program information, the dataset contains utility billings information for all participants in the energy efficiency programs as well as a random sampling of 19,029 non-program participants.¹² The billings data includes ID numbers for each customer and premise, the date of the meter reading (which occurs approximately once per month), the zip code, and the kWh usage for that reading. The dataset also includes limited information on the dwelling's age and the dwelling type,¹³ as well as the date that the customer established their account with the utility.

The Table 3.3 summarizes difference between energy use in customers who never participate in the heat pump program, and customers who do participate (divided pre-and post participation). Prior to participation, program participants use more energy than non-participants.¹⁴ Program participants use an average of 7.11kWh more energy before participation than they do after participation.¹⁵ Even after participation, participants use more energy than non-participants.¹⁶

Table 3.3. Average daily energy usage for participants and non-participants

Customer type	Average daily energy usage	Observations
Non-participants	33.36 kWh	742,261
Participants, pre-participation	41.79 kWh	56,579
Participants, post-participation	34.67 kWh	58,840

In March of 2013, the utility's energy efficiency subsidy programs abruptly and unexpectedly ended. This was due to the utility reaching its energy efficiency targets for the year much earlier than anticipated. The utility's website stated that the programs were temporarily suspended, gave no restart date, and specified that certain programs could change or be eliminated. In fact, of the utility's six primary programs,¹⁷ all but the

¹²Note that not all customers appear for every month in the data. This is due to the fact that some customers moved into or out of the utility's service area.

¹³Dwelling age is truncated at 1980 due to data inaccuracies prior to that point.

¹⁴T-test that the participant mean is greater has P-value=0.0000.

¹⁵T-test that pre-participation energy usage is bigger has a P-value of 0.0000.

¹⁶P-value=0.0000.

¹⁷Prior to 2013, the primary program categories were weatherization, heating systems, appliances, heat

heat pump and weatherization programs were either eliminated or severely scaled back. In October 2013, the utility restarted several of the programs, including the heat pump programs. Although households could submit applications during the program gap, no one was admitted into the programs (and there was a significant risk that the programs would be discontinued all together).

It is useful to consider the patterns in the program participation data and compare them to the predictions from the theoretical model. Figure 3.7 is a histogram of participation in both heat pump programs over time, while Figures 3.8 and 3.9 break down the data by program.¹⁸ Both programs see a jump in participation following the program gap, as predicted by the model. Note that there is no spike at the beginning of the study period because the programs have been running for several years before the beginning of the study period.¹⁹

3.4 Empirical Model

I estimate a binary choice model for households deciding whether or not to participate in the program.²⁰ Let P_{it} be an indicator variable that is one if household i participates in month t and is zero otherwise.²¹ Let G_t be an indicator variable that is 1 if t is after the program gap, and 0 if t is before the program gap. The basic estimation

pumps, thermostats, and energy efficient lighting.

¹⁸“Start Date” refers to the date when the participant begins their upgrade project. The utility does not collect program application dates.

¹⁹The heat pump programs have existed since before 2007.

²⁰For the empirical model, a household is defined by both the customer and premise. Thus a given premise or customer could appear multiple times if people move.

²¹Once households have participated in the program, they are no longer eligible and thus they no longer appear in the regression.

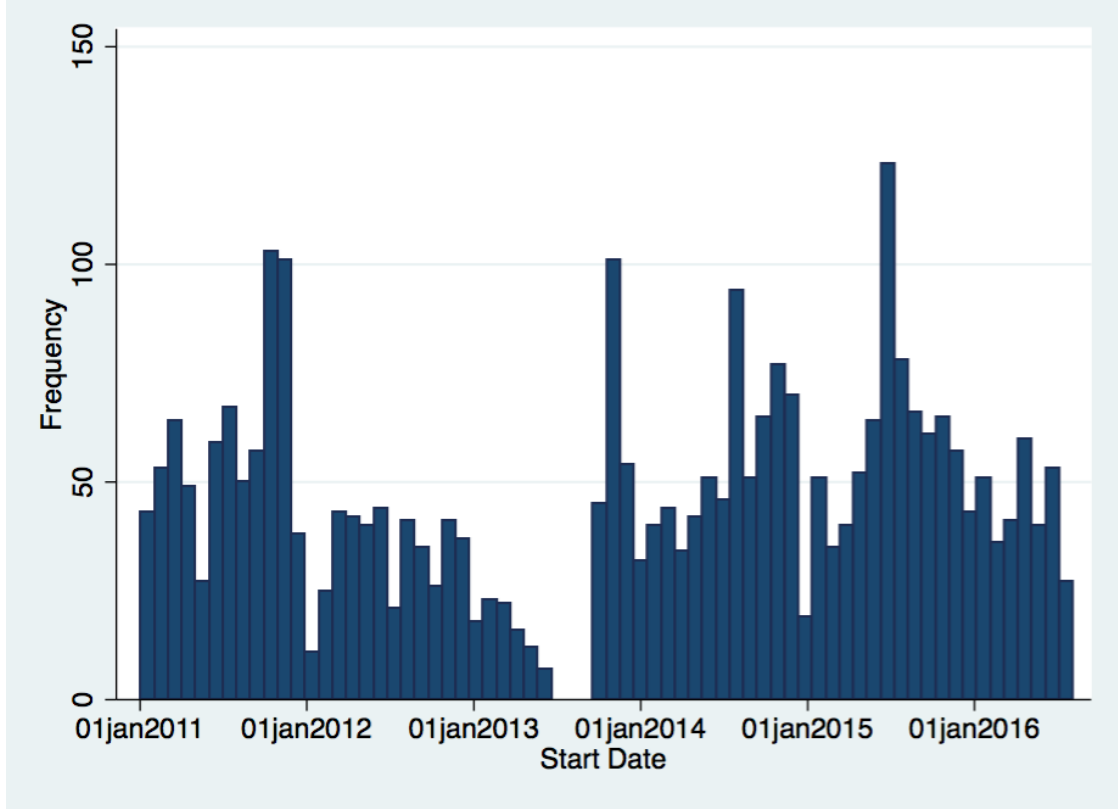


Figure 3.7. Total participation in all heat pump programs

equation is:

$$P_{it} = \alpha + \beta_1 kWh_{it} + \beta_2 kWh_lag_{it} + \beta_3 G_t + \beta_4 G_t \cdot Time_t + \beta_5 G_t \cdot Time_t^2 + \beta_6 Time_t + \beta_7 Time_t^2 + \beta_8 Season_t + \beta_9 Zipcode_i + \varepsilon_{it} \quad (3.1)$$

where α is a constant, and kWh_{it} is the average daily electricity use for household i for month t , kWh_lag_{it} is a moving average of monthly use for the previous year.²² I also include a quadratic time trend to allow for the curvature in participation trends predicted by the theoretical model. Note that the time trend is allowed to be different before and

²²Where an entire year of pre-data are not available, the monthly average simply ignores the missing value. This avoids the problem of losing a large amount of data when customers only appear in the dataset for short amounts of time.

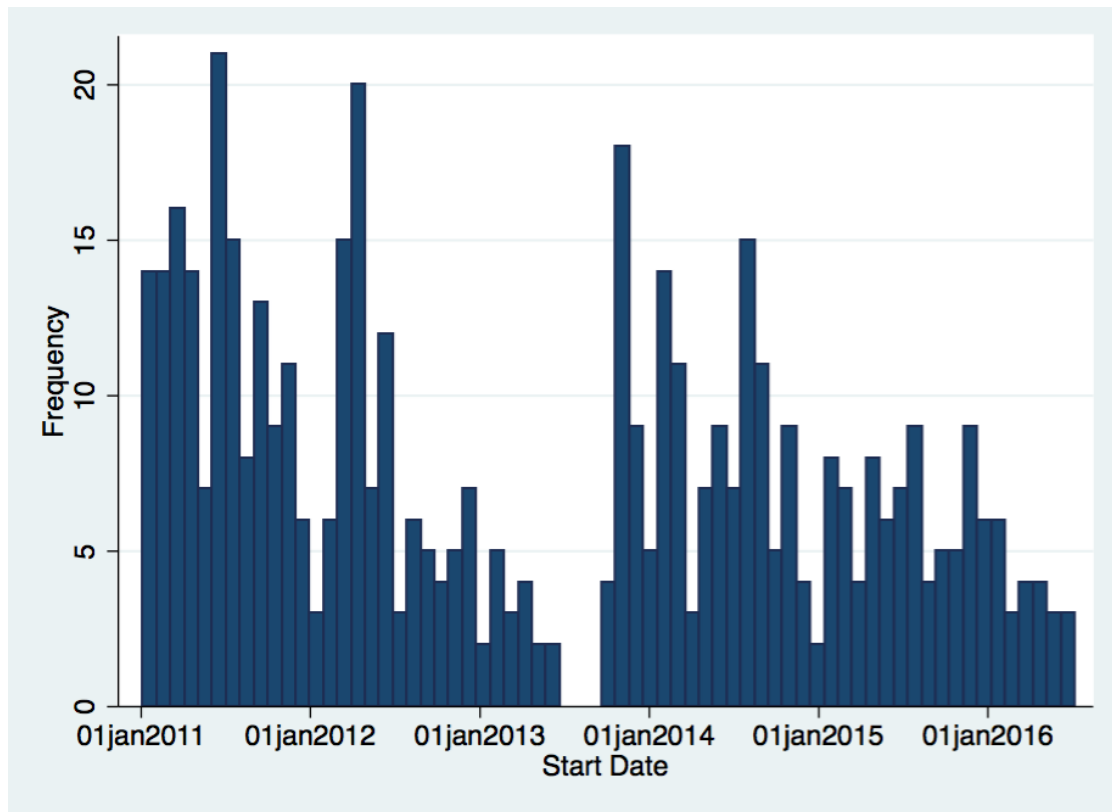


Figure 3.8. Participation in ducted heat pump programs

after the program gap. Lastly, I include sets of dummy variables for both season and zip code. I drop data points during the program gap, as participation was not possible during that time.

I run two additional variations of this model. The first includes additional controls, in particular sets of dummy variables for premise type and the date the house was built. Note that the house building date is truncated at 1980 due to data inaccuracies prior to 1980. Also note that including these additional control reduces the sample size by approximately 30% due to missing data.

The second variation on the model includes household fixed effects. I do not use household fixed effects in the primary model because the goal is to project participation past the program gap. Because not all households appear in every time period, household

rare events.²⁴ In the dataset, out of 830,083 household-month observations, there are only 2,285 instances of participation (about 0.28% of observations). For this reason, I also estimate the following alternate specifications.

For this alternate specification, instead of estimating individual household decisions, I instead estimate the total number of program participants. Let $Participants_t$ be the number of participants in the heat pump programs in month t . The estimation equation is:

$$Participants_t = \alpha + \beta_1 Customers_t + \beta_2 G_t + \beta_3 G_t \cdot Time_t + \beta_4 G_t \cdot Time_t^2 + \beta_5 Time_t + \beta_6 Time_t^2 + \beta_7 Season_t + \epsilon_t \quad (3.2)$$

Where α is a constant, $Customers_t$ is the number of eligible customers in month t ,²⁵ and G_t is an indicator for being after the program gap as before. I also include a quadratic time trend and seasonal dummy variables.

Equation 3.2 assumes that underlying demand for heat pumps stays constant over time. The following estimation equation relaxes this assumption and attempts to control for factors that potentially affect underlying demand:

$$Participants_t = \alpha + \beta_1 Customers_t + \beta_2 Energy_price_t + \beta_3 Heater_price_t + \beta_4 G_t + \beta_5 G_t \cdot Time_t + \beta_6 G_t \cdot Time_t^2 + \beta_7 Time_t + \beta_8 Time_t^2 + \beta_9 Season_t + \epsilon_t \quad (3.3)$$

Where $Energy_price_t$ is a vector of the components of the utilities pricing,²⁶ and $Heater_price_t$ controls for the price of heating upgrades. For $Heater_price_t$, I use a

²⁴See Vilalta and Ma [2002] for a description of some of the difficulties in predicting rare events.

²⁵This is constructed from the total number of residential customers, adjusted by the number of prior participants in the program. Because the utility does not keep a record of total residential customers every month, I extrapolate between the available data points.

²⁶This includes a basic flat charge, a per kWh delivery charge, and three tiers of energy use charges. I adjust all energy prices to account for inflation.

PPI for heating equipment prices.²⁷

I also estimate a variation of Equation 3.3 that controls for the average incentive. This includes five additional variables: average ductless heat pump rebate, average ductless heat pump loan, average ducted heat pump rebate, average ducted heat pump loan, and the federal tax incentive. The state incentive is omitted because it does not vary at all over time.

Estimation equations 3.2 and 3.3 rely on the fact that the demographics of the customer pool remain largely the same over time. If large numbers of customers were participating, there would be a concern that this outflow of customers would alter the demographics of the remaining pool. However, because such a low number of customers participate in each time period (and are therefore removed from the eligible applicant pool), this concern is mitigated.

3.4.1 Potential Threats to Identification

As outlined in Section 3.2.1, the identification technique requires that the dataset have certain features, such as an unanticipated interruption in the program. Failure to demonstrate these features can invalidate or bias the identification technique.

The first concern is anticipation of the program stop or restart. Consider the trend in program participation prior to the program gap. If households anticipated the program ending, we would expect to see an uptick in participation before the gap starts (due to participants shifting from during the gap to before). Looking at the actual trends in participation (see Figure 3.7), this does not appear to be the case as participation is trending downwards prior to the gap. Additionally, the utility has stated that the program stop was not planned, thus there is no reason to believe households would anticipate it.

Next, if households anticipate the restarting of the program, then this will bias

²⁷This is the Bureau of Labor Statistics PPI series for “heating equipment manufacturing,” and is seasonally adjusted.

the estimate of non-additional households downwards. Publicly available materials from the utility (in particular, their website²⁸) from the time period of the program gap do not make any indication of the programs restart date, although they mention that the program suspension may be temporary. In addition, several of the utility's other programs ended up being canceled and never restarted after the gap, indicating that households should not have been certain of a program restart. If some households did wait for the programs to restart, this will bias the estimate of non-additional households downwards.

The next potential issue concerns the specification of T_{end} . The data continue for almost three years after the program gap. This is likely long enough that participation trends have mostly returned to normal. I calculate additionality for a variety of T_{end} values and find that the results are sensitive to the choice of T_{end} . I discuss this in more detail in section 3.5.1.

Other potential threats to identification include other changes co-occurring with the program gap. The first of these is unaccounted for shifts in demand. The program gap occurred after the great recession, and therefore major macroeconomic trends are unlikely to cause a problem. Future extensions of this paper will include discussions with local heat pump salespeople and contractors to ascertain if anything unusual was occurring in those markets during the program gap. I attempt to control for changes in the underlying demand for heat pumps in my empirical specification using energy prices and heater prices.

Other demand shifts could come from changes in the incentives for heat pumps. As illustrated in Figure 3.6, the incentive levels do not remain perfectly flat over time. If this had a major impact on participation, it would confound the results. Changes in federal tax incentives are another potential concern. From 2009 to 2013, certain types of

²⁸ Accessed via the Internet Archive's WayBack Machine.

highest-efficiency electric heat pumps were eligible for up to a \$300 tax credit.²⁹ To test for impacts of incentive changes, I include an empirical specification that controls for incentive levels (including average incentives by month for the utility's programs and the federal incentive level). I find that including these additional variables has minimal impact on the estimated coefficients, and that none of the incentive coefficients are significant at the 10% level.

Lastly, because the utility suspended several other rebate programs concurrent to the gap, any substitution households make between the programs could cause problems. Future extensions of this paper will examine the questions of substitution in more detail, in part by examining surveys from program participants.³⁰

3.5 Results

Table 3.4 has regression results for the first estimation equation,³¹ using a linear probability model and clustering standard errors at the household level. As expected, the R-squared is incredibly low, making the regression useless for projecting participation without the program gap.

Estimates are fairly stable between the estimates from versions 1 and 2 of the first estimation equation, and the fixed effects model varies only somewhat from the other two versions. As expected, the lag of kWh has a positive effect, meaning that the more energy a household has used in the past, the more likely they are to participate in the program. The current month's kWh usage has a negative effect. Although this could be partly due to the way the data are aggregated monthly (i.e., it could be picking up some of the savings from the heat pump), this is mitigated by the fact that the average

²⁹This was referred to as the Residential Energy Efficient Property Credit, and was altered in 2014 to no longer apply to electric heat pumps.

³⁰These surveys were provided by the utility. Although they cannot be linked to individuals in the data and are from a more limited sample, they may provide additional insights.

³¹3.7.1 contains more extensive regression results for all estimation equations.

Table 3.4. Regression results for the first estimation equation

VARIABLES	(1) Model 1 Basic Model	(2) Model 1 Extra Controls	(3) Model 1 Fixed Effects
kWh	-1.80e-05*** (5.42e-06)	-1.88e-05*** (6.44e-06)	-3.23e-05*** (5.57e-06)
kWh lag	4.95e-05*** (6.87e-06)	1.88e-05** (7.57e-06)	5.59e-05*** (9.65e-06)
G	-0.00687*** (0.00260)	-0.00753*** (0.00288)	-0.00351 (0.00266)
G·Time	0.000290** (0.000123)	0.000296** (0.000138)	-0.000197 (0.000125)
G·Time ²	-4.41e-07 (2.40e-06)	-4.62e-08 (2.78e-06)	1.22e-05*** (2.39e-06)
Time	2.71e-05 (6.31e-05)	2.28e-05 (7.31e-05)	0.000685*** (6.25e-05)
Time ²	-2.70e-06 (2.15e-06)	-3.07e-06 (2.51e-06)	-1.57e-05*** (2.14e-06)
Constant	0.00110** (0.000463)	0.00100* (0.000527)	-0.000313 (0.000573)
Seasonal Controls	Yes	Yes	Yes
Zip Code Controls	Yes	Yes	
Extra Controls		Yes	
Fixed Effects			Yes
Observations	681,636	473,763	681,636
R ²	0.001	0.002	

Robust standard errors in parentheses

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

participant takes over two months between when they start their heat pump project and when they complete it.³² One possible explanation for the negative coefficient is that households may be engaging in multiple energy savings projects around the same time.

³²Specifically, the average duration between the project start date and project completion date is 77 days.

Table 3.5. Regression results for the second and third estimation equations

VARIABLES	(1) Model 2	(2) Model 3	(3) Model 3
Customers	0.354*** (0.113)	0.641*** (0.137)	0.684*** (0.156)
Heater price		13.93*** (3.528)	8.818* (4.518)
G	-274.0** (120.1)	-823.8*** (274.4)	-788.3** (371.8)
Time	-13.56*** (4.450)	-41.31*** (10.92)	-37.37*** (11.71)
Time ²	-0.117 (0.0755)	0.0605 (0.155)	0.0365 (0.181)
G·Time	3.626 (5.582)	27.51** (11.21)	23.18 (15.56)
G·Time ²	0.136 (0.107)	-0.153 (0.171)	-0.0910 (0.218)
Constant	-27075*** (8651)	-53250*** (10863)	-55163*** (12212)
Seasonal controls	Yes	Yes	Yes
Energy price controls		Yes	Yes
Incentive controls			Yes
Observations	60	60	55
R ²	0.485	0.708	0.807

Bootstrapped standard errors (1000 reps) in parentheses

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

Table 3.5 presents the regression results for equations 3.2 and 3.3.³³ These models have much higher predictive power than the first estimation equation. None of the energy price coefficients (base charge, delivery charge, and Tiers 1 through 3 charges) are significant at the 5% level,³⁴ however the price index for heater prices is significant. The coefficient for the heater PPI is positive, meaning that as heaters get more expensive

³³3.7.1 contains the full regression results.

³⁴See 3.7.1 for these coefficients.

people are more likely to participate in the heat pump program. This is possibly because as heaters become more expensive, it is relatively more attractive to install a heat pump instead of undergoing an entire heating system upgrade. The various components of the time trend vary somewhat across the specifications—this is likely because the time trend is absorbing some of the variation from changes in demand in Model 3.2. For all specifications, the coefficient on the number of eligible customers is highly significant and positive, as expected.

I use estimation model 3.3 without incentive controls as the preferred specification (this is column 2 in Table 3.5). The additionality calculations rely mainly on the ability of the model to accurately predict participation trends. Specification 3.2 has an R-squared of 0.485, while specification 3.3 increases the R-squared to 0.708 by controlling for underlying trends in demand for heat pumps. Although adding the incentive controls increases the R^2 to 0.807, it is unclear how to project the incentive level variable across the program gap. Thus I mainly use the estimation equation with incentive controls as a robustness check: adding these controls does not drastically impact the other estimated coefficients.

3.5.1 Additionality Results

Figure 3.10 graphs the actual program participation and the projected program participation had there not been a program gap.³⁵ As expected, actual participation spikes at the end of the gap, and the projected participation trend (without a program gap) falls below actual participation levels. Over time, the two trends appear to converge towards each other.

Recall that the identification of non-additional households involves comparing the area between the projected and actual curves, as demonstrated in Figure 3.5. Area

³⁵The additionality results in this section all use specification 3.3 with no incentive controls (column 2 in Table 3.5).

$A+B$ (projected participation during the gap) equals 249, meaning that had there not been a program gap, I estimate that 249 people would have chosen to participate during that time period. Table 3.6 presents the estimated number of participants in Area C (i.e., the area between actual and projected participation, see Figure 3.5) for different levels of T_{end} . The fraction of non-additional households is estimated as $\frac{A+B-C}{A+B}$, and is also presented in Table 3.6 as a function of T_{end} .

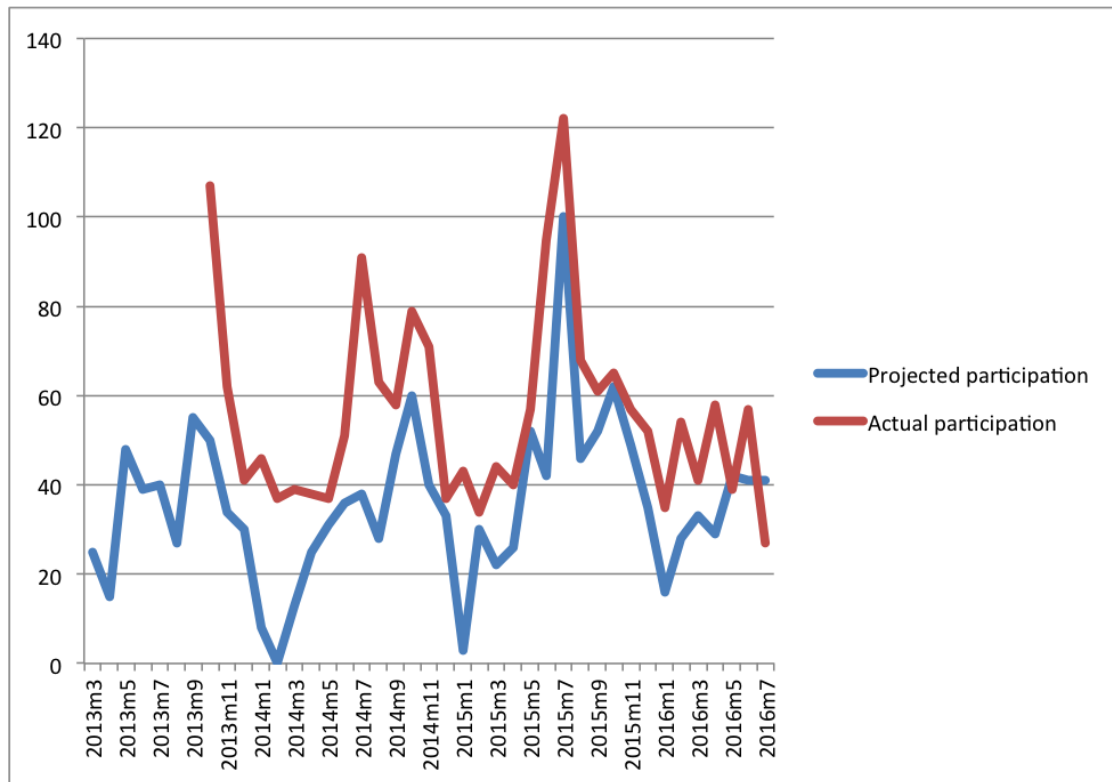


Figure 3.10. Projected participation (in the absence of a program gap) and actual participation

As seen in Table 3.6, the estimated fraction of non-additional households is highly sensitive to the window used (T_{end}), ranging from 0% to 77.3%. Recall that the theoretical model recommends using a T_{end} as far out as possible. That said, the theoretical model also assumes that underlying demand remains constant over time, and if this is not the case, using too long of a window will be inaccurate and will bias the results. A long T_{end}

Table 3.6. Estimated number of participants in Area C and estimated fraction of non-additional participants as T_{end} varies

T_{end}	Participants in Area C	Fraction Non-additional
1 month	57	77.3%
2 months	85	65.8%
3 months	97	61.2%
4 months	135	46.0%
5 months	172	31.2%
6 months	197	21.0%
7 months	210	15.6%
8 months	217	13.0%
9 months	231	7.1%
10 months	285	0.0%
1 year	330	0.0%
End of study period	684	0.0%

can also be a concern because standard errors on the projections become very large over time. Figure 3.11 includes 95% confidence bands around the projection. Note that by the end of the study period, the confidence interval around the participation level in the last month ranges from zero to almost 900 participants.

This leads to the question: what is the appropriate window to use for T_{end} ? An ideal T_{end} would be long enough for the trends to converge, but short enough that it will not pick up noise past the convergence point. Thus picking a T_{end} involves determining where the two trends converge. To do this, I conducted individual tests comparing each projected point to actual participation. That is, I test H_0 : projected and actual participation are different versus H_1 : projected and actual participation are the same. At the 5 percent level, I fail to reject H_0 for months 1, 2, and 4, but reject for all other months. The results of this test can be easily seen by graphing 5 percent confidence bands around the projection, as illustrated in Figure 3.12. This test is not ideal, as it tests points individually rather than jointly. For instance, it is not clear from this test how to interpret months three (reject) and month four (fail to reject). That said, the most

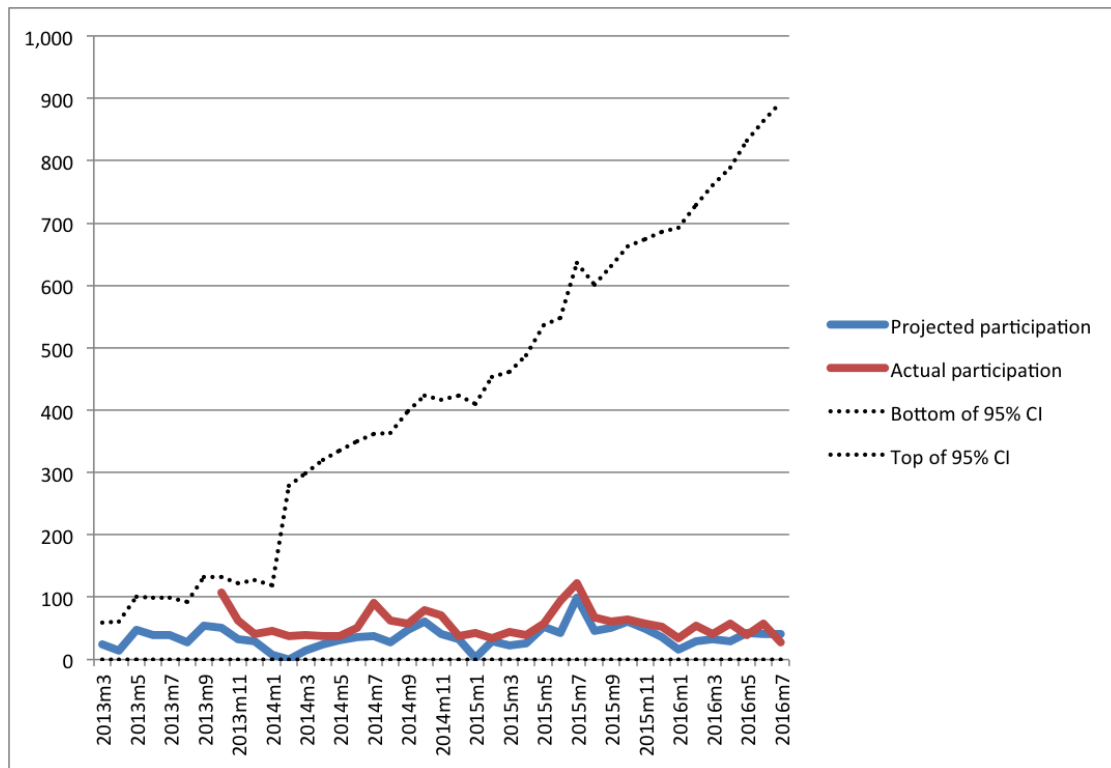


Figure 3.11. Confidence bands on projected participation become very large over time

reasonable conclusion from this test is that the trends converge roughly around month five. This translates into an estimate of 31.2% non-additional participants.

3.6 Conclusion

Using a theoretical model of household behavior, I construct a new identification technique for measuring non-additional households. Unlike previous approaches, this technique does not rely on engineering estimates or on self-reported data, both of which can be inaccurate. However, the identification technique is sensitive to several other factors, including the ability of participants to anticipate program re-starts, changes in underlying demand, and the length of the window used for the additionality calculations.

I apply the identification technique to a dataset of participation in heat pump subsidies from a medium-sized utility in Oregon. The data appear to be consistent with

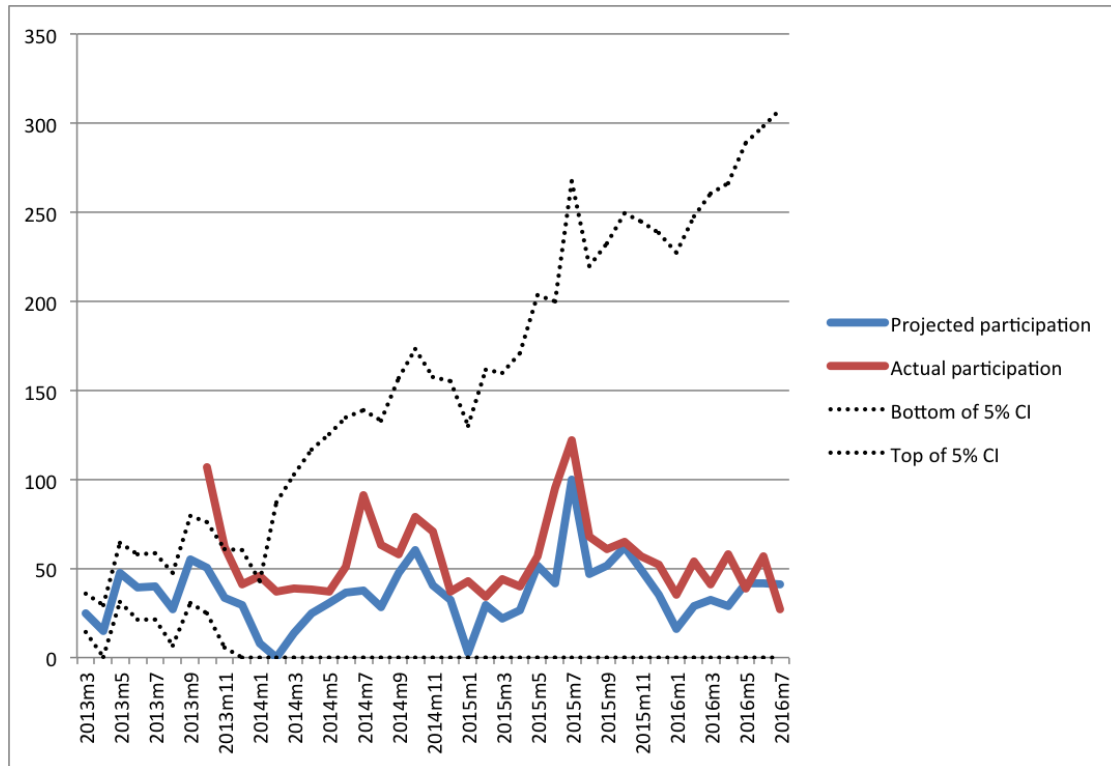


Figure 3.12. Results of the convergence test

the predictions of the theoretical model. I find that the estimated number of non-additional households is sensitive to the window used to measure additionality. Estimates range from 0% to 77% non-additional participants, depending on the window used. Convergence tests suggest that a five month window is appropriate for T_{end} . I conclude that 31.2% of households are non-additional.

3.7 Appendix

3.7.1 Full regression Results

Table 3.7. Full regression results for the first estimation equation

VARIABLES	(1) Model 1 Basic Model	(2) Model 1 Extra Controls	(3) Model 1 FE
kWh	-1.80e-05*** (5.42e-06)	-1.88e-05*** (6.44e-06)	-3.23e-05*** (5.57e-06)
kWh lag	4.95e-05*** (6.87e-06)	1.88e-05** (7.57e-06)	5.59e-05*** (9.65e-06)
G	-0.00687*** (0.00260)	-0.00753*** (0.00288)	-0.00351 (0.00266)
G·Time	0.000290** (0.000123)	0.000296** (0.000138)	-0.000197 (0.000125)
G·Time ²	-4.41e-07 (2.40e-06)	-4.62e-08 (2.78e-06)	1.22e-05*** (2.39e-06)
Time	2.71e-05 (6.31e-05)	2.28e-05 (7.31e-05)	0.000685*** (6.25e-05)
Time ²	-2.70e-06 (2.15e-06)	-3.07e-06 (2.51e-06)	-1.57e-05*** (2.14e-06)
Quarter 1	-0.00135*** (0.000176)	-0.00104*** (0.000200)	-0.00119*** (0.000164)
Quarter 2	-0.000858*** (0.000207)	-0.000575** (0.000236)	-0.000962*** (0.000198)
Quarter 3	-0.000132 (0.000217)	-1.32e-05 (0.000246)	-0.000547*** (0.000207)
Constant	0.00110** (0.000463)	0.00100* (0.000527)	-0.000313 (0.000573)
Zip Code Controls	Yes	Yes	
Extra Controls		Yes	
Fixed Effects			Yes
Number of household FE			30,150
Observations	681,636	473,763	681,636
R ²	0.001	0.002	

Robust standard errors in parentheses

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

Additional controls are sets of dummy variables for premise type and premise age.

Table 3.8. Regression results for the second estimation equation.

VARIABLES	(1) Model 2	(2) Model 3	(3) Model 3
Customers	0.354*** (0.113)	0.641*** (0.137)	0.684*** (0.156)
Base charge		-4.398 (10.68)	-10.66 (28.85)
Delivery charge		9530 (33404)	-10110 (70092)
Tier 1 price		-5131 (17069)	-4092 (96828)
Tier 2 price		15887 (17710)	13591 (74790)
Tier 3 price		-5937 (20553)	1580 (37762)
Heater price		13.93*** (3.528)	8.818* (4.518)
G	-274.0** (120.1)	-823.8*** (274.4)	-788.3** (371.8)
Time	-13.56*** (4.450)	-41.31*** (10.92)	-37.37*** (11.71)
Time ²	-0.117 (0.0755)	0.0605 (0.155)	0.0365 (0.181)
G·Time	3.626 (5.582)	27.51** (11.21)	23.18 (15.56)
G·Time ²	0.136 (0.107)	-0.153 (0.171)	-0.0910 (0.218)
Quarter 1	-27.51*** (6.214)	-25.02*** (7.384)	-27.86*** (8.721)
Quarter 2	-31.05*** (9.313)	-28.15*** (9.855)	-31.22*** (10.50)
Quarter 3	-15.48* (9.035)	-21.48*** (7.872)	-13.20 (8.894)
Federal incentive			-0.00375 (2.004)
Ducted rebate			-0.0372 (0.0300)
Ducted loan			0.00374 (0.00422)
Ductless rebate			0.0622 (0.0536)
Ductless loan			-0.00314 (0.00868)
Constant	-27075*** (8651)	-53250*** (10863)	-55163*** (12212)
Observations	60	60	55
R ²	0.485	0.708	0.807

Bootstrapped standard errors (1000 reps) in parentheses

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

Bibliography

Alberta Agriculture and Forestry. Agricultural carbon offsets: Information for Alberta's offset market: Tillage offset credit update. 2011.

Hunt Allcott. Consumers' perceptions and misperceptions of energy costs. *The American Economic Review: Papers & Proceedings*, 101(3):98–104, 2011.

Hunt Allcott and Michael Greenstone. Is there an energy efficiency gap? *The Journal of Economic Perspectives*, 26(1):3–28, 2012.

Hunt Allcott, Christopher Knittel, and Dmitry Taubinsky. Tagging and targeting of energy efficiency subsidies. *The American Economic Review: Papers & Proceedings*, 105(5): 187–191, 2015.

Silvia Banfi, Mehdi Farsi, Massimo Filippini, and Martin Jakob. Willingness to pay for energy-saving measures in residential buildings. *Energy Economics*, 30(2):503–516, 2008.

Antonio M Bento, Ravi Kanbur, and Benjamin Leard. Designing efficient markets for carbon offsets with distributional constraints. *Journal of Environmental Economics and Management*, 70:51–71, 2014.

Tilman Börgers, Daniel Krähmer, and Roland Strausz. *An introduction to the theory of mechanism design*. Oxford University Press, USA, 2015.

Phillip H Brown and Christopher L Lant. The effect of wetland mitigation banking on the achievement of no-net-loss. *Environmental Management*, 23(3):333–345, 1999.

California Department of Fish and Game. Staff report on burrowing owl mitigation. 1995.

Richard T Carson and Theodore Groves. Incentive and informational properties of preference questions. *Environmental and Resource Economics*, 37(1):181–210, 2007.

- Richard T. Carson and Brigitte Roth Tran. Discounting behavior and environmental decisions. *Journal of Neuroscience, Psychology, and Economics*, 2(2):112–130, 2009.
- SE Chidiac, EJC Catania, E Morofsky, and S Foo. A screening methodology for implementing cost effective energy retrofit measures in Canadian office buildings. *Energy and Buildings*, 43(2):614–620, 2011.
- Rae Kwon Chung. A CER discounting scheme could save climate change regime after 2012. *Climate Policy*, 7(2):171–176, 2007.
- Connecticut Department of Environmental Protection. Connecticut’s nitrogen credit exchange: An incentive-based water quality trading program. 2010.
- Kenton R Corum and Dennis L O’Neal. Investment in energy-efficient houses: An estimate of discount rates implicit in new home construction practices. *Energy*, 7(4): 389–400, 1982.
- T.E. Dahl and S.M. Stedman. Status and trends of wetlands in the coastal watersheds of the conterminous United States: 2004 to 2009. U.S. Department of the Interior, Fish and Wildlife Service and National Oceanic and Atmospheric Administration, National Marine Fisheries Service, 2013.
- Lucas W Davis, Alan Fuchs, and Paul J Gertler. Cash for coolers. *National Bureau of Economic Research Working Paper Series*, 2012.
- JR DeShazo, Tamara L Sheldon, and Richard T Carson. Designing policy incentives for cleaner technologies: Lessons from California’s plug-in electric vehicle rebate program. Technical report, Working paper, 2014.
- Paul V. Ellefson, Antony S. Cheng, and Robert J. Moulton. Regulation of private forestry practices by state governments. *State Bulletin Minnesota Agricultural Experiment Station*, 605-1995, 1995.
- Environmental Law Institute. Banks and fees: The status of off-site wetland mitigation in the United States. 2002.
- European Commission. Emissions trading: Commission welcomes vote to ban certain industrial gas credits. *European Commission Press Release Database*, 2011.
- Nicolás Figueroa and Vasiliki Skreta. The role of outside options in auction design. 2007.
- Finn R Førsund and Eric Nævdal. Efficiency gains under exchange-rate emission trading.

- Environmental and Resource Economics*, 12(4):403–423, 1998.
- Meredith Fowlie, Michael Greenstone, and Catherine Wolfram. Do energy efficiency investments deliver? Evidence from the weatherization assistance program. *National Bureau of Economic Research Working Paper Series*, 2015a.
- Meredith Fowlie, Michael Greenstone, and Catherine Wolfram. Are the non-monetary costs of energy efficiency investments large? Understanding low take-up of a free energy efficiency program. *The American Economic Review: Papers & Proceedings*, 105(5):201–204, 2015b.
- Heidi Gjertsen, Dale Squires, Peter H Dutton, and Tomoharu Eguchi. Cost-effectiveness of alternative conservation strategies with application to the Pacific leatherback turtle. *Conservation Biology*, 28(1):140–149, 2014.
- Peter Grösche and Colin Vance. Willingness to pay for energy conservation and free-ridership on subsidization: Evidence from Germany. *The Energy Journal*, pages 135–153, 2009.
- Jerry A Hausman. Individual discount rates and the purchase and utilization of energy-using durables. *The Bell Journal of Economics*, pages 33–54, 1979.
- Pete Hodgson. Sea lion kills force closure of squid fishery. New Zealand Ministry of Fisheries, 2002.
- Rebecca L Kihlslinger. Success of wetland mitigation projects. *National Wetlands Newsletter*, 30(2):14–16, 2008.
- D King and C Bohlen. Estimating the costs of restoration. *National Wetlands Newsletter*, 16(3):3–5, 1994a.
- Dennis M King and Curtis C Bohlen. Making sense of wetland restoration costs. University of Maryland Center for Environmental and Estuarine Studies, 1994b.
- Dennis M King and Elizabeth W Price. Developing defensible wetland mitigation ratios: A companion to “The five-step wetland mitigation ratio calculator”. Technical report, prepared for NOAA Office of Habitat Conservation, Habitat Protection Division, 2004.
- George J Mailath. Incentive compatibility in signaling games with a continuum of types. *Econometrica*, pages 1349–1365, 1987.
- Eric Malm. An actions-based estimate of the free rider fraction in electric utility DSM

- programs. *The Energy Journal*, pages 41–48, 1996.
- Alejandro M Manelli and Daniel R Vincent. Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *Journal of Economic theory*, 137(1):153–185, 2007.
- Alan K Meier and Jack Whittier. Consumer discount rates implied by purchases of energy-efficient refrigerators. *Energy*, 8(12):957–962, 1983.
- Axel Michaelowa and Pallav Purohit. Additionality determination of Indian CDM projects: Can Indian CDM project developers outwit the CDM Executive Board? *Climate Strategies*, Discussion Paper CDM-1, 2007.
- Juan-Pablo Montero. Optimal design of a phase-in emissions trading program. *Journal of Public Economics*, 75(2):273–291, 2000.
- Andrew Moxey, Ben White, and Adam Ozanne. Efficient contract design for agri-environment policy. *Journal of agricultural economics*, 50(2):187–202, 1999.
- Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.
- National Research Council. Compensating for wetland losses under the Clean Water Act. 2001.
- Efe A Ok. *Real analysis with economic applications*. Princeton University Press, 2007.
- Imran Rasul and Silvia Sonderegger. The role of the agent’s outside options in principal–agent relationships. *Games and Economic Behavior*, 68(2):781–788, 2010.
- James T Robb. Assessing wetland compensatory mitigation sites to aid in establishing mitigation ratios. *Wetlands*, 22(2):435–440, 2002.
- Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill New York, 3rd edition, 1976.
- Richard Schmalensee and Robert Stavins. Lessons learned from three decades of experience with cap-and-trade. *National Bureau of Economic Research*, NBER Working Paper No. 21742, 2015.
- Lambert Schneider. Assessing the additionality of CDM projects: Practical experiences and lessons learned. *Climate Policy*, 9(3):242–254, 2009.

- Lambert Richard Schneider. Perverse incentives under the CDM: An evaluation of HFC-23 destruction projects. *Climate Policy*, 11(2):851–864, 2011.
- Igor Shishlov and Valentin Bellassen. Review of the experience with monitoring uncertainty requirements in the Clean Development Mechanism. *Climate Policy*, 16(6): 703–731, 2016.
- R Eugene Turner, Ann M Redmond, and Joy B Zedler. Count it by acre or function: Mitigation adds up to net loss of wetlands. *National Wetlands Newsletter*, 23(6):5–6, 2001.
- US EPA. Water quality trading toolkit for permit writers. 2009.
- Arthur Van Benthem and Suzi Kerr. Scale and transfers in international emissions offset programs. *Journal of Public Economics*, 107:31–46, 2013.
- William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961.
- Ricardo Vilalta and Sheng Ma. Predicting rare events in temporal domains. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 474–481. IEEE, 2002.
- Virginia Department of Environmental Quality. Trading nutrient reductions from non-point source best management practices in the Chesapeake Bay watershed: Guidance for agricultural landowners and your potential trading partners. 2008.
- Michael W Wara and David G Victor. A realistic policy on international carbon offsets. *Program on Energy and Sustainable Development Working Paper*, #74, 2008.
- Washington Department of Ecology. Wetlands in Washington state: Volume 2: Guidance for protecting and managing wetlands. 2005.
- Washington Department of Ecology. Wetland mitigation in Washington state: Part 1: Agency policies and guidance. 2006.
- Nada Wasi and Richard T Carson. The influence of rebate programs on the demand for water heaters: The case of New South Wales. *Energy Economics*, 40:645–656, 2013.
- Chris Wilcox and C Josh Donlan. Compensatory mitigation as a solution to fisheries bycatch: Biodiversity conservation conflicts. *Frontiers in Ecology and the Environment*, 5(6):325–331, 2007.

RX Zerbe, Xi Han, David Layton, and Tom Leshine. A history of discount rates and their use by government agencies. *Working Paper*, 2002.

Junjie Zhang and Can Wang. Co-benefits and additionality of the clean development mechanism: An empirical analysis. *Journal of Environmental Economics and Management*, 62(2):140–154, 2011.