Economics of Air Pollution: Policy, Mortality Concentration-Response, and Increasing Marginal Benefits of Abatement

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Andrew Lloyd Goodkind

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Jay S. Coggins

December, 2014

Acknowledgements

I wish offer my greatest appreciation and thanks to those that have shaped my thinking and supported my education. My adviser, Jay Coggins, for inspiring me to study environmental economics and for providing me unwavering encouragement. To the members of my dissertation committee, Julian Marshall, Stephen Polasky and Frances Homans, my education has been greatly enhanced by learning from each of you in classes, as an assistant, and as an adviser to me. My deepest thanks to Chris Tessum, for patiently explaining the details of air dispersion models and selflessly sharing his models and data that were integral to the completion of my dissertation. I want to credit the members of Jay Coggins' environmental economics special project group, for the fruitful discussions of increasing marginal benefits that helped develop my research topics.

I am forever grateful to my parents, and my wife Amanda for supporting my decisions, providing needed encouragement, and making everything I do possible.

Dedication

To Amanda, my loving wife and best friend.

Abstract

This dissertation examines the economics of air pollution in three essays. The first two essays consider the implications of the possibility of increasing marginal benefits to pollution abatement. The third essay integrates a new model of air dispersion with an economic model to estimate the marginal damage caused by criteria pollutants in the United States.

In the first essay, the optimal abatement policy is derived for a scenario with increasing marginal benefits of abatement and uncertainty in the marginal cost of abatement. Pollution taxes are preferred over quantity restrictions when marginal benefits are increasing in abatement.

The second essay uses simulations of fine particulate matter ($PM_{2.5}$) dispersion and compares optimal source-specific pollution control policies with pollution concentration standards and uniform pollution taxes. Optimal policies for $PM_{2.5}$ regulation yield substantial advantages over uniform policies that do not discriminate based on the location of emissions. The simulations also consider the shape of the concentration-response (C-R) relationship between $PM_{2.5}$ pollution and mortality. With a log-log C-R, where marginal benefits of $PM_{2.5}$ abatement are increasing, society should prefer fewer emissions and lower $PM_{2.5}$ concentrations than if the C-R is log-linear, where marginal benefits of abatement are decreasing.

The third essay estimates the marginal damages of criteria pollutant emissions for hundreds of the most heavily polluting sources in the U.S. Marginal damages vary substantially depending on the location of the emission source. The calculation of marginal damages is highly dependent on the choice of air dispersion modeling, the C-R relationship, and the value assigned to mortality caused by environmental risks.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation discusses issues of air pollution and environmental risk. Air pollution that impacts human health is a topic critical to policy makers, businesses and economists. The environmental risks from air pollution, especially fine particulates (particles with diameter less than 2.5 microns, $PM_{2.5}$), are large and affect almost all people.

Epidemiologists have identified fine particulate air pollution as a key environmental risk (Pope *et al.* 1995; Dockery *et al.* 1993; Pope, Ezzati and Dockery 2009b). Exposure to all outdoor air pollutants, globally and in the U.S., has been linked with several detrimental health impacts leading to morbidity and mortality including: respiratory infections; lung, trachea and bronchus cancers; ischaemic heart disease; cerebrovascular disease; and chronic obstructive pulmonary disease (Lim *et al.* 2012). Globally, 3.4 million deaths are attributed to outdoor air pollution annually, *95% of which are from exposure to fine particulates.* In the U.S., risks from fine particulate air pollution is the 8th leading cause of mortality, just ahead of alcohol use, responsible for over 100,000 deaths annually, leading to 1.6 million years of life lost (IHME 2014).

With many lives at stake, controlling air pollution is a pressing issue for government regulators and lawmakers. Making comprehensive decisions to limit the external costs of air pollution is difficult for a number of reasons: there are many sources of air pollution, and the connection between emissions and the people impacted is highly complex. With businesses bearing only a fraction of the environmental costs associated with their emissions, air pollution represents an externality that affects an enormous number of people given dispersion of air pollutants and the difficultly avoiding pollution in the air we breathe. Economists play a prominent role in advising policy makers how to appropriately handle the externalities from emissions of pollutants that originate from many sources. Businesses that emit air pollution represent important and central industries in the economy. Given the magnitude of the externalties associated with $PM_{2.5}$ air pollution, it may be optimal to require businesses to incur large costs to limit their contributions to pollution, even if it results in dramatic changes to the economy.

U.S. government agencies assign large values to health risk reductions of existing and proposed regulations. The Office of Management and Budget estimates benefits of \$19 to \$167 billion per year from Environmental Protection Agency (EPA) regulations of fine particulates, against costs of \$7 billion (OMB 2013). In EPA regulatory impact analyses of ozone (EPA 2008), mercury and air toxics (EPA 2011), and carbon (EPA 2014b), the health benefits of reduced mortality risks attributable to fine particulates are included to justify the regulations.

Although these regulations save many lives, an efficient policy of fine particulate air pollution could reap benefits to society between \$50 and \$220 billion annually, from additional pollution reductions and cost-effective abatement strategies (Muller and Mendelsohn 2009). Across all outdoor air pollutants with identified monetary damages from increased morbidity and mortality, decreased visibility, and impacts to timber production, agriculture and recreational activies, the lion's share of damages are attributable to mortality from chronic exposure to fine particulates. This results from the large mortality risks associated with fine particulate air pollution and the large value placed on avoiding risks to human life.

The impacts of fine particulate air pollution have been identified as an issue of utmost importance by epidemiologists, regulators, and economists, but additional analysis is necessary. This dissertation examines some of the important issues that affect our understanding of the damages from fine particulate air pollution and how they should be regulated.

There exists a disconnect between economic theory of efficient regulation of externalities and most regulations of air pollution. Economics advises pricing emissions of air pollutants at the value of the marginal external damages caused by the emissions. Yet regulators often implement command and control policies requiring emission reductions to meet pollution concentration limits at locations of concern, generally without providing financial incentives for polluters to make cost effective pollution abatement decisions. In the U.S., the National Ambient Air Quality Standards (NAAQS) set pollution concentration limits that cannot be exceeded in any location.

The reliance on concentration limits for criteria pollutants, in particular $PM_{2.5}$, suggests that regulators assume that there are no damages below a particular concentration level. In other words, there is a threshhold concentration level, below which human health risks are not impacted by further concentration reductions. For exposure to $PM_{2.5}$, results from the two most prominent studies into the link with human mortality, the American Cancer Society (ACS) study (Pope *et al.* 2005; Pope *et al.* 2002; Krewski *et al.* 2009) and the Harvard Six City (H6C) study (Dockery *et al.* 1993;

Laden *et al.* 2006; Lepeule *et al.* 2012), indicate that no safe threshhold exists, or that the threshhold is below the concentration levels experienced by most people. If no safe threshhold exists, then there are opportunities to improve public health by focusing attention on reducing risks to people exposed across the entire range of $PM_{2.5}$ concentrations.

The issue of no safe threshhold of $PM_{2.5}$ exposure is related to the concentration-response (C-R) relationship between $PM_{2.5}$ concentrations and the risk of mortality. In the most recent analysis of the ACS study, Krewski *et al.* (2009), estimate two possible functional forms of the C-R: log-linear, a commonly assumed relationship with a constant relative-risk of mortality between any fixed difference in concentration; and log-log, an alternative relationship where the risk of mortality decreases at a faster rate from reductions at low compared to high concentrations. With log-log, the marginal benefits of abating fine particulate pollution are increasing in concentration reductions, suggesting that not only is there no safe threshhold, but the first unit of exposure is the most damaging. The implications of the difference between the log-linear and log-log C-R relationships, for policy and damages from pollution, are explored in the following chapters.

With log-log, a central tenet of environmental economics, that the marginal benefits of abatement are decreasing, is violated. Chapter 2 evaluates the theory of optimal pollution control given increasing marginal benefits of abatement. Here price and quantity instruments are compared under marginal abatement cost uncertainty, a framework first developed by Weitzman (1974). The optimal price and quantity policies are derived depending on the relative slopes of the marginal benefits of abatement and the marginal abatement costs. The key finding is that, with increasing marginal benefits, a quantity policy is never preferred. Unlike Weitzman (1974), who assumed downward sloping marginal benefits of abatement, the optimal price and quantity policies may yield solutions either at the corners (zero abatement or complete abatement) or in the interior. Another surprising finding is that the optimal price policy is not necessarily at the intersection of the marginal benefit and the expected marginal abatement cost functions. In addition, the optimal price policy is a function of the level of uncertainty in marginal costs. As the level of uncertainty increases, the advantage of the price policy over the quantity policy increases.

A simulation model is developed in Chapter 3 that analyzes pollution control policies for $PM_{2.5}$. Dispersion of air pollution from hundreds of sources of $PM_{2.5}$ emissions are simulated using a Gaussian plume model, and the resulting pollution concentrations are calculated in receptors for a region of the U.S. Midwest. The model calculates mortality risks at each receptor with the log-linear and the log-log C-R from Krewski *et al.* (2009). Three air pollution policies are examined: efficient emission abatement, where the marginal costs of abatement are equated with the marginal benefits of abatement at each source; a uniform pollution limit that cannot be exceeding in any receptor; and a uniform emissions tax.

Important differences are found in the outcomes across the three policies and the two C-R relationships. With a log-log C-R, each policy calls for lower emissions and lower $PM_{2.5}$ concentrations than with a log-linear C-R. If the true C-R is log-log lower emissions and concentrations are preferred to take advantage of the larger mortality risk reductions possible in the cleaner locations.

For both C-R relationships, the efficient abatement policy achieves substantially greater welfare for society than the uniform pollution standard. This result highlights the importance of regulating emissions that impact more than just those that face the greatest risks. Finally, substantial advantages exist with the efficient policy compared with the uniform emissions tax. While the unform tax achieves a cost-effective outcome, the efficient policy differentiates between the damages of emissions at each location. The spatial heterogeneity of marginal benefits of abatement indicates the importance of applying source-specific regulations. One possibile approach to implement an approximation of the efficient abatement policy, with low information requirements of abatement costs by regulators, is a set of discriminating emissions taxes, different for each source equal to the marginal damage of emissions.

In Chapter 4 the marginal damages of emissions of certain criteria pollutants are estimated using a newly developed air dispersion model for the U.S. The impact of emissions of $PM_{2.5}$, SO_X, NO_X and NH_3 (pollutants that contribute to the total fine particulate concentration), are modeled from hundreds of the largest elevated and ground-level sources of emissions across the U.S. The estimates of the marginal damages of an additional ton of emissions show orders of magnitude differences depending on source location and pollutant. In addition, damages in receptors far from the source of emission can be substantial, highlighting the interconnected nature of the many sources of emissions that contribute to the $PM_{2.5}$ concentrations in a receptor.

Marginal damages were calculated for each source with the both the Krewski *et al.* (2009) log-linear and log-log C-R relationships. In 2005, the baseline year modeled, fine particulate concentrations were substantially higher than today, and marginal damages with a log-linear and log-log C-R were very similar. However, when the calculations were made with lower 2013 $PM_{2.5}$ concentrations, the marginal damages with a log-log C-R were much larger than with a log-linear C-R. At lower fine particulate concentrations, the value of additional emission reductions are larger if the true C-R is log-log.

The results found in Chapters 2, 3 and 4 indicate that using a log-log C-R (or increasing marginal benefits of abatement) can substantially change the analysis of $PM_{2.5}$ air pollution impacts and regulation. Identifying the true shape of the C-R between exposure to fine particulates and mortality thus has significant implications for society.

Regardless of the shape of the C-R, fine particulate air pollution is a significant risk to human health that requires attention from epidemiologists and economists. Despite the substantial improvements in air quality in the United States over the last several decades, additional reductions may be required.

Chapter 2

Prices *vs.* Quantities With Increasing Marginal Benefits*

*Chapter 2 originally as unpublished manuscript with authors:

Andrew L. Goodkind, Jay S. Coggins, Timothy A. Delbridge, Milda Irhamni, Justin Andrew Johnson, Suhyun Jung, Julian D. Marshall, Bijie Ren, Martha H. Rogers and Joshua S. Apte.

2.1 Introduction

Environmental policy is often built upon quantity restrictions. In the U.S., at least, these usually take the form of direct quantity standards, with a few noteworthy allowance-trading schemes and emissions taxes mixed in. In this paper we offer new reasons to favor taxes. We find that, for the model under study, quantity restrictions are never preferred and the relative advantage to taxes can be large. In our theoretical model, taxes are preferred because of the flexibility they grant to the polluting industry. The greater the level of uncertainty regarding control costs, the greater the advantage to taxes. Recycling of tax revenue, which we ignore, would tilt the comparison still more decisively in favor of taxes.

In the standard economic model of environmental policy, one assumes that marginal benefits are decreasing or, in the limit, constant in abatement.[1] One assumes too that marginal costs are increasing in abatement.[2] In this setting, the obvious optimum is found where marginal benefits and marginal costs meet. Familiar textbook treatments of the problem adhere to the standard view. Baumol and Oates (1988, p. 59), to take one prominent example, justify their assumption of downward-sloping marginal benefits this way: "In accord with the usual observations, [MB] has a negative slope, indicating that the greater the degree of purity of air or water that has already been achieved, the less the marginal benefit of a further 'unit' of purification."

This understanding of the curvature of abatement benefits is no longer obviously correct. A benefit function for air pollution abatement is, after all, a reduced form in which is embedded a good deal of nontrivial science. The mapping runs first from abatement to a change in ambient pollution concentration, next from a change in concentration to a change in health outcomes, and finally from a change in health outcomes to monetary benefits.[3] If (*i*) the first and third mappings are approximately linear; (*ii*)

[1]The limiting case of linear benefits, constant marginal benefits, is employed by Muller and Mendelsohn (2009).

[2]But see Andreoni and Levinson (2001), who argue that the production function for abatement of fine particulates exhibits increasing returns to scale, and thus that marginal costs are decreasing in abatement.

[3]This emphasis on health outcomes reflect that fact that the lion's share of benefits associated with air-quality rules flows from improvements in human health. Of that, the lion's share flows from avoided mortality. If the curvature of mortality benefits goes against type, it is unlikely that other components such as benefits to avoided hospitalization and morbidity will reverse it.

the mapping from concentration to health outcomes (a concentration-response function that relates mortality or another negative health outcome to pollution levels) takes the classical logistic form; and (*iii*) the range of concentration relevant to air policy falls in the convex part of the curve; then one does indeed find support for the "usual observations." The first unit of abatement is most valuable, the last unit least valuable: marginal benefits are decreasing in abatement.

To be fair to Baumol and Oates, their view reflected the conventional scientific wisdom of the time. Dockery *et al.* (1993) and Pope *et al.* (1995), showed for the first time that there is no safe minimum level of concentration of fine particulates ($PM_{2.5}$). Concentration-response (C-R), both papers said, is not logistic but linear. Health damage goes all the way down to the lowest observed concentration. Recent evidence suggests that linearity itself is now questioned. The C-R, some prominent studies suggest, is strictly concave. Not only is the first unit of concentration not safe, it is the most harmful.

Unexpected benefits curvature can arise from any component of the mapping from abatement to benefits. Nonlinearity due to atmospheric chemistry in the first mapping is a distinct possibility. A curious example, featured in Repetto (1987) and in Muller and Mendelsohn (2012), is that of ozone. Titration of ozone by excess nitrogen oxides means that, in conditions where NO_X is plentiful, abatement of NO_X can lead to *increased* concentration of harmful ozone. Here we have negative and increasing marginal abatement benefits for the first units of abatement, at least for a portion of the downwind landscape.[4]

If, contrary to the ozone case, the rest of the problem *is* approximately linear, the curvature of benefits is dual to that of the underlying C-R. Here, if the C-R curve is strictly concave in concentration, then the first unit of concentration is the most harmful, the last unit of abatement the most valuable. In such a case benefits are strictly convex and marginal benefits therefore slope upward in abatement.

This possibility, an unusual observation, forms the intellectual basis for the present

[4]Strange curvature and increasing marginal benefits arise in other situations as well, at least over a range of the given policy choice. Examples due to externalities of an intervention include incidence rates for vaccination (Boulier, Datta, and Goldfarb 2007) and for bed nets to prevent malaria (Hawley *et al.* 2003). In Anderson, Laxminarayan, and Salant's (2012) dynamic model of the optimal expenditure of a treatment budget for an infectious disease in multiple villages, unexpected curvature and surprising corner solutions result.

paper. Why might it be interesting? For $PM_{2.5}$, according to a growing chorus of environmental-health experts the curvature of the C-R for this deadly pollutant might be strictly concave.[5] Examples include Ostro (2004), Pope *et al.* (2009a, 2011), and Smith and Peel (2010). These studies rely in part on data representing very high concentrations, for active smokers, that lie outside the range of ambient concentrations observed in U.S. cities.

More compelling evidence, because it relies exclusively upon ambient levels of $PM_{2.5}$ that are directly relevant to U.S. clean-air policy, is found in Crouse *et al.* (2012) and Krewski *et al.* (2009). The Crouse *et al.* study, in which $PM_{2.5}$ concentrations range from 1.9 to 19.2μg/m^3, fits natural-spline and logarithmic C-R functions, for four causes of mortality, to a large cohort of Canadian residents. For three of the four categories they cannot reject a linear relationship, but for the fourth (ischemic heart disease) they reject linearity in favor of strict concavity. There, the first unit of concentration is the most harmful, the last unit of abatement the most valuable.[6]

Krewski *et al.* (2009) is especially important because of the role their results play in the U.S. Environmental Agency's recent (2012b) Regulatory Impact Analysis supporting a new proposed national standard for $PM_{2.5}$ concentration. Krewski *et al.* contains an extended analysis of the influential American Cancer Society longitudinal study (Pope *et al.* 1995) of the effects of air pollution on human health. In the extended analysis, where $PM_{2.5}$ concentrations range as low as 5 μg/m^3, Krewski *et al.* estimated a log-linear and a log-log C-R relating $PM_{2.5}$ to five causes of death (all causes, cardiopulmonary disease, ischemic heart disease, lung cancer, and all other causes). Their point estimate of 1.06 for the constant hazard ratio (HR) assocated with all causes of death (see Table 11 of Krewski *et al.* 2009, p. 28) serves as an essential parameter in the EPA's (2012b, p. 5-27) $PM_{2.5}$ benefits assessment.[7]

[5]Strict concavity is neither unusual nor, apparently, controversial in the case of environmental health related to toxins found in workplaces, a threat that lies outside the purview of the U.S. EPA. See, for example, Steenland *et al.* (2011) and especially Stayner *et al.* (2003).

[6]In the Crouse *et al.* cohort of 2.1 million subjects, ischemic heart disease claimed 43,400 lives between 1991 and 2001. The corresponding number for all non-accidental causes is 192,300.

[7]In the U.S. EPA $PM_{2.5}$ RIA, the 1.06 linear HR estimate from Krewski *et al.* (2009) helps determine the lower end of the range of EPA's benefits estimate. The corresponding number from Laden *et al.* (2006) is 1.16, which helps determine the higher end of the range of EPA's benefits estimate. Average concentrations in the six cities, all in the eastern U.S., are higher than in the much larger ACS cohort.

The three right-most columns in Krewski *et al.*'s Table 11 report the results of estimating a log-log function for each cause of death. These results reflect a strictly concave fitted relationship: the first units of concentration are the most harmful.[8] Though they do not report the results of a statistical test to determine which function gives the better fit, Krewski *et al.* (2009, p. 27) observe that the log-log form "was a slightly better predictor of the variation in survival."[9] The EPA chose to use the results of the log-linear rather than the log-log model.

Let us state carefully what we endeavor to claim: that a strictly concave C-R, and attendant increasing marginal benefits, might sometimes be true, for some important pollutants and some health endpoints. The scientific results are mixed, and so a stronger claim on our part would be unwarranted. Our point, though, is that this science is unsettled and relatively new. The experts disagree.[10] The evidence conflicts.

Our paper is rooted in the following question: *What if* marginal benefits of abatement are increasing. What then is the proper policy response, and which if any of our most familiar recommendations need to be reconsidered?

We examine the effect of increasing marginal benefits on the comparison of tax and quantity instruments in environmental policy (Weitzman 1974). Our focus is not on technology or dynamics or on how different permit-market arrangements affect policy choice.[11] Rather, in the presence of increasing marginal benefits we study how corner

[8]It is perhaps worth noting the terminology adopted by Krewski *et al.* in describing their results. The column headings in their Table 11 are given as "Linear" and "Log." These labels refer to the way in which the $PM_{2.5}$ variable enters the right side of their regression models. It is either untransformed (the linear model) or log-transformed (the log model). In both cases, however, the dependent variable is the log of the hazard ratio. (See the unnumbered equations on their p. 27.) Here we adopt the language "log-linear" and "log-log" to refer to the two alternatives. Over the range of ambient $PM_{2.5}$ concentrations, their log-linear results yield a C-R that is very nearly linear and their log-log results yield a C-R that is markedly concave.

[9]Noting the importance of the choice of functional form, they write (2009, p. 28), "The choice of functional relationship between PM exposure and mortality can make a considerable difference in the predicted risk at lower concentrations. For example, the HR for lung cancer adjusted for the ecologic covariates based on the [log-linear] formulation is 1.142 (95% CI, 1.057–1.234), whereas the HR based on the [log-log] formulation is 1.236 (95% CI, 1.114–1.372), a 66% increase in risk." The corresponding increase for all causes, from 1.078 to 1.128, is 64%.

[10]For an interesting glimpse into the way the scientists talk to each other about curvature, see Schwartz (2011).

[11]Karp and Zhang (2012) present a nice overview of the large literature related to the question. They study the problem when a regulator behaves strategically. See also Jaffe *et al.* (2003). Various aspects of the problem of instrument choice are take up by, among others, Moledina *et al.* (2003), Fisher *et al.* (2003), and Tarui and Polasky (2005). The connection between technology adoption and instrument is

solutions affect the usual Weitzman-style results.

Given an initial level of air quality, we assume that the marginal benefits of abatement and the marginal costs of abatement are both increasing in abatement. We allow for the possibility that reducing emissions to zero may be technologically infeasible, perhaps due to background natural sources. In this case, "complete" abatement means a reduction to the minimum feasible level. Expected marginal costs are assumed to meet marginal benefits where abatement is positive but not complete. Then we analyze two sets of cases: those in which marginal benefits are less steep than marginal costs (we call this case "crossing from above"); and those in which marginal benefits are steeper than marginal costs ("crossing from below").

A model of optimal abatement policy, largely familiar excepting the slope of marginal benefits, is outlined in the following section. There we explain two additional assumptions, which are adopted throughout: (*i*) linearity of marginal costs and benefits; and (*ii*) a uniform distribution on uncertainty. The first is common in the literature since the work of Adar and Griffin (1976). The second is less common, but our wish to characterize the optimal tax requires the selection of a specific functional form for the denisity on the stochastic term. (A normal distribution offers advantages, but it defeats attempts at analytical solutions.) We believe that our results survive in a more general setup, but we defer a detailed analysis of that situation to future work.

Sections 2.3 and 2.4 contain a preliminary analysis of first the optimal quantity policy and then the optimal price policy. If marginal benefits are steeper than marginal costs, we show that the optimal quantity policy is either zero or maximum possible abatement. The optimal price policy is more complicated because for a given tax it is possible that for some realizations of the stochastic term the industry will choose either zero abatement or maximum possible abatement. The corners that come into play add a significant degree of complexity to the tax-setting regulator's optimization problem.

Section 2.5 addresses the problem for the situation of marginal benefits crossing from above. In this case the optimal price is always preferred. This finding is not at odds with Weitzman, but we find that if the level of uncertainty becomes sufficiently large, the optimal emissions tax is no longer at the intersection of marginal benefits and

explored by Montero 2002, Milliman and Prince (1989), Biglaiser *et al.* (1995), Gersbach and Glazer (1999), and Requate and Unold (2003).

expected marginal costs.

Section 2.6 contains the heart of the paper. There marginal benefits cross marginal costs from below. The all-or-nothing nature of the quantity policy means that the welfare stakes are especially high, in that the regulator might choose a policy of zero abatement when the optimal policy was maximum possible abatement. If the level of uncertainty is low, the price and quantity policies are equivalent. If the level of uncertainty is sufficiently high, the price policy becomes strictly preferred. At the threshold, the optimal tax can jump discretely from extreme (either high or low) to an intermediate level.

The intuition for this result goes as follows. If the regulator had perfect information about costs, the appropriate policy, either zero abatement or maximum possible abatement, could be selected with confidence. This is essentially the outcome achieved by either policy if uncertainty is low. If the level of uncertainty is high, though, the quantity policy is likely to produce the incorrect level of abatement. The wrong choice, in either direction, can be quite costly. An intermediate tax, however, allows the regulator to exploit the fact that the industry will choose low abatement (if the realization of uncertainty is high) precisely when zero abatement was optimal and will choose high abatement (if the realization of uncertainty is low) precisely when maximum possible abatement was optimal. The tax policy confers an advantage via the flexibility granted to the industry, whose interests are, in a crucial sense, aligned with those of the regulator.

2.2 The basic model

Consider the problem facing a regulator who contemplates limiting emissions of a single pollutant from a single polluting industry. Current total emissions are $e^T > 0$. Due to technological constraints the minimum achievable level of emissions is $e^{\min} > 0$. The corresponding maximum level of abatement is denoted $e^0 < e^T$ and abatement is $a \in [0, e^0]$. (Abatement levels greater than e^0 are ignored throughout, and abatement at e^0 is referred to as complete abatement.) The nonstochastic benefit to abatement is described by the quadratic function $B : [0, e^T] \to \mathbb{R}_+$ given by $B(a) = \alpha a + (\beta/2)a^2$, with $\alpha \geq 0$, $\beta > 0$, and $B(0) = 0$. Marginal benefits are written

$$\text{MB}(a) = \alpha + \beta a$$

and are known by the regulator with certainty.

The industry's cost of abatement is quadratic on $[0, e^0]$, but because further abatement is infeasible (and so infinitely costly) we ignore $a > e^0$ and write $C : [0, e^0] \times \mathbb{R} \to \mathbb{R}$. The cost function depends upon abatement and upon a random variable u, and is additively separable in its two arguments: $C(a, u) = \eta a + (\delta/2)a^2 + ua$, with $\eta \geq 0$, $\delta > 0$ and $C(0, u) = 0$ for any u. Marginal costs are written

$$\text{MC}(a, u) = \eta + \delta a + u.$$

Let the vector of structural parameters be denoted $\theta = (\alpha, \beta, \eta, \delta, e^0)$.

Uncertainty enters only through the distribution on the intercept of marginal costs. Let the support for the intercept be on the finite interval $[\eta - \nu, \eta + \nu]$, with $\nu > 0$. The density function for u within this interval is assumed to be uniform, with density $f(u) = 1/2\nu$ and with $E(u) = 0$. We assume that the regulator knows the density function $f(u)$. Social welfare is quadratic in abatement, and is given by

$$\text{SW}(a, u) = B(a) - C(a, u)$$

We shall assume throughout that marginal benefits and expected marginal costs intersect exactly once, in the interior of $[0, e^0]$. We then distinguish between situations in which marginal benefits cross marginal costs from above and from below. These are formalized as follows.

Assumption 1. (Crossing-from-above condition.) There exists $\hat{a} \in (0, e^0)$ such that $\text{MB}(a) > E\big(\text{MC}(a, u)\big)$ for all $a \in [0, \hat{a})$ and $\text{MB}(a) < E\big(\text{MC}(a, u)\big)$ for all $a \in (\hat{a}, e^0]$.

Assumption 2. (Crossing-from-below condition.) There exists $\hat{a} \in (0, e^0)$ such that $\text{MB}(a) < E\big(\text{MC}(a, u)\big)$ for all $a \in [0, \hat{a})$ and $\text{MB}(a) > E\big(\text{MC}(a, u)\big)$ for all $a \in (\hat{a}, e^0]$.

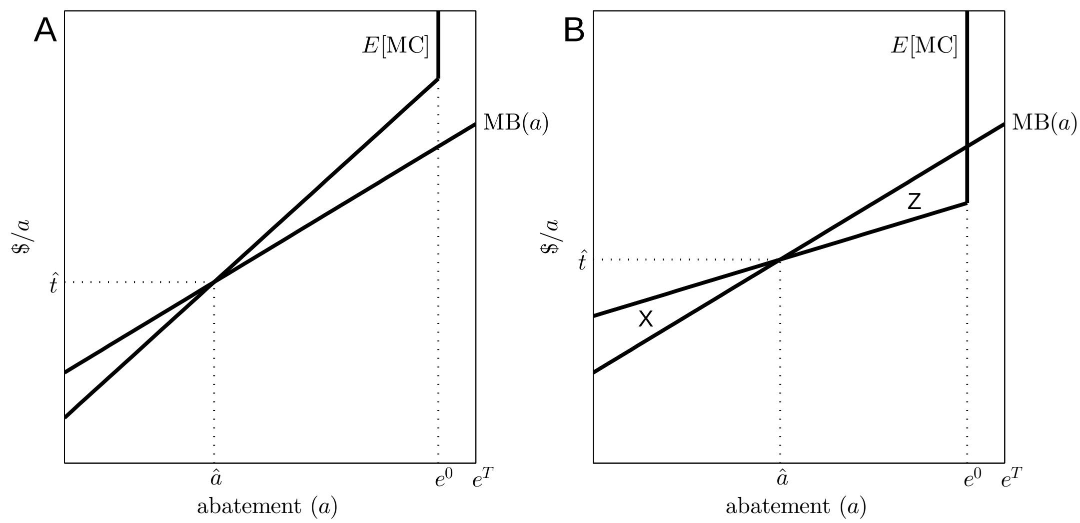


Figure 2.1: Depiction of upward-sloping marginal benefits. **A**: crossing from above. **B**: Crossing from below.

Assumption 1, which implies that $\alpha > \eta$ and $\beta < \delta$, is depicted in Figure 2.1A, where $\mathrm{MC}(a, 0)$ represents marginal cost with degenerate u. A quantity-setting regulator in this situation would maximize expected social welfare by setting abatement at $\hat{a}$. A price-setting regulator would maximize expected social welfare by setting an emissions tax at or near $\hat{t}$.

Assumption 2, which implies that $\alpha < \eta$ and $\beta > \delta$, is depicted in Figure 2.1B. There, the optimal abatement level depends upon whether area X is less than or greater than area Z. In either case, abatement at $\hat{a}$, the crossing point, results in negative social welfare, the minimum to the regulator's optimization problem. This will be true of the intersection any time the curves satisfy Assumption 2. We will see that under Assumption 2, with uncertainty the optimal emissions tax is ill-behaved in the face of uncertain costs.

2.3 The optimal quantity policy

Suppose that the random term in $\mathrm{MC}(a, u)$ is degenerate at $u = 0$. Under Assumption 1, the problem without uncertainty is straightforward. The optimal quantity policy is set where $\mathrm{MB}(a) = \mathrm{MC}(a, 0)$. Under Assumption 2, the optimal quantity policy is either maximum possible abatement or no abatement. In either case, there exists a critical value η^* such that the social welfare associated with maximum possible abatement is zero. This value is given by

$$\eta^* = \frac{e^0}{2}(\beta - \delta) + \alpha. \tag{2.1}$$

The optimal quanity policy is defined by the relative value of the intercept of the expected marginal cost curve and $\eta*$

$$q^*(\eta) = \begin{cases} e^0 & \text{if } \eta < \eta^* \\ \{0, e^0\} & \text{if } \eta = \eta^* \\ 0 & \text{if } \eta > \eta^*. \end{cases} \tag{2.2}$$

With uncertainty in marginal cost, whether she chooses to pursue a price policy or a quantity policy, our regulator is assumed to maximize expected social welfare. The optimal quantity policy is based entirely upon the intersection of marginal benefits and *expected* marginal costs. The quantity-setting regulator's optimal decision rule $q^*(\nu;\theta)$ maximizes expected social welfare:

$$q^*(\nu;\theta) = \underset{q \in [0, e^0]}{\operatorname{argmax}} \; E_u \left[\int_0^q \Big(\mathrm{MB}(a) - \mathrm{MC}(a, u) \Big) \, da \right]. \tag{2.3}$$

This constraint set for q is compact in $\mathbb{R}$ and the objective function is continuous in q. By the Weierstrass Theorem it achieves a maximum. If marginal benefits cross from above (Assumption 1 holds), the optimal quantity policy is at $\hat{a}$, where the two curves meet. If marginal benefits cross from below (Assumption 2 holds), the optimal quantity policy is once again given by equation (2.2) and is either zero or e^0.

2.4 The optimal price policy

If there is no uncertainty, the price policy, an emissions tax of t, mimics the quantity policy. If marginal benefits cross from above, it is set where $\mathrm{MB} = E[\mathrm{MC}(a, u)]$. If marginal benefits cross from below, it is set either where $t \geq \mathrm{MC}(e^0)$ (if $\eta < \eta^*$) or where $t \leq \mathrm{MC}(0)$ (if $\eta > \eta^*$). Because any tax at or less than $\mathrm{MC}(0)$ is equivalent, and any tax at or greater than $\mathrm{MC}(e^0)$ is equivalent, we limit attention to those inner thresholds.[12] For the knife-edge case with $\eta = \eta^*$, the regulator is indifferent between the upper and lower taxes. The optimal tax policy with no uncertainty is

$$t^*(\eta) = \begin{cases} \mathrm{MC}(e^0) & \text{if } \eta < \eta^* \\ \big\{\mathrm{MC}(0), \mathrm{MC}(e^0)\big\} & \text{if } \eta = \eta^* \\ \mathrm{MC}(0) & \text{if } \eta > \eta^*. \end{cases} \tag{2.4}$$

If there is uncertainty in costs, the price policy depends upon more than expected marginal cost, for the industry's chosen level of abatement depends upon the realization of u. The abatement response creates a price corner that is quite different from the all-or-nothing quantity corner that arises, even without uncertainty, when Assumption 2 is satisfied.

Clearly, if $t \leq \mathrm{MC}(0, -\nu)$ the outcome will be zero abatement. Define $t_{\min} = \mathrm{MC}(0, -\nu)$ and note that any $t \leq t_{\min}$ serves as a zero-abatement policy. We ignore $t < t_{\min}$ as redundant. If $t \geq \mathrm{MC}(e^0, \nu)$ the outcome will be maximum possible abatement. Define $t_{\max} = \mathrm{MC}(e^0, \nu)$ and note that any $t \geq t_{\max}$ serves as a full-abatement policy. We ignore $t > t_{\max}$ as redundant. The interesting cases lie between these two extremes. There, the industry will attempt to respond to t by choosing $\tilde{a}$ so that $t = \mathrm{MC}(\tilde{a}, u)$. Given that marginal cost, being continuous and strictly monotone increasing, is invertible, on the interval $u \in \Big[t - E\big[\mathrm{MC}(e^0, u)\big], t - E\big[\mathrm{MC}(0, u)\big]\Big]$ the function describing the abatement level at which realized marginal cost equals t is given by

$$\tilde{a}(t, u) = \frac{t - \eta - u}{\delta}. \tag{2.5}$$

[12] Here we hew to the literature in assuming implicitly that there is no deadweight loss associated with taxation. The tax revenue of $t^* e^0$ in the zero-abatement situation is simply a transfer from polluters to the regulator. Under Assumption 2, the question of deadweight taxation losses can be mooted by imposing a tax of zero whenever $\eta > \eta^*$. At the other extreme, of course, if abatement is at e^0 there are no emissions and so no tax is paid at all.

For the industry, the optimal abatement level given t is therefore

$$a^*(t,u) = \begin{cases} e^0 & \text{if } u \leq t - E\big[\text{MC}(e^0,u)\big] \\ \tilde{a}(t,u) & \text{if } t - E\big[\text{MC}(e^0,u)\big] < u < t - E\big[\text{MC}(0,u)\big] \\ 0 & \text{if } u \geq t - E\big[\text{MC}(0,u)\big]. \end{cases} \tag{2.6}$$

The interval of interesting taxes may be partitioned usefully, as illustrated in Figure 2.2. Consider first Figure 2.2A, where uncertainty is relatively low. For taxes below $t_{\max}$ but above $\text{MC}(e^0, -\nu)$ (this threshold is denoted $\bar{t}_4 = \underline{t}_3$ in the figure), the industry's abatement response will be either e^0 with positive probability (for low realizations of u) or interior (for high realizations). Denote by T_3 the interval of tax levels at which this corner can arise. One price corner, where abatement turns from the maximum possible to interior, resides in T_3. For taxes above $t_{\min}$ but below $\text{MC}(0, \nu)$ (this threshold is denoted $\underline{t}_4 = \bar{t}_2$ in the figure), with positive probability the industry's abatement response will be zero (for high realizations of u) or it will be interior (for low realizations). Denote by T_2 the interval of tax levels at which this corner can arise. Another price corner, where abatement turns from zero to interior, resides in T_2. Between T_2 and T_3 is a band of tax levels at which abatement is sure to be interior. Denote this interval T_4.

Now consider a situation in which uncertainty has risen. In Figure 2.2B, this is depicted as an increase in ν. If, as here, the increase is large enough that $\text{MC}(0, \nu) \geq \text{MC}(e^0, -\nu)$, then the inner endpoints of T_2 and T_3 switch places, though tax levels in the two intervals encounter the same corners as before. When it is nonempty, as in Figure 2.2B, the region between T_2 and T_3, now denoted T_1, is fundamentally different from the T_4 region in Figure 2.2A. At any $t \in T_1$, according to the response function in (2.6) there is positive probability of $a = 0$ (for high realizations of u), of interior abatement $a \in (0, e^0)$ (for intermediate realizations of u), and of $a = e^0$ (for low realizations of u). In fact, whenever $T_1 \neq \emptyset$ there is no tax at which abatement is guaranteed to be interior.[13]

The price corners are usefully depicted in Figure 2.3, which shows the industry's

[13] Avoiding the price corner, Weitzman (1974) considers only tax levels in T_4, where abatement is sure to be interior.

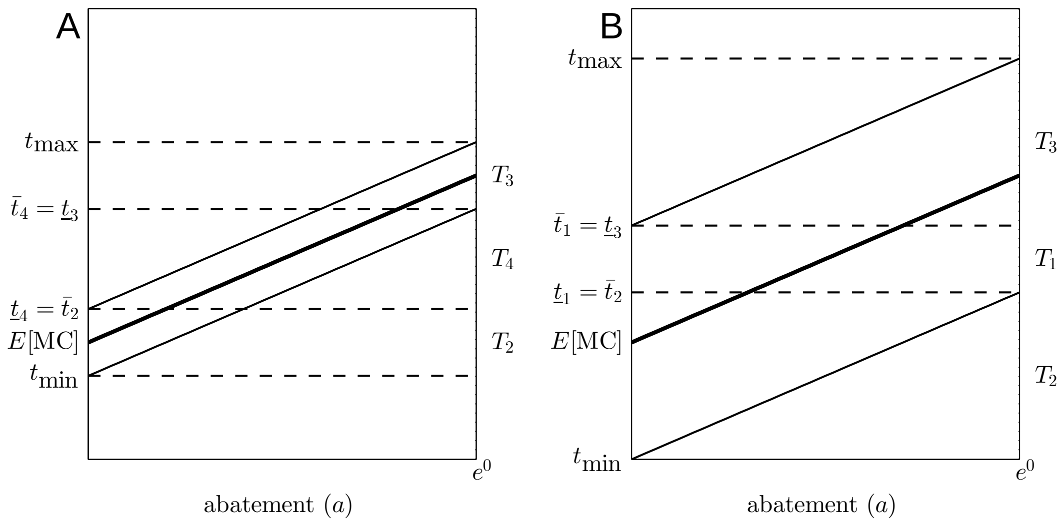


Figure 2.2: Tax ranges for four cases. **A**: uncertainty is low, T_4 is nonempty. **B**: uncertainty is high, T_1 is nonempty.

response to different tax levels. (The dashed lines represent possible realizations of marginal costs.) In Figure 2.3A, which corresponds to Figure 2.2A, three different taxes are illustrated. At $t' \in T_2$, abatement will be zero for high and interior for low realizations of u. At $t''' \in T_3$, abatement will be at e^0 or interior. At $t'' \in T_4$, abatement must be interior. In Figure 2.3B, which corresponds to Figure 2.2B, we see that at $t'' \in T_1$ all three responses are possible: zero, interior, or maximum possible abatement for high, intermediate, or low realizations of u.

More formally, take a permissible parameter vector (θ, ν), and define $T = [t_{\min}, t_{\max}]$. Partition T as described above, where the suprema and infima of the T_j are given by

$$\underline{t}_1 = \mathrm{MC}(e^0, -\nu) \qquad \bar{t}_1 = \mathrm{MC}(0, \nu) \tag{2.7a}$$

$$\underline{t}_2 = t_{\min} \qquad \bar{t}_2 = \min\left\{\mathrm{MC}(e^0, -\nu), \mathrm{MC}(0, \nu)\right\} \tag{2.7b}$$

$$\underline{t}_3 = \max\left\{\mathrm{MC}(e^0, -\nu), \mathrm{MC}(0, \nu)\right\} \qquad \bar{t}_3 = t_{\max} \tag{2.7c}$$

$$\underline{t}_4 = \mathrm{MC}(0, \nu) \qquad \bar{t}_4 = \mathrm{MC}(e^0, -\nu). \tag{2.7d}$$

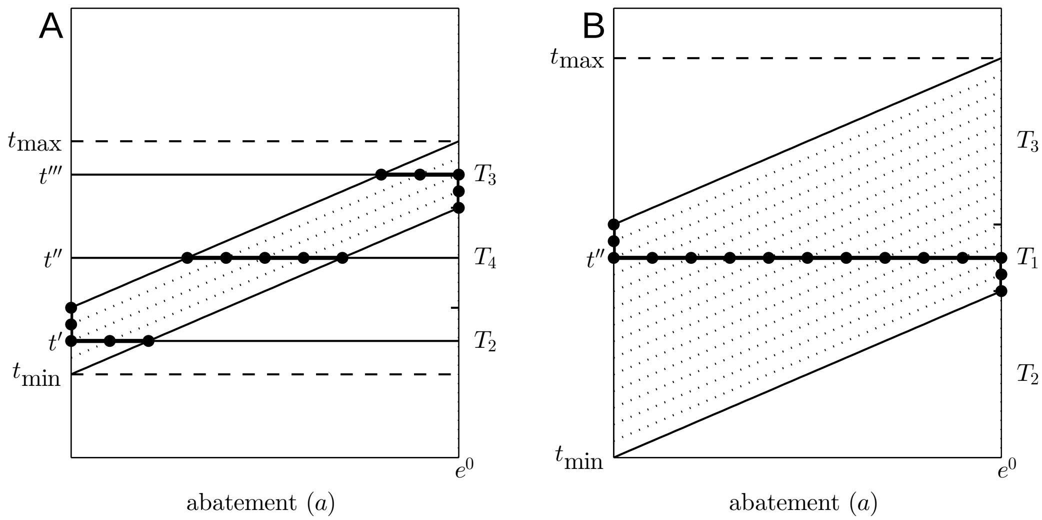


Figure 2.3: Abatement responses. **A**: T_4 is nonempty. Abatement is zero or interior at t'; interior at t''; complete or interior at t'''. **B**: T_1 is nonempty. Abatement is zero, interior or complete at t''.

The elements of the partition are

$$
T_1 = \begin{cases} [\underline{t}_1, \bar{t}_1] & \text{if } \underline{t}_1 \leq \bar{t}_1 \\ \emptyset & \text{otherwise} \end{cases} \qquad T_2 = \begin{cases} [\underline{t}_2, \bar{t}_2] & \text{if } T_1 = \emptyset \\ [\underline{t}_2, \bar{t}_2) & \text{if } T_1 \neq \emptyset \end{cases}
$$
$$
T_3 = \begin{cases} [\underline{t}_3, \bar{t}_3] & \text{if } T_1 = \emptyset \\ (\underline{t}_3, \bar{t}_3] & \text{if } T_1 \neq \emptyset \end{cases} \qquad T_4 = \begin{cases} (\underline{t}_4, \bar{t}_4) & \text{if } \underline{t}_4 < \bar{t}_4 \\ \emptyset & \text{otherwise.} \end{cases} \tag{2.8}
$$

Because of the way they are defined, $T_1 = \emptyset$ precisely when $T_4 \neq \emptyset$. This is true if $\mathrm{MC}(0, \nu) < \mathrm{MC}(e^0, -\nu)$. Conversely, $T_4 = \emptyset$ precisely when $T_1 \neq \emptyset$. This is true if $\mathrm{MC}(0, \nu) \geq \mathrm{MC}(e^0, -\nu)$. To see that equations (2.8) constitute a partition, note that $T_j \cap T_{j'} = \emptyset$ for all $j \neq j'$ and also that $\cup_j T_j = T$. The two-part definitions for T_2 and T_3 ensure that the partitions do not intersect, as uncertainty increases, at the point at which T_1 appears as a singleton set with $t = \mathrm{MC}(0, \nu) = \mathrm{MC}(e^0, -\nu)$.

Social welfare can be expressed, using (2.6), as a three-part function of t and u, according to whether abatement is at e^0, interior, or zero:

$$\mathrm{SW}(t,u) = \begin{cases} B(e^0) - C(e^0,u) & \text{if } u \le t - E\big[\mathrm{MC}(e^0,u)\big] \\ B\big(\tilde{a}(t,u)\big) - C\big(\tilde{a}(t,u),u\big) & \text{if } t - E\big[\mathrm{MC}(e^0,u)\big] < u < t - E\big[\mathrm{MC}(0,u)\big] \\ B(0) - C(0,u) & \text{if } u \ge t - E\big[\mathrm{MC}(0,u)\big]. \end{cases} \tag{2.9}$$

The tax-setting regulator's optimization problem may be obtained as follows. Integrate (2.9) over the support of u, separating the three subintervals in $[-\nu,\nu]$. One or two of these intervals may be empty for some values of t. The first integral in (2.10) is over all realizations of u that yield maximum possible abatement and the second is over realizations for which abatement is interior. The third integral, over realizations for which abatement is zero, yields zero expected welfare and so can be discarded. The optimal tax, which maximizes expected social welfare, is the solution correspondence

$$\begin{aligned} t^*(\nu;\theta) = \underset{t\in[t_{\min},t_{\max}]}{\operatorname{argmax}} & \int_{-\nu}^{\max\left\{-\nu,\min\left\{\nu,t-E\left[\mathrm{MC}(e^0,u)\right]\right\}\right\}} \big[B(e^0) - C(e^0,u)\big] f(u)\,du \\ & + \int_{\max\left\{-\nu,\min\left\{\nu,t-E[\mathrm{MC}(e^0,u)]\right\}\right\}}^{\min\left\{\nu,\max\left\{-\nu,t-E[\mathrm{MC}(0,u)]\right\}\right\}} \big[B\big(\tilde{a}(t,u)\big) - C\big(\tilde{a}(t,u),u\big)\big] f(u)du \\ & + \int_{\min\left\{\nu,\max\left\{-\nu,t-E[\mathrm{MC}(0,u)]\right\}\right\}}^{\nu} [0] f(u)du \end{aligned} \tag{2.10}$$

where $f(u) = 1/2\nu$. The $\min\{\cdot,\cdot\}$ and $\max\{\cdot,\cdot\}$ operators in the limits of integration ensure that each integral is evaluated over the correct interval. The outer $\max\{\cdot,\cdot\}$ function on the upper limit of the first integral, for example, determines whether the infimum of T_3 is given by $\mathrm{MC}(e^0,-\nu)$ (when $T_4 \neq \emptyset$) or by $\mathrm{MC}(0,\nu)$ (when $T_1 \neq \emptyset$). The inner $\min\{\cdot,\cdot\}$ function then determines whether the limit is determined by the upper boundary ν of u or by the value of u at which abatement moves into the interior. This is a price corner.

The constraint set for t in (2.10) is compact in $\mathbb{R}$ and the objective function is continuous in t. By the Weierstrass Theorem it achieves a maximum. But the task of selecting the optimal tax is not trivial. This is because of the non-differentiable functions, containing the choice variable t, that define the limits of integration. Were those functions differentiable, first-order necessary conditions for (2.10) could be obtained via

Leibniz's rule. Because they are not, the only available strategy for deriving the optimal tax is to define separate functions for expected social welfare on the T_j. For $t \in T_1$ there are possible tax levels in all three of the components of integral (2.10). For $t \in T_2$ there is no realization of u at which abatement will be e^0. There, only the second and third integrals in (2.10) are relevant. For $t \in T_3$ there is no realization of u at which abatement will equal zero. The first two components of (2.10) are relevant. For $t \in T_4$ it must be true that $a \in (0, e^0)$, so only the second component of (2.10) is relevant.

Define the following four functions describing expected social welfare on the corresponding T_j:

$$\Gamma_1 : T_1 \to \mathbb{R}, \text{ with } \Gamma_1(t) = \int_{-\nu}^{t-E[\mathrm{MC}(e^0,u)]} \big[B(e^0) - C(e^0,u)\big] f(u)du + \int_{t-E[\mathrm{MC}(e^0,u)]}^{t-E[\mathrm{MC}(0,u)]} \big[B\big(\tilde{a}(t,u)\big) - C\big(\tilde{a}(t,u),u\big)\big] f(u)du \tag{2.11}$$

$$\Gamma_2 : T_2 \to \mathbb{R}, \text{ with } \Gamma_2(t) = \int_{-\nu}^{t-E[\mathrm{MC}(0,u)]} \big[B\big(\tilde{a}(t,u)\big) - C\big(\tilde{a}(t,u),u\big)\big] f(u)du \tag{2.12}$$

$$\Gamma_3 : T_3 \to \mathbb{R}, \text{ with } \Gamma_3(t) = \int_{-\nu}^{t-E[\mathrm{MC}(e^0,u)]} \big[B(e^0) - C(e^0,u)\big] f(u)du + \int_{t-E[\mathrm{MC}(e^0,u)]}^{\nu} \big[B\big(\tilde{a}(t,u)\big) - C\big(\tilde{a}(t,u),u\big)\big] f(u)du \tag{2.13}$$

$$\Gamma_4 : T_4 \to \mathbb{R}, \text{ with } \Gamma_4(t) = \int_{-\nu}^{\nu} \big[B\big(\tilde{a}(t,u)\big) - C\big(\tilde{a}(t,u),u\big)\big] f(u)du. \tag{2.14}$$

The functions Γ_1 and Γ_4 in (2.11) and (2.14) are quadratic in t; Γ_2 and Γ_3 in (2.12) and (2.13) are cubic in t. All four are therefore continuously differentiable in t. Piecing them together yields the function describing expected social welfare on all of T. This function is also continuously differentiable, though there may be multiple local maxima and minima:

$$E[\mathrm{SW}(t,u)] = \begin{cases} \Gamma_1(t) & \text{if } t \in T_1 \\ \Gamma_2(t) & \text{if } t \in T_2 \\ \Gamma_3(t) & \text{if } t \in T_3 \\ \Gamma_4(t) & \text{if } t \in T_4. \end{cases}$$

It is useful to isolate two classes of situations: those in which $T_4 = \emptyset$ and those in which $T_1 = \emptyset$. In either case, the relationships between equations (2.11)–(2.14) at the boundaries of the T_j ensure that the price-setting regulator's problem is well behaved: the $E[\mathrm{SW}(t,u)]$ function is twice differentiable.

Our goal is to identify the values of t at which expected social welfare might be maximized. This requires finding all values at which the derivatives of (2.11)–(2.14) are equal to zero. Differentiating all four expressions with respect to t, setting the results equal to zero, and solving for t in each case produces six candidate zeros and leads to the following collection of candidate optima:

$$\hat{t}_1 = \alpha + \frac{\beta e^0}{2} \tag{2.15}$$

$$\hat{t}_{2a} = \eta - \nu \qquad \text{and} \quad \hat{t}_{2b} = \frac{\beta(\eta - \nu) - 2\alpha\delta}{\beta - 2\delta} \tag{2.16}$$

$$\hat{t}_{3a} = \eta + \delta e^0 + \nu \quad \text{and} \quad \hat{t}_{3b} = \frac{\beta(\eta + \nu) - \beta\delta e^0 - 2\alpha\delta}{\beta - 2\delta} \tag{2.17}$$

$$\hat{t}_4 = \frac{\beta\eta - \alpha\delta}{\beta - \delta}. \tag{2.18}$$

Note that neither $\hat{t}_1$ nor $\hat{t}_4$ depends upon ν, and $\hat{t}_4$ is the tax that equates marginal benefits and expected marginal costs. In all cases, we have $t_{\min} \equiv \hat{t}_{2a}$ and $t_{\max} \equiv \hat{t}_{3a}$. Depending on the parameter vector, some of the conditions in (2.15)–(2.18) might describe either a local maximum or a local minimum. Separating them, essential in order to identify the optimal policy overall, involves analyzing a set of second-order sufficient conditions. That analysis, lengthy and somewhat tedious, is available in Appendix B.

2.5 Prices *vs.* quantities: Crossing from above ($\beta < \delta$)

Suppose that Assumption 1 is satisfied, so that marginal benefits cross marginal costs from above. That is, $\beta < \delta$. In this case the following threshold values of ν are relevant. [See Appendix B.] They are important in determining (2.23), the function describing the optimal tax:

$$\nu_{1A} = E[\mathrm{MC}(e^0, u)] - \mathrm{MB}(e^0) - \frac{e^0\beta}{2} \tag{2.19}$$

$$\nu_{1B} = \mathrm{MB}(0) - E[\mathrm{MC}(0, u)] + \frac{e^0\beta}{2} \tag{2.20}$$

$$\nu_{4A} = \frac{\delta[E[\mathrm{MC}(0, u)] - \mathrm{MB}(0)]}{\beta - \delta} \tag{2.21}$$

$$\nu_{4B} = \frac{\delta[\mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)]]}{\beta - \delta}. \tag{2.22}$$

As ν increases beyond the relevant thresholds, progressively more probability is pushed away from the neighborhood of the intersection of $\mathrm{MB}(a)$ and $E[\mathrm{MC}(a, u)]$ and, eventually, piles up at $a = 0$ and at $a = e^0$. There, the price policy is unambiguously preferred to the quantity policy, which is locked immovably at $q^* = \hat{a}$. More probability at the corners means that the advantage to the price policy is greater.

The optimal tax rule under Assumption 1 is given by

$$t^*_{[\beta<\delta]}(\nu; \theta) = \begin{cases} \hat{t}_1 & \text{if } [\,\eta \geq \eta^* \text{ and } \nu \geq \nu_{1A}\,] \text{ or } [\,\eta \leq \eta^* \text{ and } \nu \geq \nu_{1B}\,] \\ \hat{t}_{2b} & \text{if } [\,\eta > \eta^* \text{ and } \nu \in [\nu_{4A}, \nu_{1A})\,] \\ \hat{t}_{3b} & \text{if } [\,\eta < \eta^* \text{ and } \nu \in [\nu_{4B}, \nu_{1B})\,] \\ \hat{t}_4 & \text{if } [\,\eta \geq \eta^* \text{ and } \nu < \nu_{4A}\,] \text{ or } [\,\eta \leq \eta^* \text{ and } \nu < \nu_{4B}\,]. \end{cases} \tag{2.23}$$

Recall that at the threshold η^*, found in equation (2.1), expected social welfare at maximum possible abatement is zero. If $\eta < \eta^*$, expected social welfare at maximum possible abatement is positive; if $\eta > \eta^*$ it is negative. The elements that make up equation (2.23) may be divided along these lines.

The optimal price is $\hat{t}_4$ when the level of uncertainty ν is sufficiently small, with the threshold depending upon whether η is greater than or less than η^*. This price is at the intersection of marginal benefits and expected marginal costs. When ν is sufficiently large the optimal price, $\hat{t}_1$, diverges away from this intersection, with the threshold again depending upon whether η is greater than or less than η^*. The optimal price is intermediate, either $\hat{t}_{2b}$ or $\hat{t}_{3b}$, when ν is also intermediate.

Two primary results arise in this setting. The first, Proposition 1, is that a price policy strictly dominates a quantity policy, and the advantage grows with ν, when Assumption 1 holds. (Proofs of this and the remaining propositions appear in Appendix A.) Under the optimal price policy, the level of emissions selected by the polluting industry moves toward the optimum. Under the optimal quantity policy this adjustment is impossible. Expected social welfare under the quantity policy is unchanged as ν increases, but expected social welfare under the price policy grows as ν increases.

Proposition 1. *For any permissible parameter vector* (θ, ν) *at which* $\beta < \delta$*, the expected social welfare resulting from the optimal price policy is not less than that associated with the optimal quantity policy:* $E[SW(t^*(\nu;\theta), u)] \geq E[SW(q^*(\nu;\theta), u)]$*. If* $\nu > 0$*,* $E[SW(t^*(\nu;\theta), u)] - E[SW(q^*(\nu;\theta), u)]$ *is strictly positive and strictly increasing in* ν*.*

A central tenet of the literature comparing price and quantity policies is that the price policy should be set at $\tilde{t}$, where marginal benefits equal expected marginal costs. As the level of uncertainty increases, for the linear-uniform case examined here this is not always true. This is due to the fact that, for high levels of uncertainty, a price at the intersection cannot capture the welfare gains associated with realizations far from expected marginal costs. The following result establishes conditions under which the received wisdom is overturned.

Proposition 2. *Consider a permissible parameter vector* (θ, ν) *at which* $\beta < \delta$*.*

(i.) Suppose $\eta > \eta^*$*. If* $\nu > \nu_{4A}$ *then* $t^* > \tilde{t}$*.*
(ii.) Suppose $\eta < \eta^*$*. If* $\nu > \nu_{4B}$ *then* $t^* < \tilde{t}$*.*

Example 1. A numerical example might help to illuminate these results and the complex nature of the regulator's optimal tax-setting rule. For the linear-uniform situation under consideration, suppose that the parameter values are

$$\alpha = 80, \qquad \beta = 1, \qquad \eta = 60, \qquad \delta = 2, \qquad e^0 = 50.$$

Figure 2.4 shows the four $\Gamma_j(t)$ curves that combine to form the regulator's objective function $E[\mathrm{SW}(t, u)]$, which is the bold curve. It is constant outside of the interval $[t_{\min}, t_{\max}]$. This figure illuminates several important features of the optimal-price problem. One is that $\Gamma_2(t)$ and $\Gamma_3(t)$ are indeed cubic and $\Gamma_1(t)$ and $\Gamma_4(t)$ are

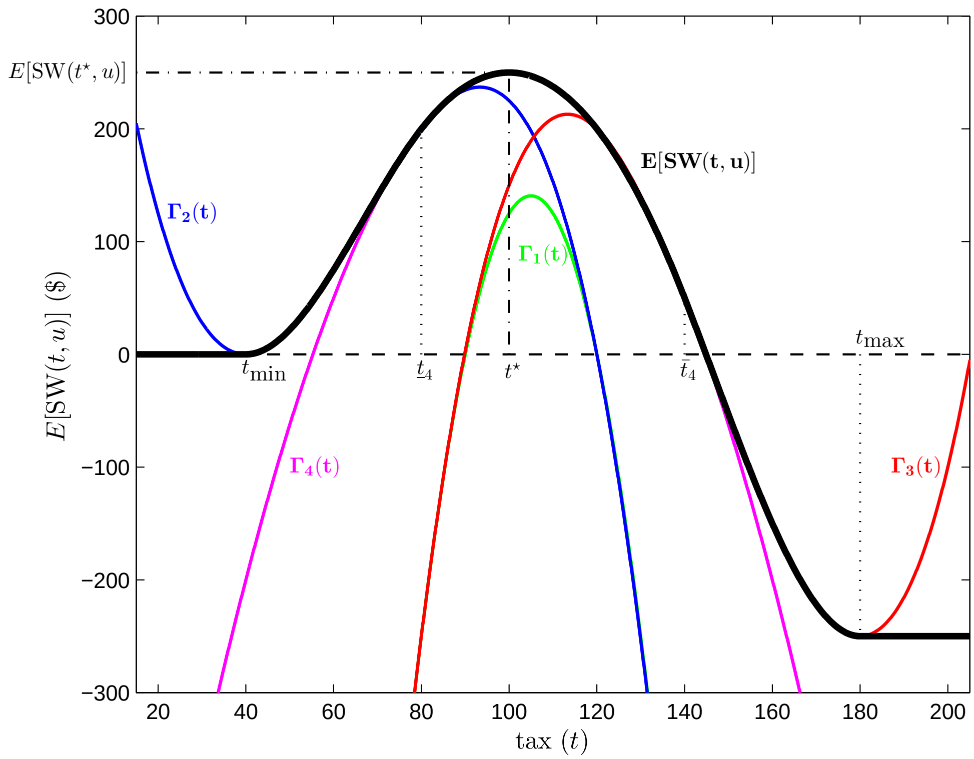


Figure 2.4: Outcome from all possible taxes, and the four Γ_j functions with $\nu = 20$, $T_1 = \emptyset$, $\alpha = 80$, $\beta = 1$, $\eta = 60 > \eta^\star$, and $\delta = 2$.

indeed quadratic functions. Another is that the three feasible SW curves join as claimed, and are also differentiable where they join at $\underline{t}_4 = 80$ and $\bar{t}_4 = 140$. Thus, the global $E[\mathrm{SW}(t,u)]$ function is differentiable. Note that, with $T_1 = \emptyset$, $\Gamma_1(t)$ is not a part of the $E[\mathrm{SW}(t,u)]$ function. In this example, $\nu = 20 < \nu_{4A}$, the optimal price is $\hat{t}_4$ at the intersection of marginal benefits and marginal costs. For higher values of ν, T_4 becomes empty and the T_1 interval comes into play.

Figure 2.5 shows how things change as ν, the level of uncertainty, increases. There, the five curves represent expected social welfare, as a function of the tax, for ν ranging from 10 to 60. For any $\nu > 0$, the optimal price policy yields higher expected social welfare than does the optimal quantity policy. As uncertainty grows with ν, the expected social welfare associated with the optimal tax also grows and, at $\nu = 60$, is almost thrice

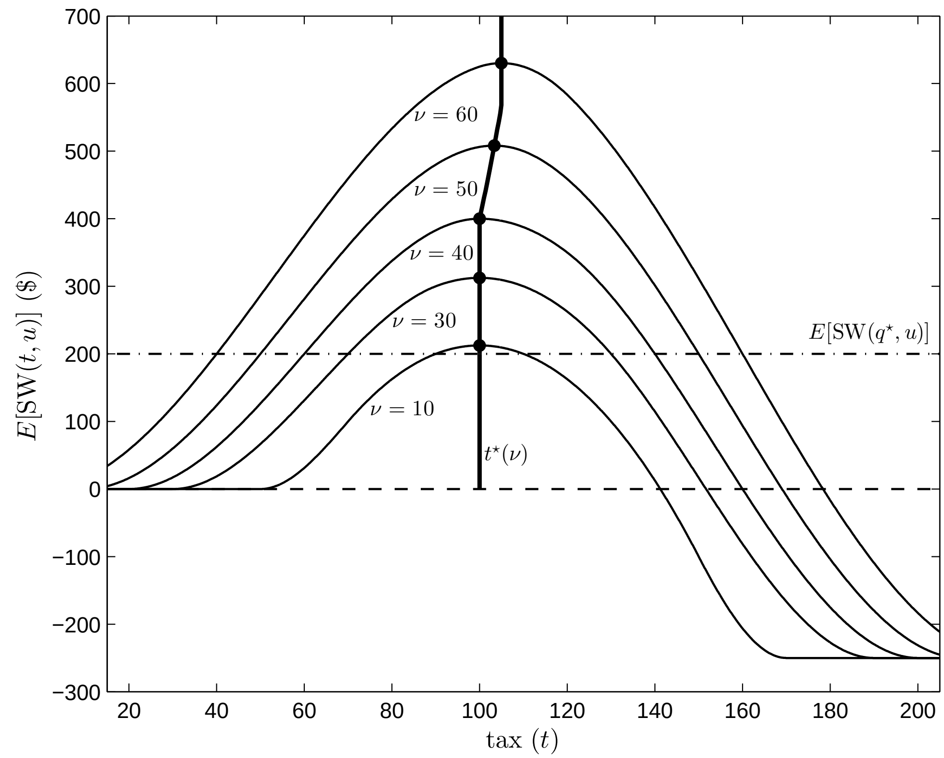


Figure 2.5: $E[\mathrm{SW}(t,u)]$ and $t^\star$ as uncertainty increases with $\alpha = 80$, $\beta = 1$, $\eta = 60 > \eta^\star$, and $\delta = 2$.

that at $\nu = 10$. Higher uncertainty, it turns out, is better.[14] Meanwhile, the expected social welfare of the optimal quantity policy remains at 200.

The vertical and piecewise-linear object labelled $t^*(\nu)$ in Figure 2.5 shows how the optimal tax changes continuously as ν grows. For $\nu \leq 40$ we find that $t^* = \hat{t}_4 = 100$, the price at which $\mathrm{MB}(a)$ and $E[\mathrm{MC}(a,u)]$ intersect. For $\nu \geq 55$, $t^* = \hat{t}_1$ is constant at 105. Between $\nu = 40$ and $\nu = 55$, the optimal tax, $\hat{t}_{2b}$, rises linearly in ν.[15] (See also

[14]Greater uncertainty can be an advantage in Weitzman (1974, p. 485) too: "The *ceteris paribus* effect of increasing σ^2 is to magnify the expected loss from employing the planning instrument with comparative disadvantage." In his model the expected social welfare under a quantity policy is unchanged in the face of increasing σ^2. Therefore, expected social welfare under a price policy must increase with σ^2. One might reasonably observe that increased uncertainty is unlikely to make society better off. The fact that it does so here, as in Weitzman, is a result of the assumptions that the polluting industry knows its cost function with certainty and the regulator knows the distribution of uncertainty exactly. If the industry were itself uncertain about its costs, as is more likely to be the case in practice, the effect of greater uncertainty may be quite different.

[15]In this example we have $\eta > \eta^*$. If $\eta < \eta^*$, for sufficiently high values of ν, Proposition 2 shows

Figure 2.9, where the relationship between t^* and ν is even more clear.)

2.6 Prices *vs.* quantities: Crossing from below

The situation examined in the previous section is not different in essence from the Weitzman framework with marginal benefits sloping downward but less steeply sloped than marginal costs. Even though our marginal benefits slope upward, Weitzman's basic insight is preserved: when marginal benefits are more nearly horizontal than marginal costs, a price policy is strictly preferred.

In this section, which concerns cases in which Assumption 2 is satisfied (that is, $\alpha < \eta$ and $\beta > \delta$), we enter unexplored terrain. Because the optimal quantity policy is now discrete, all or nothing, the comparison to a price policy becomes both more complicated and more important. Complicated because a new kind of corner solution and a new discontinuity appear, and important because the stakes involved in choosing the right policy become greater. Choosing maximum possible abatement when zero abatement is optimal, or *vice versa*, can lead to very large welfare losses. And choosing anything in the middle, which can occur when an intermediate tax is selected, risks the greatest losses of all. We will see that this risk is sometimes worth taking.

It turns out that another surprise awaits us. The problem, and especially the behavior of the optimal tax policy, depend crucially on the difference between the slope of marginal costs and the slope of marginal benefits. Our results are for the linear-uniform case to which we have restricted attention, but it appears that a similar result will go through for a more general setup. Here, the behavior depends upon whether β is less than or greater than twice δ. Much of the section is divided along these lines. The reader should bear in mind that, in the entire section, equations (2.6) through (2.18) are still in force.

Before turning to an examination of the two cases, we pause to establish an initial result that applies whenever marginal benefits cross from below, whether $\beta < 2\delta$ or $\beta \geq 2\delta$. Proposition 3 shows that a quantity policy can never outperform the optimal price policy when marginal benefits cross from below.[16]

that the optimal tax moves *below* the intersection of MB and E[MC].

[16]Note that we have not restricted realized marginal costs to remain positive, which they do not whenever $\eta < \nu$. In that case any negative tax, including importantly $t_{\min} = \eta - \nu < 0$, can be thought

Proposition 3. *For any permissible parameter vector* (θ, ν) *at which* $\beta > \delta$*, the expected social welfare resulting from the optimal price policy is not less than that associated with the optimal quantity policy:* $E[SW(t^*(\nu;\theta), u)] \geq E[SW(q^*(\nu;\theta), u)]$.

2.6.1 Similar slopes: $\delta < \beta < 2\delta$

We turn first to the case in which marginal benefits cross from below, but the slope is less than twice that of marginal costs: $\beta \in (\delta, 2\delta)$. Throughout, whenever Assumption 2 is satisfied and so $\beta > \delta$, from equation (2.18) we know that $\hat{t}_4$ is a local minimum when T_4 is feasible. We can ignore $\hat{t}_4$ in our search for the optimal tax.

The optimal tax rule, analogous to equation (2.23) is given by

$$t^*_{[\beta\in(\delta,2\delta)]}(\nu;\theta) = \begin{cases} t_{\min} & \text{if } \big[\eta \geq \eta^* \text{ and } \nu \leq E[\mathrm{MC}(0,u)] - \mathrm{MB}(0)\big] \\ \hat{t}_{2b} & \text{if } \big[\eta \geq \eta^* \text{ and } \nu \in \big(E[\mathrm{MC}(0,u)] - \mathrm{MB}(0), \nu_{1A}\big)\big] \\ \hat{t}_1 & \text{if } \big[\eta \geq \eta^* \text{ and } \nu \geq \nu_{1A}\big] \text{ or } \big[\eta \leq \eta^* \text{ and } \nu \geq \nu_{1B}\big] \\ \hat{t}_{3b} & \text{if } \big[\eta \leq \eta^* \text{ and } \nu \in \big(\mathrm{MB}(e^0) - E[\mathrm{MC}(e^0,u)], \nu_{1B}\big)\big] \\ \hat{t}_{\max} & \text{if } \big[\eta \leq \eta^* \text{ and } \nu \leq \mathrm{MB}(e^0) - E[\mathrm{MC}(e^0,u)]\big]. \end{cases} \tag{2.24}$$

All five of the possibilities from equations (2.15)–(2.17) are now relevant. If uncertainty is low, the tax should be set at $t_{\min}$ or at $t_{\max}$. These extreme tax levels mimic the optimal quantity policy by guaranteeing zero or maximum possible abatement respectively.

As uncertainty increases, the optimal tax moves away from the extremes and into the interior of its feasible range. And when ν exceeds the relevant threshold (either ν_{1A} or ν_{1B} depending on whether η is less than or greater than η^*), the optimal tax becomes $\hat{t}_1$ and remains there for further increases in ν.

The next two results, for $\eta > \eta^*$ (Proposition 4) and for $\eta < \eta^*$ (Proposition 5), provide parametric conditions under which the optimal tax is interior to $T = [t_{\min}, t_{\max}]$. They also establish that when this is true, the price policy is strictly preferred and is strictly increasing in ν.[17] In both of these propositions, notice that when the tax policy

of as a subsidy.

[17]The special case in which $\eta = \eta^*$, and so the planner is indifferent between zero and maximum possible abatement, is neglected here. This is not because it is uninteresting but because it leads to an additional layer of conditionality on the optimal tax that only clutters our notation further.

is preferred, the advantage increases as ν increases.

Proposition 4. *Consider a permissible parameter vector* (θ, ν) *with* $\beta \in (\delta, 2\delta)$ *and suppose that* $\eta > \eta^*$, *which means that* $q^*(\nu;\theta) = 0$.

(i.) *If* $\nu \le E[MC(0,u)] - MB(0)$, *then* $t^*(\nu;\theta) = t_{\min}$.

(ii.) *If* $\nu > E[MC(0,u)] - MB(0)$, *then* $t^*(\nu;\theta) \in (t_{\min}, t_{\max})$ *and, for* ν *sufficiently large,* $t^*(\nu;\theta) = \hat{t}_1$. *The decision rule* $t^*(\nu;\theta)$ *is a continuous function of* ν.

(iii.) *If* $\nu > E[MC(0,u)] - MB(0)$, *then* $E[SW(t^*(\nu;\theta), u)] - E[SW(q^*(\nu;\theta), u)]$ *is strictly positive and strictly increasing in* ν.

Proposition 5. *Consider a permissible parameter vector* (θ, ν) *with* $\beta \in (\delta, 2\delta)$ *and suppose that* $\eta < \eta^*$, *which means that* $q^*(\nu;\theta) = e^0$.

(i.) *If* $\nu \le MB(e^0) - E[MC(e^0,u)]$, *then* $t^*(\nu;\theta) = t_{\max}$.

(ii.) *If* $\nu > MB(e^0) - E[MC(e^0,u)]$, *then* $t^*(\nu;\theta) < t_{\max}$ *and, for* ν *sufficiently large,* $t^* = \hat{t}_1$. *The decision rule* $t^*(\nu)$ *is a continuous function of* ν.

(iii.) *If* $\nu > MB(e^0) - E[MC(e^0,u)]$, *then* $E[SW(t^*(\nu;\theta), u)] - E[SW(q^*(\nu,\theta), u)]$ *is strictly positive and strictly increasing in* ν.

Example 2. Once again we explain the unwieldy tax-setting rule through an example. Consider the following vector of parameter values:

$$\alpha = 60, \qquad \beta = 2, \qquad \eta = 70, \qquad \delta = 1.5, \qquad e^0 = 50.$$

The example is illustrated in Figure 2.6. In this case $\eta^* = 72.5$, which means that we have $\eta < \eta^*$ and so the optimal quantity policy is $q^* = e^0$, at which $E[\mathrm{SW}(q^*, u)] = 125$. It also means that the first two terms in (2.24), those for $t_{\min}$ and $\hat{t}_{2b}$, can be ignored. For any $\nu \le 15$, we can be sure that $t^* = t_{\max}$, which increases linearly in ν from 145 for $\nu = 0$ to 160 for $\nu = 15$. Above that value, t^* equals $\hat{t}_{3b}$ and so moves into the interior of the interval $[t_{\min}, t_{\max}]$, declining until $\nu = 40$. There, $t^* = 110$, where it remains for further increases in ν.

In Figure 2.6, the four curves represent expected social welfare as a function of t for $\nu = 10, 20, 30$, and 40. The many local maxima and minima are readily apparent. Look at the curve for $\nu = 10$, and note the way in which expected social welfare for an

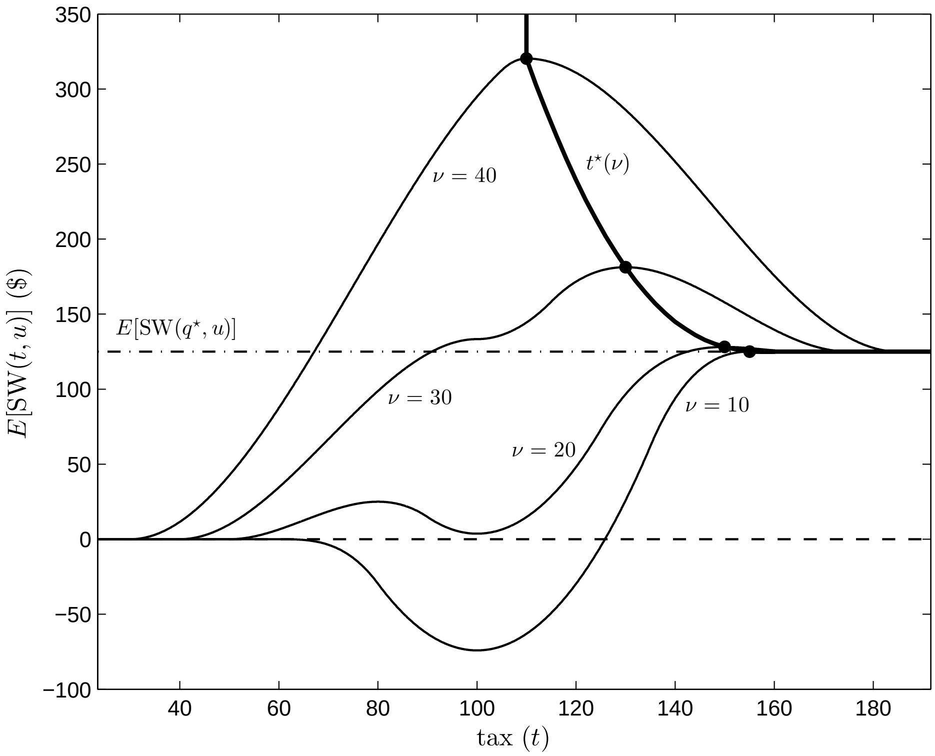


Figure 2.6: $E[\mathrm{SW}(t,u)]$ and $t^\star$ as uncertainty increases with $\alpha = 60$, $\beta = 2$, $\eta = 70 < \eta^\star$, and $\delta = 1.5$.

intermediate price is much lower than for the optimal quantity. Indeed, it goes negative there. To see why, note that at the optimal price any relization of u near zero means that the industry will choose an intermediate abatement level. This is the worst possible outcome: each unit of abatement costs more than the benefit it confers.

The relative gains to a price policy occur for extreme realizations, where $q^* = 0$ or $q^* = e^0$ can be very wrong, and in the example this cannot happen when uncertainty is low. Thus, for small ν the optimal tax mimics the optimal quantity of maximum possible abatement and so $E[\mathrm{SW}(q^*, u)] = E[\mathrm{SW}(t^*, u)]$. In this example with $\nu = 10$ the optimal price policy is $\hat{t}_{\max}$. On the $\nu = 20$ curve, though, the optimal tax is $t^* = \hat{t}_{3b} = 150$ and has already begun its gradual descent toward 110, where $t^* = \hat{t}_1$. For any $\nu > 15$ expected social welfare is higher for the tax than for the quantity policy, and expected social welfare associated with t^* increases monotonically. The optimal tax is piecewise linear, and changes continuously, in ν. The relationship appears as the bold curve in Figure 2.6; the curvature is generated by the $E[\mathrm{SW}(t,u)]$ function. (The piecewise linearity of $t^*(\nu)$ is apparent in Figure 2.9.)

2.6.2 Dissimilar slopes: $\beta \geq 2\delta$

Finally, consider the most unusual case, in which $\beta \geq 2\delta$. When marginal benefits are at least twice as steep as marginal costs, both $\hat{t}_{2b}$ and $\hat{t}_{3b}$ are local minima and so along with $\hat{t}_4$ play no role. The optimal tax is now driven discontinuously from an extreme of $t_{\min}$ or $t_{\max}$, which matches the optimal quantity, to an intermediate value at $\hat{t}_1$. Define

$$\nu^*_{\min} = \eta - \eta^* + \frac{e^0}{6}\sqrt{3\delta(2\beta-\delta)} \tag{2.25}$$

as the value of ν that equates $\Gamma_2(\hat{t}_{2a})$ and $\Gamma_1(\hat{t}_1)$. Similarly, define

$$\nu^*_{\max} = \eta^* - \eta + \frac{e^0}{6}\sqrt{3\delta(2\beta-\delta)} \tag{2.26}$$

as the value of ν that equates $\Gamma_3(\hat{t}_{3a})$ and $\Gamma_1(\hat{t}_1)$. These are threshold levels of uncertainty above which, for $\eta > \eta^*$ and $\eta < \eta^*$ respectively, the regulator is indifferent between an extreme tax and the intermediate tax $\hat{t}_1$. The optimal tax rule is

$$t^*_{[\beta\geq 2\delta]}(\nu;\theta) = \begin{cases} t_{\min} & \text{if } [\,\eta \geq \eta^* \text{ and } \nu \leq \nu^*_{\min}\,] \\ \hat{t}_1 & \text{if } [\,\eta \geq \eta^* \text{ and } \nu \geq \nu^*_{\min}\,] \text{ or } [\eta \leq \eta^* \text{ and } \nu \geq \nu^*_{\max}\,] \\ \hat{t}_{\max} & \text{if } [\,\eta \leq \eta^* \text{ and } \nu \leq \nu^*_{\max}\,]. \end{cases} \tag{2.27}$$

We will see that, for some parameter values, this rule is multi-valued and so is a correspondence rather than a function.

We have already shown, in Proposition 3, that the quantity policy can never outperform the price policy strictly when $\beta \geq 2\delta$. The two results of this section, for $\eta > \eta^*$ (Proposition 6) and for $\eta < \eta^*$ (Proposition 7), provide parametric conditions under which the optimal tax is interior to $T = [t_{\min}, t_{\max}]$. They also establish that when this is true, the price policy is strictly preferred and expected social welfare under the optimal tax increases in ν. At the threshold, both the optimal tax and the expected level of abatement are discontinuous in ν.

Proposition 6. *Consider a permissible parameter vector* (θ, ν) *with* $\beta \geq 2\delta$ *and suppose that* $\eta > \eta^*$*, which means that* $q^*(\nu; \theta) = 0$.

(i.) *If* $\nu < \nu^*_{\min}$*, then* $t^*(\nu; \theta) = t_{\min}$.

(ii.) *If* $\nu = \nu^*_{\min}$*, then* t^* *is multi-valued with* $t^*(\nu; \theta) \in \{t_{\min}, \hat{t}_1\}$*. The correspondence* $t^*(\nu; \theta)$ *is not continuous in* ν.

(iii.) *At the threshold value* $\nu^*_{\min}$*, the expected level of abatement changes discontinuously from* 0 *to* $E_u[a^*(t^*(\nu, \theta), u)] > 0$.

(iv.) *If* $\nu > \nu^*_{\min}$*, then* $E[SW(t^*(\nu, \theta), u)] - E[SW(q^*(\nu, \theta), u)]$ *is strictly positive and strictly increasing in* ν.

Proposition 7. *Consider a permissible parameter vector* (θ, ν) *with* $\beta \geq 2\delta$ *and suppose that* $\eta < \eta^*$*, which means that* $q^*(\nu; \theta) = e^0$.

(i.) *If* $\nu < \nu^*_{\max}$*, then* $t^*(\nu; \theta) = t_{\max}$.

(ii.) *If* $\nu = \nu^*_{\max}$*, then* t^* *is multi-valued with* $t^*(\nu; \theta) \in \{\hat{t}_1, t_{\max}\}$*. The correspondence* $t^*(\nu; \theta)$ *is not continuous in* ν.

(iii.) *At the threshold value* $\nu^*_{\max}$*, the expected level of abatement changes discontinuously from* e^0 *to* $E_u[a^*(t^*(\nu, \theta), u)] < e^0$.

(iv.) *If* $\nu > \nu^*_{\max}$*, then* $E[SW(t^*(\nu, \theta), u)] - E[SW(q^*(\nu, \theta), u)]$ *is strictly positive and strictly increasing in* ν.

Example 3. A final example shows how the optimal tax rule behaves when marginal benefits are at least twice as steep as marginal costs. Consider the following vector of parameter values:

$$\alpha = 0, \qquad \beta = 2, \qquad \eta = 35, \qquad \delta = 0.5, \qquad e^0 = 50.$$

The example is illustrated in Figures 2.7 and 2.8. In Figure 2.7, where $\nu = 19$, we see that T_1 ranges from 41 to 54. Also, $\eta < \eta^* = 37.5$, so that the optimal quantity policy is $q^*(\nu; \theta) = e^0$, where expected social welfare is 125. With $\nu = 19$, the optimal tax is at $t_{\max}$, where abatement is sure to be at e^0. Figure 2.7 is akin to Figure 2.4, except that here $T_4 = \emptyset$. For $\nu = 19$ the price $\hat{t}_1$ is a local, but not a global, maximum.

In Figure 2.8 we see how things change as ν increases from 16 to 19, to 21.594, and then to 26. The threshold value of $\nu^*_{\max} = 21.594$ is relevant for this case with $\eta < \eta^*$.

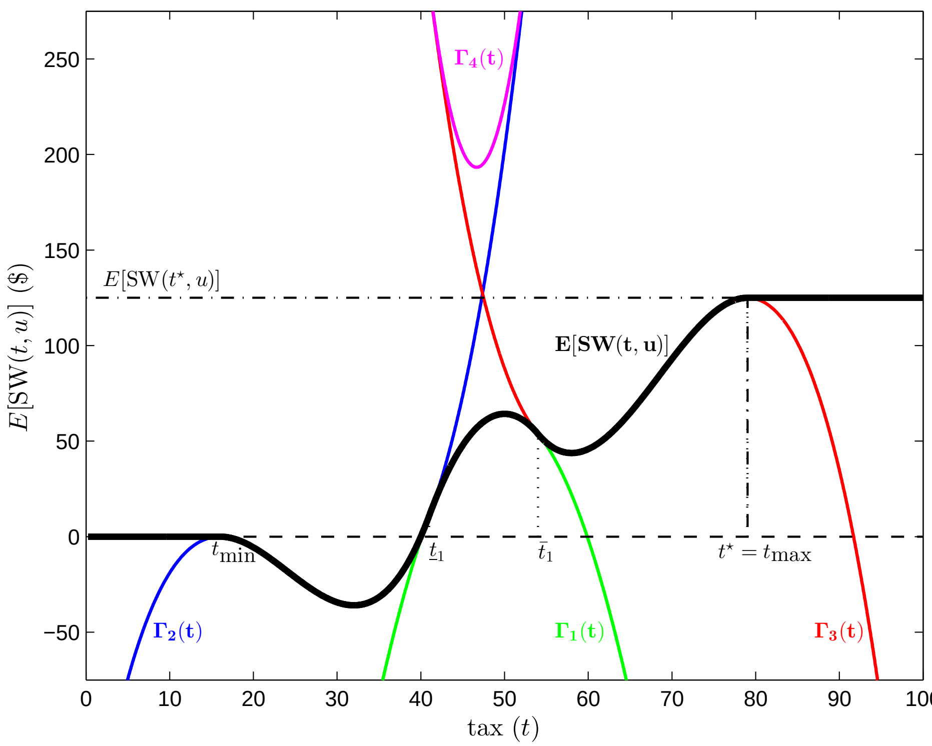


Figure 2.7: Outcomes from all possible taxes, and the four Γ_j functions with $\nu = 19$, $T_4 = \emptyset$, $\alpha = 0$, $\beta = 2$, $\eta = 35 < \eta^\star$, and $\delta = 0.5$.

At $\nu^*_{\max}$ the regulator is indifferent between choosing $t_{\max}$ or $\hat{t}_1$, here equal to 50. Thus, $t^*(\nu;\theta)$, represented by the two bolded segments, is multi-valued at $\nu^*_{\max}$. The optimal price policy, and the resulting expected level of abatement, are discontinuous at $\nu^*_{\max}$, but the level of expected social welfare is not. Figure 2.9 depicts the optimal price as a function of ν. There one can see the discontinuity and the multi-valuedness of $t^*(\nu;\beta)$ at $\nu^*_{\max}$. At $\nu = 26$ the optimal price is unique at $\hat{t}_1$ and the expected social welfare is higher.

2.7 Conclusions

When marginal benefits do not adhere to the basic rules of economic analysis, environmental policy becomes both more complicated and more interesting. When marginal

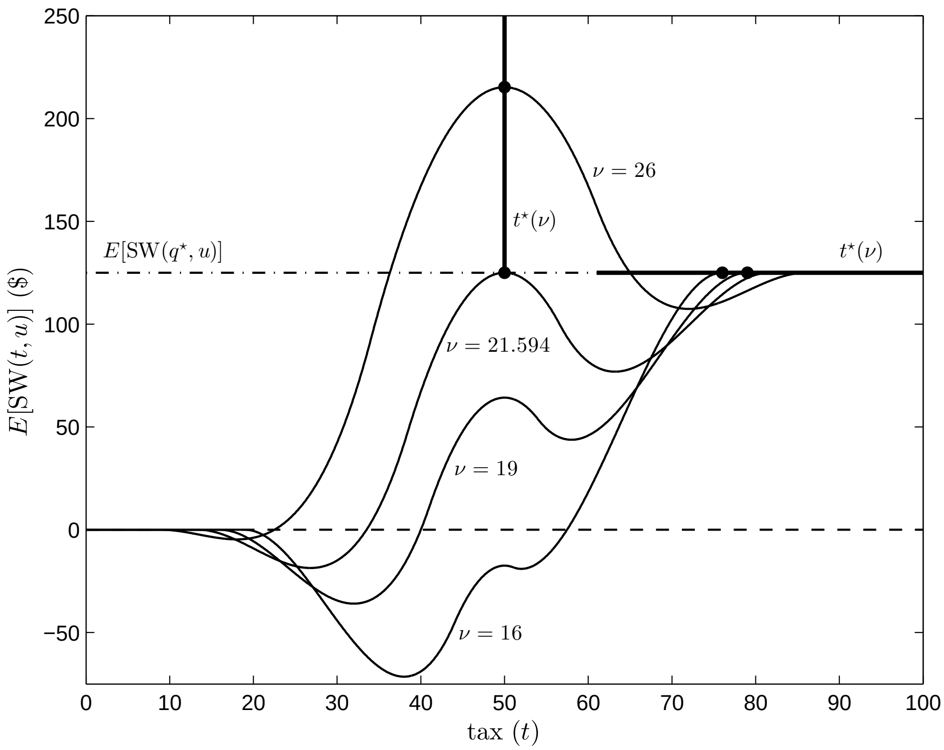


Figure 2.8: $E[\mathrm{SW}(t,u)]$ and $t^\star$ as uncertainty increases with $\alpha = 0$, $\beta = 2$, $\eta = 35 < \eta^\star$, and $\delta = 0.5$.

benefits slope upward, there is never an advantage to a quantity policy. Understanding the degree of uncertainty in marginal costs takes on a new and increased level of importance.

We contend that these findings are not merely mathematical curiosities. The scientific literature suggesting a strictly concave C-R is real, it is important, and it is familiar to scientists who study the relationship between pollution and public health. Economics has some catching up to do. And, for fine particulate matter especially, the stakes are high. Each year tens of thousands of Americans die prematurely due to exposure to ambient concentrations of particulates. Using standard estimates of the value of a statistical life, the value of foregone social welfare due to $\mathrm{PM}_{2.5}$ alone reaches into the tens of billions of dollars annually (U.S. EPA 2012b). By any measure, getting this right is important.

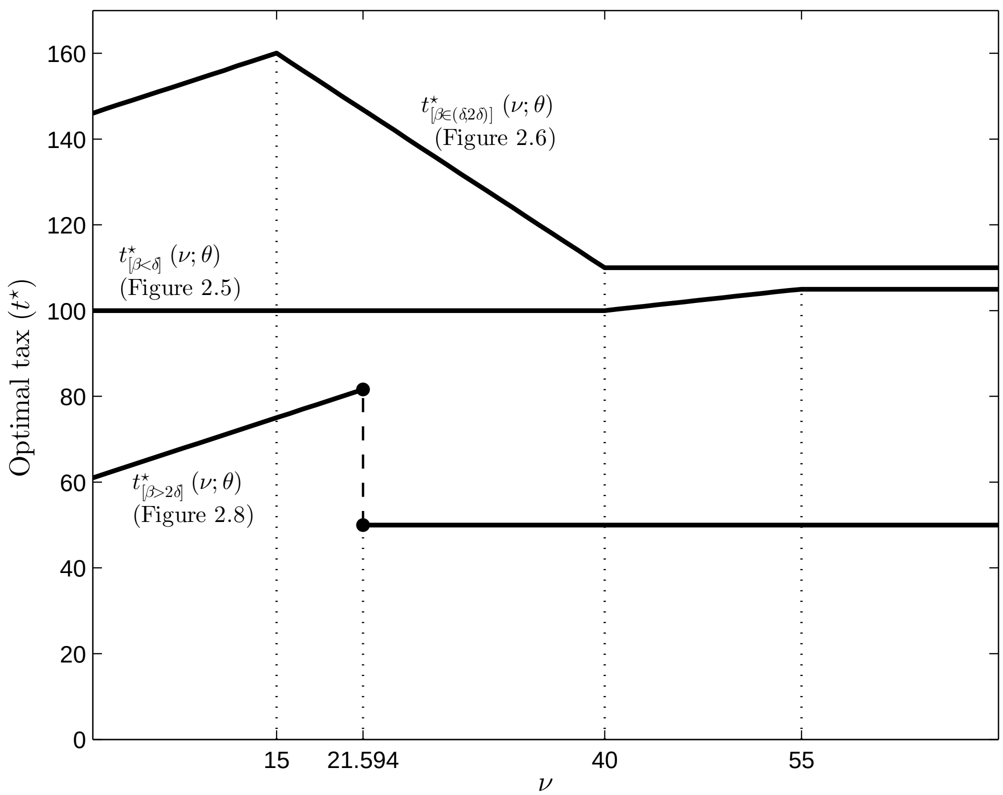


Figure 2.9: Optimal tax policy as a function of ν. Figure 2.5 is crossing from above, Figure 2.6 from below with $\beta < 2\delta$, and Figure 2.8 from below with $\beta > 2\delta$.

But one does not need to accept the recent scientific findings in order to see that anomalous situations, in which our analysis may be relevant, can be concocted using only ingredients from the mainline economics literature. Muller and Mendelsohn (2009) employ a linear damage function for fine particulates and thus assume that marginal benefits are constant. Their β is zero. Add to that Andreoni and Levinson (2001), who argue that the marginal cost of particulate abatement slopes downward. Their δ is negative. The combination yields marginal benefits crossing from below and, indeed, $\beta > 2\delta$.

Whether the findings of the paper prove to be important is, ultimately, an empirical question. But our insight that the level of uncertainty is more important than has been realized before seems likely to matter in the end. Without question, climate policy faces us with huge uncertainties.

Chapter 3

A Spatial Model of Air Pollution: the Impact of the Concentration-Response Function*

---

*Chapter 3 published as:

Goodkind, Andrew L., Jay S. Coggins and Julian D. Marshall. **2014**. A spatial model of air pollution: The impact of the concentration-response function. *Journal of the Association of Environmental and Resource Economists*, 1 (4), 451–479.

3.1 Introduction

The standard approach to analyzing clean-air policy assumes that the benefit associated with a unit of abatement declines as the air becomes cleaner.[1] That is, the greatest marginal improvements in human health are to be achieved in the dirtiest places. This assumption is based in a particular understanding of the relationship between pollution and human health: that the marginal harm to health grows ever more severe as the level of pollution rises. The traditional view is appealing from an ethical perspective because it means we should clean the dirtiest places first, thereby protecting those most who are most at risk. In line with this understanding, regulation of harmful pollutants has generally been based on uniform standards such as, in the U.S., the National Ambient Air Quality Standards (NAAQS). The NAAQS set limits on pollution concentration that are not to be exceeded in any location.

In this paper we ask, what if the basic understanding regarding the link between pollution exposure and health outcomes is wrong? This question, we argue, is both interesting and relevant to environmental policy. Recent estimates of the concentration-response (C-R) relationship between fine particulates (particulate matter with a diameter less than 2.5 microns, $PM_{2.5}$) and several causes of adult mortality suggest that the C-R function for $PM_{2.5}$ might be strictly concave in concentration. If true, this C-R function would mean that the first unit of abatement yields the smallest improvement in health risk, while the last unit, taking us to a pristine environment, yields the greatest improvement. Crouse *et al.* (2012), for example, find that the C-R relationship for $PM_{2.5}$ on ischemic heart disease is strictly concave ("supralinear") over ambient concentrations. Supralinear C-R functions across a wide range of $PM_{2.5}$ concentrations (including oberservations from exposure to ambient air, second-hand smoke and active smokers) were also found in Ostro (2004), and Pope *et al.* (2009a, 2011).

An interesting question, which economists are hardly equipped to answer, is which physiological pathways could lead to supralinearity. This matter is not well understood in the relevant health literature, but there are some tentative suggestions. Ambrose and Barua (2004, p. 1735) posit that the "underlying biochemical and cellular processes may become saturated with small doses of toxic components from cigarette smoke causing a

[1] Or, as in Muller and Mendelsohn (2009), that marginal benefits are constant.

nonlinear dose-response on cardiovascular function." Whatever the physiological explanation, supralinear C-R functions are common in studies of mortality from exposure to workplace toxins (Stayner *et al.* 2003). Among several explanations of the attenuation of the relative risk at higher concentrations, Stayner *et al.* suggest the possibility of a saturation of enzyme systems, where relative-risks increase faster before saturation is reached, and less thereafter. Birnbaum (2012) discusses the "low-dose hypothesis" which indicates that the impacts to human health from exposure to low doses of chemicals may be fundamentally different than what would be expected from the impact at higher doses.

Krewski *et al.* (2009), in their follow-up to the influential study by Pope *et al.* (1995), estimate the health risks associated with $PM_{2.5}$ exposure. They present the results of estimates based on two functional forms (see Figure 3.1). The first form is a log-linear C-R relationship (they call this relationship the "linear" version), which is convex and so leads to the usual form: a marginal benefit function that decreases in abatement. The second form is a log-log relationship (they call this relationship the "log" version), which is concave, or supralinear, and so leads to a marginal benefit function that increases in abatement. Regarding which of the two functions is to be preferred statistically, Krewski *et al.* (2009, p. 27) say only that "[T]he logarithmic function was a slightly better predictor of the variation in survival among MSAs than the linear function because the MSA random-effect variance is somewhat smaller (than that for the linear function) for each cause-of-death category except all other causes."

The Krewski *et al.* study highlights the uncertainty regarding the shape of the C-R function between $PM_{2.5}$ and adult mortality. The difference between the two forms of the C-R function has significant implications for air pollution policy of possibly the most consequential environmental issue impacting human health. Despite the enormous benefits from existing regulations, ambient concentrations of fine particulates remain a major cause of premature mortality in the U.S. The Office of Management and Budget estimates that the benefits of EPA regulations on fine particulate concentrations range from $19 billion to $167 billion per year (OMB 2013). The benefits are large because reduced exposure to fine particulates has saved many lives, yet at existing concentrations substantial risks remain. Fann (2012) estimates 130,000 annual cases of premature mortality attributable to fine particulate concentrations. In contrast, the comparative

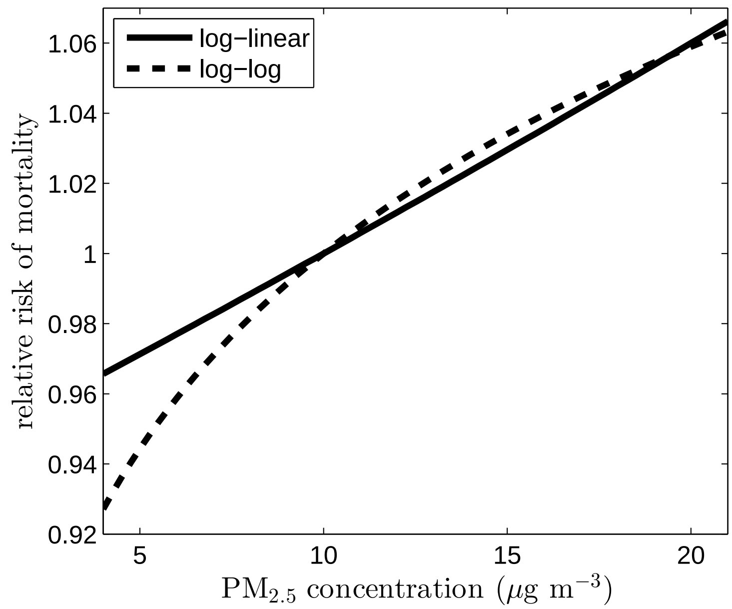


Figure 3.1: Risk of mortality from $PM_{2.5}$ concentration relative to risk at 10 μg m^{-3} for Krewski *et al.* (2009) log-linear and log-log concentration-response functions.

costs for cleanup are quite modest: by OMB's estimates, benefits are 2.6 to 22.9 times larger than costs (OMB 2013). Krewski *et al.* is central to the analysis of the impacts of exposure to fine particulates. The estimates in Krewski *et al.* are the latest from an American Cancer Society study, one of two major longitudinal analyses on the link between $PM_{2.5}$ and premature mortality. In the recent regulatory impact analysis by the U.S. EPA (2012b, page 5-32) recommending a lower standard for fine particulate concentrations, the key parameter was the Krewski *et al.* log-linear estimate.

The importance of the log-log C-R function is it calls into question the risks of mortality faced by individuals exposed to low levels of pollution, suggesting that mortality risks may be substantially lower at very low concentrations than at moderate or high concentrations. Why should one care about the risks borne by people exposed to the lowest concentrations, given that most experience higher exposures and correspondingly greater risks? Under the Clean Air Act, the primary goal of pollution control is to protect human health, and the general trend has been towards less pollution and lower risks.[2] In order to arrive at desirable policy outcomes, we must understand the risks at the lowest pollution level being considered. If the log-log C-R function is correct, then the benefits associated with achieving very low pollution levels are even higher than we have thought.[3]

A log-log C-R relationship suggests that regulators should focus not only on reducing risks for people at high concentrations, but also on reducing risks for people at low concentrations. Uniform pollution standards like the NAAQS, by their nature, have the effect of aiming abatement resources at those places where concentrations are the highest. As we show here, the uniform standard approach is not necessarily the most appealing from a social-welfare perspective if marginal benefits are increasing in abatement. An alternative policy, one that seeks to maximize the aggregate net benefits of abatement, might yield a very different outcome. A policy of maximizing social welfare, as opposed to limiting risks in the dirtiest locations, raises concerns over environmental justice as it may tend to exacerbate the disparity between the pollution faced by individuals leading to greater inequality in environmental risk.

We compare the implications for mortality risk of the two functional forms estimated by Krewski *et al.* (2009), and the socially preferred policies to regulate air pollution in both cases. We devise a simple model that captures the spatial aspects of air pollution over a region with many sources of pollution and many receptors. An efficient abatement policy is examined that controls pollution at each individual source and maximizes the social welfare of the individuals and industries in the region. The efficient policy is compared to a uniform standard, under which a cap is placed upon concentrations across the region, and a uniform tax, where a fee is levied on each unit of pollution emitted. The three policies are compared for both C-R functional forms based on total social welfare, the average concentration of pollution in receptors, and the level of environmental inequality across receptors.

We find that society should prefer significantly lower emissions, and correspondingly

---

[2] For instance, emissions of sulfur dioxide (one of the main contributors to fine particulate concentrations) in the U.S. has dropped from 23 million tons in 1990 to 5.5 million tons in 2012 (EPA 2013c). This drop has contributed to the 33% decrease in the U.S. national average concentration of $PM_{2.5}$ from 13.8 μg m^{-3} in 2000 to 9.3 μg m^{-3} in 2012 (EPA 2013b).

[3] In an analysis of the effects of lead exposure on children's IQ, Rothenberg and Rothenberg (2005), find that a model using the log of lead exposure is a significantly better fit of the data than a model with a linear lead relationship. The estimated benefits in the U.S. from the drop in lead concentrations to very low levels is 2.2 times greater with the log model than with linear.

lower ambient concentrations, of air pollution if the log-log C-R function is correct than if the log-linear C-R function is correct. Our findings underscore the importance of identifying the true shape of the C-R function between fine particulates and adult mortality.

With a log-log C-R function, we find that the efficient abatement policy performs substantially better than the uniform standard in maximizing social welfare and limiting the exposure to pollution. The efficient policy is also preferred to a uniform tax policy across the region, suggesting that there is substantial heterogeneity in marginal damages across sources. Surprisingly, the environmental inequality concerns with the efficient policy are slightly less than with the uniform standard. Pollution concentration reductions in the cleanest receptors provides benefits to all surrounding locations due to the widespread dispersion of the pollutant. In obtaining the greatest risk reductions in the cleanest locations, significant pollution concentration reductions are achieved across the map.

3.2 A multiple-receptor, multiple-source model

This paper presents a model that compares the outcomes from air pollution regulation policies using either a log-linear or log-log C-R relationship between fine particulate concentrations and adult mortality. The model simulates the dispersion of emissions from many sources (denoted by subscript j) located across a rectangular geographic region, and calculates the resulting change in pollution concentrations in all receptors (denoted by subscript i) in this region.

The region is separated into N identically sized grid squares. Each square can be both a source and receptor of pollution ($i, j = 1, \ldots, N$). In each grid square there is a population, Pop_i, and an aggregate mass emission rate, e_j, of $PM_{2.5}$, a primary conserved air pollutant. The model simulates emission and dispersion of primary $PM_{2.5}$, and the resulting concentration in each receptor.[4]

Pollution from each source is emitted at the center of the grid square and dispersed across the region as nonreactive emissions according to a Gaussian-Plume dispersion

[4] We consider only the impact of primary $PM_{2.5}$ on total $PM_{2.5}$ concentrations. Several other pollutants contribute to total $PM_{2.5}$ concentrations but these are excluded for simplicity.

model. The wind is uniform across the region and travels in many directions weighted by a representative wind rose. The pollution is assumed to be emitted from a point source with the same effective height for each grid square. The Gaussian-Plume dispersion model describes the change in average annual ground level concentration (in μg m^{-3}) at any other grid square resulting from the emission of an additional unit of pollution (tons/year) from a given grid square. Dispersion is calculated for all N grid squares to derive a source-receptor (S-R) matrix that describes the change in concentration at every grid square from the emission of a pollutant from every grid square. Let π_{ji} be the S-R coefficient from source grid square j to receptor grid square i, and let $\Pi_{N\times N}$ denote the S-R matrix for all grid squares.

The concentration of $PM_{2.5}$ is calculated for each grid square based on the emission rates from all squares, $\mathbf{e}$ (an $N \times 1$ vector), the S-R matrix, and a background concentration level, $\tilde{\mathrm{C}}$, that accounts for emissions upwind of the modeling domain. Background concentrations, which are constant across the region, are added to the pollution emitted inside the region. In the initial situation, prior to pollution regulation, the concentration in receptor square i is $\mathrm{C}_i(\mathbf{e}^0) = \tilde{\mathrm{C}} + \sum_{j=1}^{N} \pi_{ji} e_j^0$. After an abatement policy is adopted that induces abatement of $\mathbf{a}$ tons, the level of emissions are $\mathbf{e} = \mathbf{e}^0 - \mathbf{a}$, and the resulting concentration in receptor square i is $\mathrm{C}_i(\mathbf{e}) = \tilde{\mathrm{C}} + \sum_{j=1}^{N} \pi_{ji} e_j = \tilde{\mathrm{C}} + \sum_{j=1}^{N} \pi_{ji}(e_j^0 - a_j)$.

3.2.1 Krewski *et al.* concentration-response relationships

Concentration-response functions identify the relative risk of disease given exposure to concentrations of a stressor compared to some baseline. Suppose that the C-R relationship between exposure to fine particulates and adult mortality follows either the log-linear or the log-log functional form reported in Krewski *et al.* (2009), which are only two of many possible forms the relationship between $PM_{2.5}$ concentrations and risk of mortality can take. By focusing attention on the two forms in Krewski *et al.* we do not suggest that alternate forms are impossible. Rather we attempt to call attention to the divergent policy implications from these two functional forms, highlighting the importance of identifying the true shape of the C-R function.

We apply the log-linear and log-log C-R functions across the entire range of ambient $PM_{2.5}$ concentrations considered in this model. Because the data are sparse, Krewski *et al.* is silent on the shape of the relationship at even lower concentrations. It is possible that the function becomes convex in that region. There appears to be little doubt, though, that the function bends downward at very high concentrations (Smith and Peel 2010; Pope *et al.* 2009a).

Krewski *et al.* report the results of estimating two forms of a hazard function, denoted by $\lambda(\cdot)$, using a random-effects Cox proportional-hazard model. These hazard functions map a given level of concentration, and several covariates, onto the risk of developing a negative health outcome. Taking the ratio of a hazard function evaluated at two different concentration levels results in a hazard ratio (HR). The HR is reported in Table 11 of Krewski *et al.* For a certain location, or receptor, the HRs identify the relative risk of mortality between a relatively high concentration of pollution and a low concentration. The HR allows us to evaluate, for each receptor, the impact of a change in air pollution concentrations.

To understand the difference between the log-linear and log-log HRs it is necessary to examine the hazard functions (or log-hazard functions, as demonstrated in (3.1) and (3.2) below) that were estimated by Krewski *et al.* The log-linear specification, which we will often refer to as the "lin" form, is given by

$$\ln\left(\lambda^{\text{lin}}(\mathbf{X}, \text{PM}_{2.5})\right) = \ln\left(\hat{\lambda}\right) + \mathbf{X}\boldsymbol{\beta}^{\text{lin}} + \text{PM}_{2.5}\gamma^{\text{lin}}. \tag{3.1}$$

The log-log specification, which we will often refer to as the "log" form, is given by

$$\ln\left(\lambda^{\log}(\mathbf{X}, \text{PM}_{2.5})\right) = \ln\left(\hat{\lambda}\right) + \mathbf{X}\boldsymbol{\beta}^{\log} + \ln(\text{PM}_{2.5})\gamma^{\log}. \tag{3.2}$$

In (3.1) and (3.2), $\hat{\lambda}$ is the baseline risk of disease; $\mathbf{X}$ is a matrix of covariates that affect the risk of disease, with $\boldsymbol{\beta}$ the estimated effect of these variables; and $\text{PM}_{2.5}$ is the concentration of $\text{PM}_{2.5}$, with γ the estimated effect of $\text{PM}_{2.5}$ concentration. Our interest centers on γ. Notice that the log-linear specification (3.1) regresses the natural log of the risk, or hazard, on the "linear" concentration of fine particulates, whereas the log-log form (3.2) regresses the natural log of the risk on the natural log of the concentration.

The HR is defined as the ratio of the hazard function evaluated at two values of $\text{PM}_{2.5}$ concentration, $\text{PM}''_{2.5}$ and $\text{PM}'_{2.5}$. This equation is calculated by taking the ratio of the antilog of the log-hazard function at two $\text{PM}_{2.5}$ concentrations. The log-linear and log-log hazard ratios are given by

$$\mathrm{HR}^{\mathrm{lin}} = \frac{\lambda^{\mathrm{lin}}\big(\mathbf{X}, \mathrm{PM}''_{2.5}\big)}{\lambda^{\mathrm{lin}}\big(\mathbf{X}, \mathrm{PM}'_{2.5}\big)} = \frac{\hat{\lambda}\cdot\exp\big\{\mathbf{X}\boldsymbol{\beta}^{\mathrm{lin}} + \mathrm{PM}''_{2.5}\gamma^{\mathrm{lin}}\big\}}{\hat{\lambda}\cdot\exp\big\{\mathbf{X}\boldsymbol{\beta}^{\mathrm{lin}} + \mathrm{PM}'_{2.5}\gamma^{\mathrm{lin}}\big\}} = \exp\big\{\gamma^{\mathrm{lin}}\big(\mathrm{PM}''_{2.5} - \mathrm{PM}'_{2.5}\big)\big\} \quad \text{and} \tag{3.3}$$

$$\mathrm{HR}^{\mathrm{log}} = \frac{\lambda^{\mathrm{log}}\big(\mathbf{X}, \mathrm{PM}''_{2.5}\big)}{\lambda^{\mathrm{log}}\big(\mathbf{X}, \mathrm{PM}'_{2.5}\big)} = \frac{\hat{\lambda}\cdot\exp\big\{\mathbf{X}\boldsymbol{\beta}^{\mathrm{log}}\big\}\big(\mathrm{PM}''_{2.5}\big)^{\gamma^{\mathrm{log}}}}{\hat{\lambda}\cdot\exp\big\{\mathbf{X}\boldsymbol{\beta}^{\mathrm{log}}\big\}\big(\mathrm{PM}'_{2.5}\big)^{\gamma^{\mathrm{log}}}} = \left(\frac{\mathrm{PM}''_{2.5}}{\mathrm{PM}'_{2.5}}\right)^{\gamma^{\mathrm{log}}}. \tag{3.4}$$

With the HR, taking the ratio of the hazard function at two concentration levels causes all the variables in $\mathbf{X}$ to cancel out, leaving an expression comparing the risks of disease that depends only on the pollution concentration. Solving for γ in (3.3) and (3.4) yields

$$\gamma^{\mathrm{lin}} = \frac{\ln\big(\mathrm{HR}^{\mathrm{lin}}\big)}{\mathrm{PM}''_{2.5} - \mathrm{PM}'_{2.5}} \quad \text{and}$$
$$\gamma^{\mathrm{log}} = \frac{\ln\big(\mathrm{HR}^{\mathrm{log}}\big)}{\ln\big(\mathrm{PM}''_{2.5}\big) - \ln\big(\mathrm{PM}'_{2.5}\big)}.$$

The hazard ratios reported in Table 11 of Krewski *et al.* are based on a 10 μg m^{-3} difference in fine particulate concentration. Notice that for the log-linear form any 10 μg m^{-3} change will lead to the same value of the HR regardless of the baseline level of the concentration. With the log-log form, the value of the HR will change depending on the levels of the concentration. HR's reported in Table 11 of Krewski *et al.* are

$$\mathrm{HR}^{\mathrm{lin}} = 1.060 \text{ for any } 10\mu\text{g m}^{-3} \text{ change}$$
$$\mathrm{HR}^{\mathrm{log}} = \begin{cases} 1.095 \text{ for a } 10\mu\text{g m}^{-3} \text{ change from } 5\mu\text{g m}^{-3} \text{ to } 15\mu\text{g m}^{-3} \\ 1.059 \text{ for a } 10\mu\text{g m}^{-3} \text{ change from } 10\mu\text{g m}^{-3} \text{ to } 20\mu\text{g m}^{-3}. \end{cases}$$

The estimated values of γ, then, are

$$\gamma^{\mathrm{lin}} = \frac{\ln(1.060)}{10} = 0.005827 \quad \text{and} \tag{3.5}$$
$$\gamma^{\mathrm{log}} = \frac{\ln(1.059)}{\ln(20) - \ln(10)} = 0.082703. \tag{3.6}$$

Using these values of γ, we can construct a HR for any $\mathrm{PM}_{2.5}$ concentration compared to an initial baseline concentration. We will define the baseline concentration in receptor i as the concentration at the initial level of emissions, $\mathrm{PM}''_{2.5} = \mathrm{C}_i(\mathbf{e}^0)$. The concentration in receptor i for some other level of emissions, $\mathbf{e}$, is defined as $\mathrm{PM}'_{2.5} = \mathrm{C}_i(\mathbf{e})$. The HRs in receptor i become[5]

$$\mathrm{HR}_i^{\mathrm{lin}}\big(\mathrm{C}_i(\mathbf{e})\big) = \exp\big\{\gamma^{\mathrm{lin}}\big(\mathrm{C}_i(\mathbf{e}^0) - \mathrm{C}_i(\mathbf{e})\big)\big\} \quad \text{and} \tag{3.7}$$

$$\mathrm{HR}_i^{\mathrm{log}}\big(\mathrm{C}_i(\mathbf{e})\big) = \left(\frac{\mathrm{C}_i(\mathbf{e}^0)}{\mathrm{C}_i(\mathbf{e})}\right)^{\gamma^{\mathrm{log}}}. \tag{3.8}$$

3.2.2 Benefits of abatement

To calculate the benefits of pollution abatement in receptor i we first go back to the original definition of the HR in (3.3) and (3.4), the ratio of hazard functions, $\lambda_i(\cdot)$:

$$\mathrm{HR}_i\big(\mathrm{C}_i(\mathbf{e})\big) = \frac{\lambda_i^0}{\lambda_i\big(\mathrm{C}_i(\mathbf{e})\big)} \tag{3.9}$$

In (3.9) $\lambda_i^0 = \lambda_i\big(\mathrm{C}_i(\mathbf{e}^0)\big)$ is defined as the risk of disease given the initial concentration before regulation, and $\lambda_i\big(\mathrm{C}_i(\mathbf{e})\big)$ is defined as the risk given a lower concentration in receptor i after a reduction in emissions. Rearranging (3.9), the risk of mortality in receptor i for any level of emissions is $\lambda_i\big(\mathrm{C}_i(\mathbf{e})\big) = \lambda_i^0/\mathrm{HR}_i\big(\mathrm{C}_i(\mathbf{e})\big)$.

Assume that premature mortality is the only identified risk from the air pollution. We can estimate the expected number of deaths in receptor i as the receptor's population times the risk of mortality, $\mathrm{Deaths}_i\big(\mathrm{C}_i(\mathbf{e})\big) = \mathrm{Pop}_i \cdot \lambda_i^0/\mathrm{HR}_i\big(\mathrm{C}_i(\mathbf{e})\big)$. The change in deaths in receptor i after a reduction in emissions from $\mathbf{e}^0$ to $\mathbf{e}$ is $\Delta\mathrm{Deaths}_i\big(\mathrm{C}_i(\mathbf{e})\big) = \mathrm{Pop}_i \cdot \lambda_i^0\big[1 - 1/\mathrm{HR}_i\big(\mathrm{C}_i(\mathbf{e})\big)\big]$. Next we define the vector of emissions in all sources after regulation in terms of the vector of abatement from all sources, $\mathbf{e} = \mathbf{e}^0 - \mathbf{a}$. This way we can write the change in expected deaths and the concentration level after regulation as a function of a vector of abatement.[6]

Define $\mathcal{V}$ as the value society places on each human life. Therefore, the benefits in receptor i of a vector of abatement in all sources, are the changes in expected deaths,

[5] Note that the initial concentration is not an argument of the HR function because it is fixed under all abatement policies.

[6] We write $\mathrm{C}_i(\mathbf{e}) = \mathrm{C}_i(\mathbf{e}^0 - \mathbf{a})$ to indicate that the concentration in a receptor is a function of the level of abatement.

times the value of a life saved.

$$\begin{aligned}\mathrm{B}_i\big(\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})\big) &= \mathcal{V}\cdot\Delta\mathrm{Deaths}_i\big(\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})\big)\\ &= \mathcal{V}\cdot\mathrm{Pop}_i\cdot\lambda_i^0\left[1-\frac{1}{\mathrm{HR}_i\big(\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})\big)}\right].\end{aligned} \tag{3.10}$$

These benefits are the difference in monetized health damages without and with regulation, defined as a function of the hazard ratio. This implies that the benefits for receptor i from the two C-R functions in (3.7) and (3.8) are

$$\mathrm{B}_i^{\mathrm{lin}}\big(\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})\big) = \mathcal{V}\cdot\mathrm{Pop}_i\cdot\lambda_i^0\left[1-\exp\left\{-\gamma^{\mathrm{lin}}\left(\mathrm{C}_i(\mathbf{e}^{\mathbf{0}})-\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})\right)\right\}\right] \quad \text{and}$$

$$\mathrm{B}_i^{\mathrm{log}}\big(\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})\big) = \mathcal{V}\cdot\mathrm{Pop}_i\cdot\lambda_i^0\left[1-\left(\frac{\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})}{\mathrm{C}_i(\mathbf{e}^0)}\right)^{\gamma^{\log}}\right].$$

The total benefits for all receptors in the region, resulting from abatement, **a**, from all sources, are the sum of the benefits across receptors.

$$\mathrm{B}^{\mathrm{lin}}(\mathbf{a}) = \sum_{i=1}^{N}\mathrm{B}_i^{\mathrm{lin}}\big(\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})\big) \quad \text{and}$$

$$\mathrm{B}^{\mathrm{log}}(\mathbf{a}) = \sum_{i=1}^{N}\mathrm{B}_i^{\mathrm{log}}\big(\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})\big).$$

The total benefits are the value of risk reductions in all receptors resulting from a vector of abatement, **a**, at every source of pollution. Next we investigate how the benefits in all receptors change from an incremental change in abatement at any single source.

3.2.3 Interrelated marginal benefits across sources

With the total benefits in all receptors we can examine the marginal benefits of additional abatement from source j. We start with the marginal benefits in receptor i associated with a change in concentration in i (surpressing the argument of C_i):

$$\mathrm{MB}_i(\mathrm{C}_i) = \frac{\partial \mathrm{B}_i(\mathrm{C}_i)}{\partial \mathrm{C}_i} = \frac{\mathcal{V}\mathrm{Pop}_i\lambda_i^0}{[\mathrm{HR}_i(\mathrm{C}_i)]^2}\frac{\partial \mathrm{HR}_i(\mathrm{C}_i)}{\partial \mathrm{C}_i}.$$

The marginal benefits in receptor i attributable to a change in abatement at source j is just $\mathrm{MB}_i(\mathrm{C}_i)$ times π_{ji}, the incremental impact on concentrations in i from emissions at j. Summing across all receptors we obtain the combined marginal benefits of additional abatement from source j:

$$\mathrm{MB}_j(\mathbf{a}) = \sum_{i=1}^{N} \mathrm{MB}_i\big(\mathrm{C}_i(\mathbf{e}^0 - \mathbf{a})\big)\pi_{ji}.$$

The marginal benefits of abatement from source j are the sum of the change in benefits in all downwind receptors from an incremental increase in abatement at that source. For the log-linear and log-log C-R functions the expressions for the marginal benefits of abatement from source j are given by

$$\mathrm{MB}_j^{\mathrm{lin}}(\mathbf{a}) = \mathcal{V}\gamma^{\mathrm{lin}} \sum_{i=1}^{N} \frac{\mathrm{Pop}_i \lambda_i^0 \pi_{ji}}{\mathrm{HR}_i^{\mathrm{lin}}\big(\mathrm{C}_i(\mathbf{e}^0 - \mathbf{a})\big)} \quad \text{and}$$
$$\mathrm{MB}_j^{\mathrm{log}}(\mathbf{a}) = \mathcal{V}\gamma^{\mathrm{log}} \sum_{i=1}^{N} \frac{\mathrm{Pop}_i \lambda_i^0 \pi_{ji}}{\mathrm{C}_i(\mathbf{e}^0 - \mathbf{a}) \cdot \mathrm{HR}_i^{\mathrm{log}}\big(\mathrm{C}_i(\mathbf{e}^0 - \mathbf{a})\big)}.$$

The marginal benefits of abatement, or equivalently the marginal damages of emissions, are different for most or possibly all sources, as demonstrated in Muller and Mendelsohn (2009), and NRC (2010). Muller and Mendelsohn (2009) report median marginal damages across U.S. counties of \$1,170 (2000 U.S. Dollars) per ton of primary $PM_{2.5}$, with a range of \$41,000 between the 1$^{\mathrm{st}}$ and 99.9$^{\mathrm{th}}$ percentiles. In a report by the National Research Council (NRC 2010) median marginal damages across coal power plants are estimated at \$7,100 (2007 U.S. Dollars) per ton of primary $PM_{2.5}$, with a range of \$23,400 between the 5$^{\mathrm{th}}$ and 95$^{\mathrm{th}}$ percentiles.[7]

The heterogeneity of marginal benefits of abatement across sources is heavily influenced by the size of the population near the source.[8] The marginal benefits with the log-log C-R function (but not with the log-linear C-R function) will also be substantially influenced by the $PM_{2.5}$ concentration of the receptor grid squares near the source. A

[7] Muller and Mendelsohn (2009) and the NRC report (2010) use the log-linear C-R function from Pope *et al.* (2002). Krewski *et al.* (2009) is an update of Pope *et al.* (2002).

[8] Marginal benefits of abatement are closely linked with the relationship between emissions and human intake, known as the intake fraction [see Bennett *et al.* 2002]. Larger populations near sources are associated with a greater intake fraction.

source emitting pollution near receptors with low pollution concentrations will have higher marginal benefits of abatement than a source near high concentration receptors (given equal receptor populations) because of the concavity of the log-log C-R function. However, the lowest concentration receptors are also likely to have the lowest population density. If the log-log C-R function is correct, these two effects, higher population in dirtier places and greater risk reductions in cleaner places, pull in opposite directions. The overall effect might be to reduce the variance of the distribution of marginal benefits of abatement with log-log compared to log-linear.

The marginal benefits in source j are a function of the level of abatement from all sources of the pollution. The interconnected nature of the marginal benefit functions across sources turns out to be quite important. How do the marginal benefits of abatement from source j change when source k increases its abatement? The relevant effects are

$$\frac{\partial \mathrm{MB}_j^{\mathrm{lin}}(\mathbf{a})}{\partial a_k} = -\mathcal{V}(\gamma^{\mathrm{lin}})^2 \sum_{i=1}^{N} \frac{\mathrm{Pop}_i \lambda_i^0 \pi_{ji} \pi_{ki}}{\mathrm{HR}_i^{\mathrm{lin}}\big(\mathrm{C}_i(\mathbf{e}^0 - \mathbf{a})\big)} \quad \text{and} \tag{3.11}$$

$$\frac{\partial \mathrm{MB}_j^{\mathrm{log}}(\mathbf{a})}{\partial a_k} = -\mathcal{V}\gamma^{\mathrm{log}}(\gamma^{\mathrm{log}} - 1) \sum_{i=1}^{N} \frac{\mathrm{Pop}_i \lambda_i^0 \pi_{ji} \pi_{ki}}{\big[\mathrm{C}_i(\mathbf{e}^0 - \mathbf{a})\big]^2 \mathrm{HR}_i^{\mathrm{log}}\big(\mathrm{C}_i(\mathbf{e}^0 - \mathbf{a})\big)}. \tag{3.12}$$

for the log-linear and log-log C-R, respectively. Using the specific parameter values for γ found in (3.5) and (3.6) we can determine the sign of equations (3.11) and (3.12). Notice that the parameters are all positive, and most importantly for equation (3.12), $\gamma^{\mathrm{log}} < 1$. This means that, although equation (3.11) is negative, the $\big(\gamma^{\mathrm{log}} - 1\big)$ term guarantees that equation (3.12) is positive. When $k = j$ it is clear that the marginal benefit function for source j is decreasing in abatement a_j for log-linear, but increasing in a_j for log-log.

When $j \neq k$ with the log-linear C-R equation, the marginal benefits of abatement from source j are decreasing with abatement from source k. Therefore, abatement from different sources can be considered substitutes: additional abatement from one source decreases the marginal benefits of abatement from other sources. With the log-log C-R function, on the other hand, the marginal benefits of abatement from source j are increasing with abatement from source k. Thus, abatement levels across sources are complements.

Two questions arise from this differential feature of the two specifications. First, which of the two effects (substitute for log-linear or complement for log-log) is larger? And second, are either of these effects significantly different from zero? To answer the first question, consider the ratio of these two expressions:

$$\Lambda = \frac{\dfrac{\partial \mathrm{MB}_j^{\log}(\mathbf{a})}{\partial a_k}}{\dfrac{\partial \mathrm{MB}_j^{\mathrm{lin}}(\mathbf{a})}{\partial a_k}} = \frac{\gamma^{\log}\left(\gamma^{\log}-1\right)}{\left(\gamma^{\mathrm{lin}}\right)^2}\left[\frac{\displaystyle\sum_{i=1}^{N}\mathrm{Pop}_i\cdot\lambda_i^0\pi_{ji}\pi_{ki}\mathrm{C}_i(\mathbf{e}^0)^{-\gamma^{\log}}\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})^{\gamma^{\log}-2}}{\displaystyle\sum_{i=1}^{N}\mathrm{Pop}_i\cdot\lambda_i^0\pi_{ji}\pi_{ki}\cdot\exp\left\{-\gamma^{\mathrm{lin}}\left(\mathrm{C}_i(\mathbf{e}^0)-\mathrm{C}_i(\mathbf{e}^0-\mathbf{a})\right)\right\}}\right].$$

For the moment, assume that the concentration level is the same in each receptor i: $\mathrm{C}_i(\mathbf{e}^0) = \mathrm{C}(\mathbf{e}^0)$ and $\mathrm{C}_i(\mathbf{e}^0-\mathbf{a}) = \mathrm{C}(\mathbf{e}^0-\mathbf{a})$. (This assumption is not appropriate for the rest of our model, but it does help shed light on the characteristics of Λ.) This allows us to cancel out all the Pop, π and λ^0 terms, leaving

$$\Lambda = \frac{\gamma^{\log}\left(\gamma^{\log}-1\right)}{\left(\gamma^{\mathrm{lin}}\right)^2}\frac{\mathrm{C}(\mathbf{e}^0)^{-\gamma^{\log}}\mathrm{C}(\mathbf{e}^0-\mathbf{a})^{\gamma^{\log}-2}}{\exp\left\{-\gamma^{\mathrm{lin}}\left(\mathrm{C}(\mathbf{e}^0)-\mathrm{C}(\mathbf{e}^0-\mathbf{a})\right)\right\}}.$$

Plugging in the parameter values found in (3.5) and (3.6) and using concentration levels commonly found in the U.S., the absolute value of Λ can range from 10 to 60, with the largest values at low concentrations. This finding suggests that the complement effect in the log-log function is far more important than the substitution effect in the log-linear function, and at lower concentrations the difference is even more pronounced. A large complement effect suggests that for multiple sources that are interrelated (such that their pollution impacts common receptors), emission reductions from one source would increase the marginal benefits of abatement from the other sources. The incremental gains, in terms of reductions in risk, at the impacted receptors from additional abatement are larger after one source has reduced emissions. This outcome is embodied by the concave shape of the log-log C-R function in Figure 3.1, where the steepest part of the curve is found at the lowest concentration levels. Achieving low concentrations in receptors allows for the largest reductions in risk of mortality, per unit of concentration reduction.

The second question is whether the complement effect is large in absolute value, rather than just in relation to the substitution effect. This question is an empirical matter, but results given below indicate that the complement effect contributes to a preference for lower emissions and lower concentrations with the log-log C-R function than with log-linear.

3.2.4 Costs of abatement

The primary focus of this paper is the impact of the functional form of the C-R function on the benefits of pollution abatement. However, to provide an interesting analysis of pollution abatement policies it is necessary to specify the costs of abatement. The cost of abatement at each source is assumed to be independent of the other sources. The form of the marginal cost function (below) was chosen to have a relatively flat slope for the first units of abatement and a steeper slope as abatement increases, indicating that abatement becomes exceedingly expensive as a source attempts to completely eliminate their pollution. The marginal cost functional is strictly convex in a_j and is defined as

$$\mathrm{MC}_j(a_j) = \phi_{1j} - \phi_{2j} \cdot \ln\left(1 - \frac{a_j}{e_j^0}\right),$$

with $\phi_{1j} \geq 0$ and $\phi_{2j} > 0$. As seen in Figure 3.2 the marginal costs rise to infinity as abatement approaches the maximum ($\lim_{a_j \to e_j^0} \mathrm{MC}_j(a_j) = \infty$).

In the simulations, described in Section 3, marginal costs are heterogeneous across sources. This outcome is accomplished by randomly assigning values for parameters ϕ_{1j} and ϕ_{2j} for each source. The simulation assumes an initial situation with no pollution regulation, and then various policies to regulate pollution are introduced. The parameter values are calibrated to make the simulation economically interesting. That is, we ensure that marginal costs are small enough to induce abatement, and large enough to discourage nearly complete abatement. Although the cost parameters are randomly assigned to sources in the initial situation, *after* abatement policies have been implemented the largest marginal costs are found, on average, at sources that had the greatest amount of abatement.

The corresponding abatement cost function with fixed costs F is

$$\mathrm{Cost}_j(a_j) = (\phi_{1j} + \phi_{2j})\, a_j + \phi_{2j}\left(e_j^0 - a_j\right) \ln\left(1 - \frac{a_j}{e_j^0}\right) + F.$$

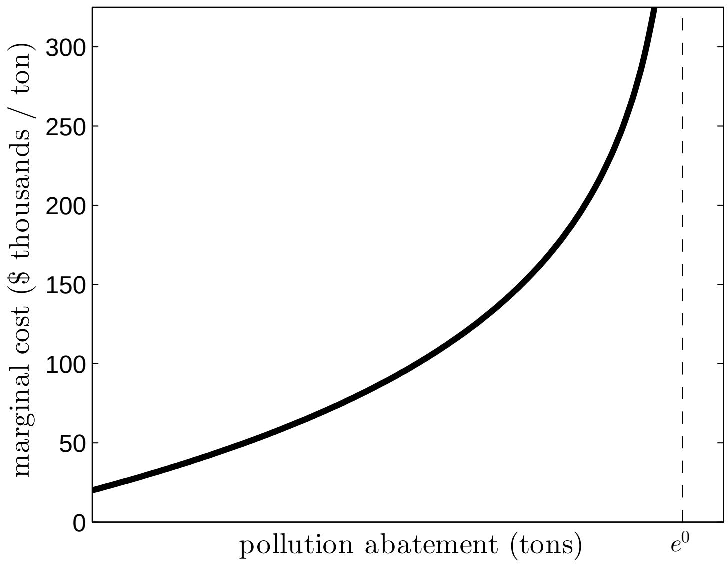


Figure 3.2: Illustration of marginal cost of abatement curve from zero abatement to complete abatement (e^0).

The total cost of a vector of abatement levels is the sum of the costs for individual sources:

$$\mathrm{Cost}(\mathbf{a}) = \sum_{j=1}^{N} \mathrm{Cost}_j(a_j).$$

3.2.5 Abatement policies

With the benefits and costs of pollution abatement established for each receptor and source in the region, we examine three approaches to abatement. The first approach (called "efficient abatement") selects abatement levels at each source to maximize the difference between the benefits and costs of abatement. The second approach is a uniform pollution concentration standard across the region, designed to emulate the NAAQS, achieved through a command-and-control approach. The third approach is a uniform tax on emissions. The uniform tax yields a cost-effective outcome, but ignores the spatial heterogeneity of marginal damages of emissions among sources. The difference in the outcomes between efficient abatement and the uniform tax identifies the importance of implementing source-specific regulation. All three approaches (policies) are considered with the log-linear and log-log C-R functions.

Efficient abatement can be represented for the log-linear C-R function, as

$$\max_{\mathbf{a}}\left\{\mathrm{B}^{\mathrm{lin}}(\mathbf{a})-\mathrm{Cost}(\mathbf{a})\right\}$$
$$\text{Subject to: } a_j \geq 0, a_j \leq e_j^0 \text{ for } j=1,\ldots,N \tag{3.13}$$

and, for the log-log C-R function, as

$$\max_{\mathbf{a}}\left\{\mathrm{B}^{\mathrm{log}}(\mathbf{a})-\mathrm{Cost}(\mathbf{a})\right\}$$
$$\text{Subject to: } a_j \geq 0, a_j \leq e_j^0 \text{ for } j=1,\ldots,N. \tag{3.14}$$

No source completely eliminates its pollution, because the slope of the marginal cost curve approaches infinity as abatement approaches e^0, but zero abatement is possible for some sources if the costs of abatement are high. The first-order conditions, then, are different depending on whether this corner comes into play:

$$\mathrm{MB}_j(\mathbf{a})=\begin{cases}\mathrm{MC}_j(a_j) & \text{if } a_j>0\\ \mathrm{MC}_j(0)-\mu_j & \text{if } a_j=0,\end{cases}$$

where the μ_j are the Lagrange multipliers on the zero abatement constraints. The N marginal benefit functions are interdependent, as abatement at each source impacts the marginal benefits at all other sources. As the number of sources becomes large, these problems pose computational challenges because of the number of equations that must be solved simultaneously.[9]

In the second policy, the uniform standard selects a concentration limit, $\bar{\mathrm{C}}$, that cannot be exceeded in any location. The optimal concentration limit is computed by setting a concentration limit, imposing emission reductions at sources to satisfy the limit, and calculating the benefits and costs of abatement. This approach is repeated for a series of limits at progressively stricter levels (lower concentrations). The uniform standard adopted (one standard for the whole region) is the concentration limit that

[9] With the log-log C-R function there is the possibility of a non-convexity because the marginal benefit and marginal cost curves are both increasing in abatement. In this model the potential for a non-convexity is very small because the absolute value of the slope of the marginal cost curve is likely greater than the absolute value of the slope of the marginal benefit curve in own-source abatement.

achieves the greatest net increase in welfare for society. This concentration is referred to as the "best" uniform standard for the region.

Under this command-and-control policy, in order to achieve any given concentration limit, emission reductions are determined based on that receptor, among all those to which the source contributes measurably, that has the highest concentration. To avoid constraining distant sources that have a minimal impact on an out-of-compliance receptor, here we defined a source as contributing pollution to a receptor if the relevant entry in the S-R matrix is above a lower threshold, ε.[10] We define θ_{ji} as an indicator variable that equals 1 if $\pi_{ji} \geq \varepsilon$ and zero if $\pi_{ji} < \varepsilon$. We employ the following approach for determining emission reductions: for each source j the emissions required to comply with the concentration limit are equal to the ratio of the proposed limit, $\bar{\mathrm{C}}$, and the maximum concentration in a receptor to which source j contributes pollution, denoted $\mathcal{C}_j$. If $\mathcal{C}_j$ is less than the proposed limit, emissions remain at the status quo, e_j^0. This maximum concentration that a source contributes pollution is defined as $\mathcal{C}_j = \max\{\mathrm{C}_1 \cdot \theta_{j1}, \ldots, \mathrm{C}_N \cdot \theta_{jN}\}$. The approach does not optimize abatement by minimizing the costs of emission reductions. Rather, each source reduces emissions by the proportion of its contribution to any downwind receptor that is out of compliance with the concentration limit.

For the log-linear functional form the problem is formally stated as

$$\begin{aligned} &\max_{\bar{\mathrm{C}}} \big\{ \mathrm{B}^{\mathrm{lin}}(\mathbf{a}) - \mathrm{Cost}(\mathbf{a}) \big\} \\ &\text{Subject to: } a_j = e_j^0 \cdot \max\left\{0, 1 - \frac{\bar{\mathrm{C}}}{\mathcal{C}_j}\right\} \text{ for } j = 1, \ldots, N \end{aligned} \tag{3.15}$$

and for the log-log functional form as

$$\begin{aligned} &\max_{\bar{\mathrm{C}}} \big\{ \mathrm{B}^{\mathrm{log}}(\mathbf{a}) - \mathrm{Cost}(\mathbf{a}) \big\} \\ &\text{Subject to: } a_j = e_j^0 \cdot \max\left\{0, 1 - \frac{\bar{\mathrm{C}}}{\mathcal{C}_j}\right\} \text{ for } j = 1, \ldots, N. \end{aligned} \tag{3.16}$$

The regional concentration standards (the NAAQS values calculated here) will depend

[10] $\varepsilon = 1.0 \times 10^{-6}$. For entries in the S-R matrix greater than this value of ε suggests that for each additional ton of emissions from a source, the annual concentration in the receptor is increased by more than 1 millionth of a μg m^{-3}.

on the C-R function because the benefits of abatement are determined by the functional form of the HR.

The third policy is a uniform tax where the regulator must choose a single tax rate on emissions to maximize the difference between total benefits of abatement and total costs of abatement. For each source the resulting quantity of abatement from any chosen tax policy is the greater of zero and the intersection of the tax and the source's marginal cost curve. For the log-linear C-R function the problem is stated as

$$\begin{aligned} &\max_t \big\{ \mathrm{B}^{\mathrm{lin}}(\mathbf{a}) - \mathrm{Cost}(\mathbf{a}) \big\} \\ &\text{Subject to: } a_j = \max \left\{ 0, \mathrm{MC}_j^{-1}(t) \right\} \text{ for } j = 1, \ldots, N \end{aligned} \tag{3.17}$$

and for the log-log functional form as

$$\begin{aligned} &\max_t \big\{ \mathrm{B}^{\mathrm{log}}(\mathbf{a}) - \mathrm{Cost}(\mathbf{a}) \big\} \\ &\text{Subject to: } a_j = \max \left\{ 0, \mathrm{MC}_j^{-1}(t) \right\} \text{ for } j = 1, \ldots, N. \end{aligned} \tag{3.18}$$

3.3 Model solution and results

The model analyzes a hypothetical geographical region that is 750 km (East/West) and 500 km North/South (an area approximately 5% the size of the contiguous U.S.). The region is separated into 25 km $\times$ 25 km grid squares, with a total of $N = 600$ grid squares that are each a source and receptor of air pollution. The populations of the grid squares in the region are modeled after a section of the U.S. Midwest that spans from Northwest West Virginia to Southeast Wisconsin.[11] The emissions from each grid square prior to regulation are artificially determined but are correlated with that grid square's population, with a correlation coefficient $\rho = 0.44$.[12] The model and this example are meant to be representative of a generic situation of air pollution and how abatement policies will affect the welfare of the region. Mapping the actual population

[11] The parameters of the dispersion model and the emissions from sources are artificial and not calibrated to the modeled geographic region. Modeled concentrations are not meant to mirror observed concentrations in this region.

[12] The correlation coefficient is derived from the correlation between criteria pollutant emissions and population from counties in Illinois, Indiana, Michigan, and Ohio.

of a region of the Midwest, and correlating the emissions with population is done to provide a reasonable reflection of a real-world situation.[13]

3.3.1 Model solution

The model assumes an initial situation without pollution regulation. The model then solves for three abatement policies: efficient abatement, a uniform pollution concentration standard across the region, and a uniform emissions tax. Each policy is analyzed with both of the Krewski *et al.* (2009) C-R functions: log-linear and log-log.

The model is run 1,000 times with different abatement cost parameter values and distributions of initial emissions at the sources, randomly selected, to provide a wide array of possible outcomes. The total quantity of emissions, prior to regulation, across the region is fixed in each iteration, but the allocation of initial emissions to each grid square is randomly assigned, with the approximate correlation between emissions and population maintained. The marginal cost parameters for each grid square source are randomly drawn from distributions in each model run.[14]

In the efficient abatement policy, which solves the maximization problems in (3.13) and (3.14), emissions are selected to maximize the difference between the benefits to society of reduced mortality and the costs of abatement for polluters. Solving this problem presents significant computational challenges, requiring as it does the simultaneous solution of $N = 600$ equations and 600 unknowns. We adopted an iterative numerical approach that yields the optimum in a computationally efficient manner.[15] The iterative

[13] Our input parameters for the Gaussian-Plume dispersion model include a constant ground level wind speed of 5.24 m/sec, based on the average annual wind speed in Minnesota at a height of 10 m., with a West to East prevailing direction. The emissions from each grid square are assumed to be emitted from a point source with an effective height of 250 m. The background concentration level of $PM_{2.5}$ is $\tilde{\mathrm{C}} = 4\mu\mathrm{g\ m}^{-3}$. Following U.S. Environmental Protection Agency (EPA) recommendations (EPA 2010) regarding the value of a statistical life (VSL), the parameter $\mathcal{V}$ is equal to \$8.43 million (2012 U.S. dollars). The baseline risk of mortality, λ_i^0, is assumed to be the same in each receptor, and is set to the 2011 national mortality rate of 806.6 deaths per 100,000 population (Hoyert and Xu 2012).

[14] The parameters ϕ_1 and ϕ_2 for each source are both independently drawn from normal distributions: $\phi_1 \sim \mathcal{N}(20{,}000, 8{,}000)$; $\phi_2 \sim \mathcal{N}(100{,}000, 50{,}000)$. Draws from the ϕ_1 distribution that are negative values are assigned a value of zero, and draws from the ϕ_2 distribution that are non-positive are assigned a value of 10. Fixed costs are assumed to be zero for all sources, $F = 0$.

[15] We confirmed this claim by solving an otherwise identical model, but with 150 grid squares instead of 600. At $N = 150$ it is just possible to compute the solution directly using a personal computer. We compared the efficient iterative solution to the fully simultaneous solution for a sample of 20 randomly selected runs. The numerical error between the two methods was very small: 0.002% with the log-linear C-R function and 0.001% with log-log.

method is based upon the algorithm suggested by Antweiler (2012), who envisions an environmental regulator who selects a set of source-specific discriminating taxes in each period, adjusting the taxes in response to the observed emissions decisions by sources. In each period (which is best interpreted as an iteration in our algorithm), the taxes are set equal to each source's marginal benefits of abatement, *computed at the previous period's abatement level*. Then the abatement level for the next iteration is determined by equating marginal benefits and marginal costs for each source (taking into account possible corner solutions). Because marginal benefits are based on abatement levels in the previous iteration, the simultaneous equation problem is avoided. In the first step, the abatement levels are not optimal, but after several iterations the solution converges to the optimum found by solving the equations simultaneously. With the large number of sources in the model, and the 1,000 model runs with different parameter values, the iterative solution method is computationally efficient and very accurate.

3.3.2 Model results

Across the 1,000 runs of the model, substantial differences appear between the outcomes from the log-linear and log-log C-R functions. With an efficient abatement policy, if the true C-R function is log-log society should prefer fewer emissions, lower average concentrations of fine particulates, and therefore, lower risks of mortality, than if the true C-R function is log-linear. This finding is attributable to the comparatively large reductions in risk of mortality that are possible from obtaining low concentrations of fine particulates with a log-log C-R function. This result highlights the importance of discovering the true shape of the C-R function between adult mortality and fine particulate exposure as the preferred abatement policies and outcomes are substantially different.

Efficient policy and uniform pollution standard

The results show that when abatement costs are sufficiently large an efficient abatement policy is usually preferred to a uniform pollution standard for fine particulate concentrations. This preference exists for both C-R functions, and suggests that policies directed at obtaining the greatest risk reductions at the lowest cost may provide enough benefits to outweigh the environmental justice concerns of not primarily focusing emission reductions at the dirtiest locations.

The following results report the outcomes from the median of the 1,000 model runs. We analyze and compare the outcomes of the model from two of the policies examined: efficient abatement policy with log-linear and log-log C-R functions; and, the "best" uniform pollution standard with log-linear and log-log C-R functions. Prior to regulation, there are 3.22×10^6 tons of emissions across the region. With the efficient abatement policy, across the 1,000 runs, the median emission reduction is 22% and 38% with log-linear and log-log C-R functions, respectively. With the uniform standards, emissions are reduced by 18% (log-linear C-R) and 30% (log-log). For both policies, total emissions and population-weighted average concentrations are lower for the log-log than for the log-linear C-R function. The largest population-weighted average concentration reduction is achieved by the efficient abatement policy with the log-log C-R function. From an initial population-weighted average concentration of 12.2 μg m^{-3}, the efficient abatement policy leads to a population-weighted average concentration of 10.0 μg m^{-3}, for log-linear, and 8.6 μg m^{-3} for log-log. The "best" uniform standards yield population-weighted average concentrations of 10.8 μg m^{-3} (log-linear) and 9.7 μg m^{-3} (log-log).

With the efficient abatement policy, concentration reductions (2.2 μg m^{-3} (log-linear) and 3.6 μg m^{-3} (log-log)) would reduce annual expected deaths across the region by 4,000 with log-linear versus 8,950 for log-log. The lower expected mortality with log-log is attributable to both the lower average concentrations and the comparatively smaller risks at low concentrations with this C-R function. The amount of abatement (and therefore in our model the total costs of abatement) is higher for log-log than for log-linear, because greater abatement is justified by the greater reductions in risk of mortality. Under the "best" uniform standards the reduction in expected mortality is 2,700 with log-linear and 5,900 with log-log.

With a log-log C-R function the total emissions from the efficient abatement policy are only 11% less than with the "best" uniform standard; however, the efficient policy is able to achieve a 52% greater reduction in expected mortality across the region compared with the uniform standard. An efficient abatement policy is able to more precisely target abatement to reduce risks of mortality.

The effectiveness of the efficient abatement policies is demonstrated by a comparison

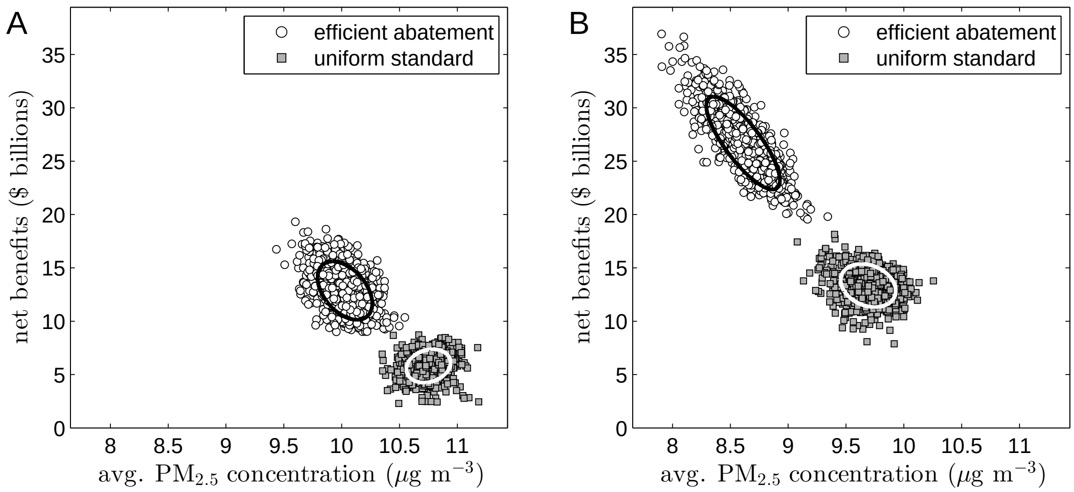


Figure 3.3: Efficient abatement policy versus uniform standard: net benefits of abatement and population-weighted average pollution concentration for each model run, (**A**) log-linear C-R function, (**B**) log-log C-R function. The four clusters each contain 1,000 points, of which 69% to 71% are encircled by the respective ellipse. Each ellipse radius represents one standard deviation from the mean. The points inside each ellipse are the outcomes nearest the center of the joint distribution of $PM_{2.5}$ concentration and net benefits.

of the net benefits of abatement between the two policies and the two C-R functions, which combines the benefits from risk reductions with the costs of abatement for the polluting sources. Figure 3.3, which displays both the net benefits of abatement and the population-weighted average concentration across the region for all model runs, shows that the efficient abatement policies are able to achieve far greater welfare for society than a uniform standard while also reducing the average concentration to lower levels. The ellipse inside each cluster of points in Figure 3.3 encircles the outcomes within one standard deviation from the mean, in each dimension, of the 1,000 model runs. In the median model run, net benefits with a log-linear C-R function are $13 billion for the efficient abatement policy and $5.8 billion for the "best" uniform standard. For the log-log C-R function, net benefits are $26 billion (efficient policy) and $13 billion (uniform standard). Under either policy, the net benefits of abatement are much larger if the C-R function is log-log compared to log-linear. If the true C-R function is identified as log-log, regulators could justify imposing a more restrictive pollution control policy.

Across the 1,000 model runs the distribution of the outcomes is tightly centered

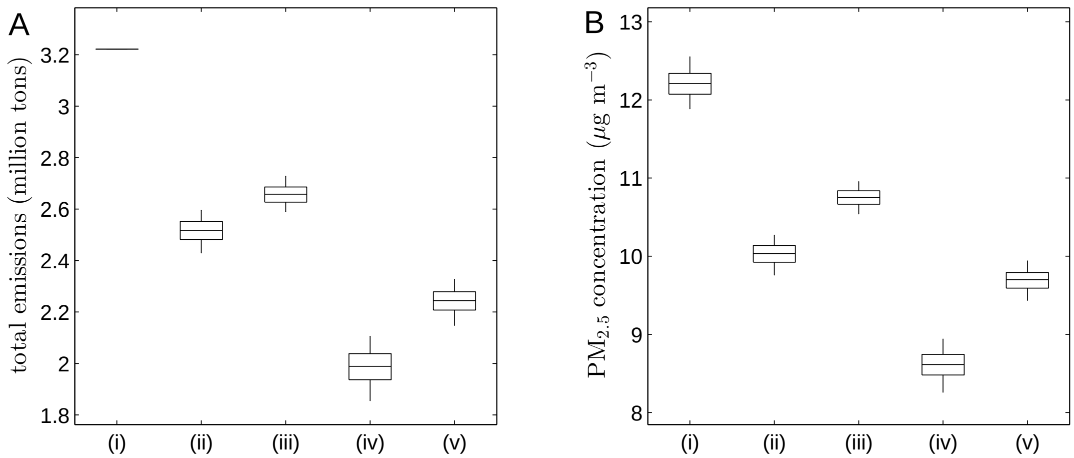


Figure 3.4: Distribution of outcomes across 1,000 model runs, (**A**) total emissions across region, (**B**) population-weighted average $PM_{2.5}$ concentration across region: (i) initial situation without regulation, (ii) log-linear efficient abatement policy, (iii) log-linear "best" uniform standard, (iv) log-log efficient abatement policy, (v) log-log "best" uniform standard. Boxplot: center line represents median, top and bottom of box represents 25^{th} and 75^{th} percentiles, and end of whiskers represent 5^{th} and 95^{th} percentiles. The total emissions boxplot appears as a single line under the initial situation without regulation because it is a constant value in all model runs.

around the median, indicating that the median results presented above provide a good representation of the array of outcomes that are possible in the model. This outcome suggests that across many profiles of initial emissions and abatement costs from sources, the general pattern holds that society prefers lower emissions with a log-log than with a log-linear C-R function. This conclusion is demonstrated in Figure 3.4 for total emissions of pollution and population-weighted average concentrations of $PM_{2.5}$ across the region. The distributions of the outcomes across the model runs from the four policies show clearly that emissions and concentrations are lower with the log-log C-R function compared with the log-linear function. There is also a less obvious, but clear distinction between the efficient policies and the uniform standards, with lower emissions and concentrations under the efficient policies.

Differences also exist between the maximum allowable concentrations from the "best" uniform standard under the two C-R functional forms. In Figure 3.5, the median standard with the log-linear C-R function is set at 13.25 μg m^{-3}, while the median standard with log-log is more stringent, at 11.75 μg m^{-3}. The primary cause of the lower

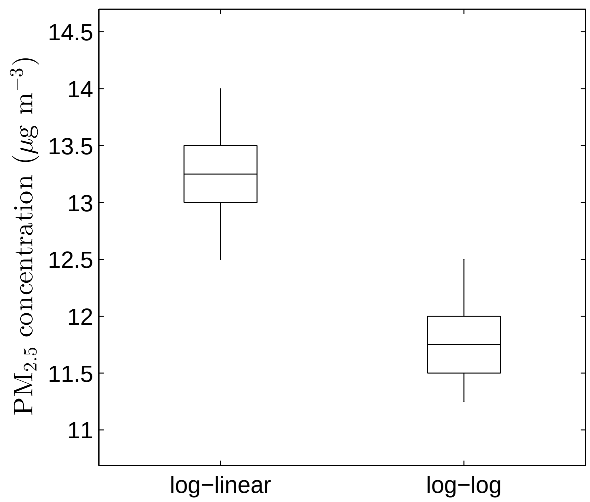


Figure 3.5: Distribution of maximum allowable PM$_{2.5}$ concentration of "best" uniform standard with log-linear and log-log C-R functions across 1,000 model runs. Boxplot: center line represents median, top and bottom of box represents 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles, and end of whiskers represent 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles.

standard with the log-log C-R function is the greater ancillary benefits that accrue to low-concentration receptors when limiting emissions to meet the standard at the dirtiest receptors. While the standard is designed to limit the pollution in the dirtiest locations, the resulting emission reductions also cleans the air in surrounding (i.e., comparatively cleaner) areas, leading to risk reductions at those cleaner areas. Because of the steep slope of the log-log C-R function at low concentrations, large risk reductions are also achieved in nearby clean receptors. The larger benefits (with log-log compared to log-linear) at the comparatively cleaner receptors offset the greater costs of setting a lower, stricter uniform concentration standard across the region.

In the U.S., uniform pollution standards under the NAAQS are used to limit the risks associated with criteria air pollutants. One apparent virtue of these policies is the perceived equity of protecting everyone from the greatest risks from pollution (i.e., the standard is the same everywhere). Under an efficient abatement policy, environmental justice concerns may arise. The focus is on making the greatest risk reductions at the lowest cost, regardless of equity among receptors. Of course, even with uniform standards inequality still exists (Brulle and Pellow 2006; Mohai, Pellow and Roberts

2009; Marshall 2008; Su *et al.* 2009; Clark, Millet and Marshall 2014). In previous studies (Marshall, Zwor and Nguyen 2014), it was found that more than 90% of articles on environmental justice in the United States reported that air pollution exposures are greater for lower- than for higher- socioeconomic status groups (e.g., based on income, race, education level, or other attributes). Dockery *et al.* (1993) reported that there is no threshold of fine particulate concentration below which risks of mortality are nonexistent (see also Pope and Dockery 2006). Receptors with concentrations below the uniform standard face less risk than those receptors that just meet the standard. The question becomes how much inequality is acceptable?

An efficient policy abates sources that would lead to the greatest risk reductions, but surrounding receptors also experience concentration reductions because of pollutant dispersion. Disregarding equity or justice under an efficient abatement policy may initially appear objectionable, but the policy should be evaluated based on a comparison of the realized level of inequality or injustice against the overall welfare gains. Figure 3.3 demonstrates the substantial advantage of the efficient abatement policies over the uniform standards both with greater net benefits to society and lower average fine particulate concentrations. Surprisingly, the level of inequality, as measured by the Gini coefficient of the differences in fine particulate concentrations across the region, is slightly lower with the efficient policy. With a log-linear C-R relationship the median Gini coefficient across the model runs is 0.077 under the efficient policy and 0.090 under the uniform standard. With log-log the Gini coefficient is 0.060 under the efficient policy and 0.066 under the uniform standard. These differences are small and show significant improvement when compared to the situation prior to regulation with a median Gini coefficient of 0.139. The distributions across the model runs of the Gini coefficient, in Figure 3.6, show considerable overlap, suggesting that the environmental inequality issues with the efficient policies, in cases considered here, are of no greater concern than with the uniform standards. Using the Atkinson coefficient, an alternative measure of inequality, the same pattern is found with inequality slightly lower under the efficient abatement policies than under the uniform standards.

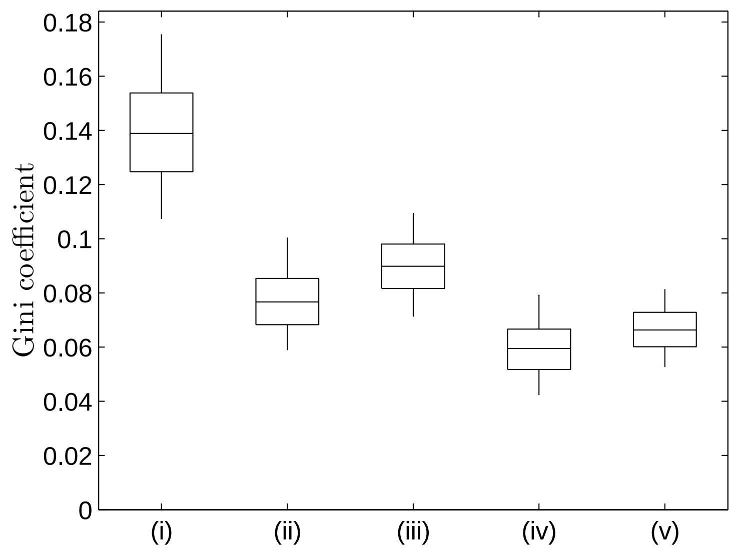


Figure 3.6: Distribution of Gini coefficient of inequality based on $PM_{2.5}$ concentration across 1,000 model runs: (i) initial situation without regulation, (ii) log-linear efficient abatement policy, (iii) log-linear "best" uniform standard, (iv) log-log efficient abatement policy, (v) log-log "best" uniform standard. Boxplot: center line represents median, top and bottom of box represents 25^{th} and 75^{th} percentiles, and end of whiskers represent 5^{th} and 95^{th} percentiles.

Efficient policy and uniform tax on emissions

The above results indicate a strong social preference for an efficient abatement policy over a command-and-control style uniform pollution standard. Next we compare the importance of source-specific emission controls in the efficient policy to a cost-effective policy that does not differentiate the impact of emissions by source. Henry, Muller and Mendelsohn (2011) find that SO_2 allowance trading in the U.S. actually increases damages relative to the no-trade baseline because it directs greater emissions to the dirtiest cities. The dirtiest cities tend to have the highest marginal costs of abatement as well as the highest marginal damages from emissions. By not differentiating between emissions at different sources, more emissions result in the areas with the highest marginal damages than is optimal.

Here we compare a uniform emissions tax to an efficient abatement policy that equates the marginal damages to the marginal costs of abatement for each source. The uniform emissions tax achieves a cost-effective outcome because sources choose a quantity of emissions to equate their marginal costs of abatement with the tax on emissions; however, because there is a single tax rate for the region, most or possibly all sources will be charged a tax that is not equal to the marginal damages from emissions. The uniform tax is a much simplier policy to administer than an efficient policy. We ask whether the distribution of marginal damages across sources is sufficiently spread out to favor an efficient policy over a uniform tax. Are the distributions of marginal damages sufficiently different between a log-linear and log-log C-R function to warrant different policies depending on the identified functional form?

The simulation results show that marginal damages vary greatly by source. Under the efficient abatement policy with both the log-linear and log-log C-R functional forms, marginal damages at the 95^{th} percentile are nearly 3.5 times larger than at the 5^{th} percentile. With a log-log C-R function the marginal damages are approximately 59% greater than with log-linear ($34,700 (log-linear), $55,100 (log-log), median marginal damages per ton). While the magnitude of the marginal benefits of abatement vary between the log-linear and log-log C-R functions, the distributions are quite similar. There is nearly perfect correlation between the marginal benefits of abatement by source with a log-linear and log-log C-R function. The form of the C-R function does not affect which sources are inflicting the greatest and least harm from pollution. Rather if the true C-R function is log-log all sources are contributing to greater damages than we previously believed.

How does the distribution of marginal benefits of abatement affect the outcomes from the efficient and uniform tax policies? Across the 1,000 model runs, the "best" uniform tax policy requires slightly greater emission reductions (23% for log-linear and 39% for log-log) than the efficient policy (22% for log-linear and 38% for log-log), yet the efficient policy has a greater impact on the population weighted average concentration (10.0 μg m^{-3} (log-linear), 8.6 μg m^{-3} (log-log)) compared to the uniform tax (10.3 μg m^{-3} (log-linear), 9.0 μg m^{-3} (log-log); see Figure 3.7). The efficient policy generates 33% greater net benefits than the uniform tax with a log-linear C-R function, and 27% greater net benefits with log-log. The level of inequality in concentration between grid squares, as measured by the Gini coefficient, is very similar between the efficient policy and the uniform tax.

The advantage of the efficient policy is that it directs more abatement to the sources

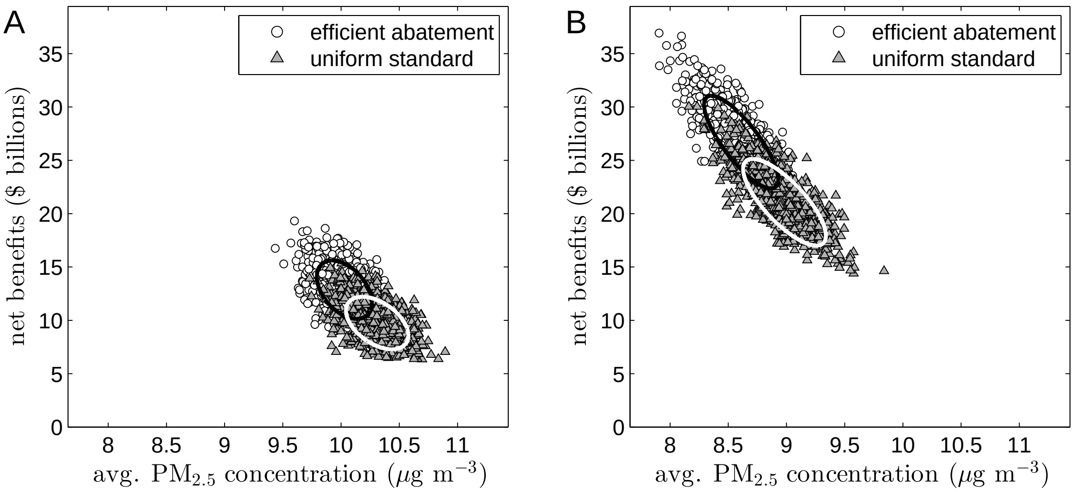


Figure 3.7: Efficient abatement policy versus uniform tax: net benefits of abatement and population-weighted average pollution concentration for each model run, (**A**) log-linear C-R function, (**B**) log-log C-R function. The four clusters each contain 1,000 points, of which 69% to 71% are encircled by the respective ellipse. Each ellipse radius represents one standard deviation from the mean. The points inside each ellipse are the outcomes nearest the center of the joint distribution of $PM_{2.5}$ concentration and net benefits.

in the most populated grid squares, where the damages from emissions are greatest. With a log-linear C-R function, the efficient policy abates 32% of emissions from sources in grid squares in the top decile of population, and only abates 9% of emissions from grid squares in the bottom decile. Under the uniform tax, with log-linear, all deciles abate 23% of emissions. With log-log, the efficient policy results in emission reductions of 51% from the top decile, and 22% from the bottom decile. Under the uniform tax, with log-log, all deciles abate 39% of emissions. The uniform tax policy, which does not differentiate between emissions by source, leads to excessive emissions in the most populated areas.

Given the wide distribution of marginal benefits of abatement, and the resulting welfare advantage from efficient abatement, a policy that differentiates between the emissions by source appears warranted. In addition, the inefficient aspects of the uniform tax policy would be magnified if the tax were applied to a larger region with a wider distribution of marginal benefits.

Table 3.1: Comparison of abatement policies to initial situation without regulation.

	Δ Emissions (tons)	Δ Concentration (μg m^{-3})	Net Benefits ($billions)
Log-linear:			
Efficient policy	-700,000	-2.2	12.8
Uniform standard	-560,000	-1.5	5.8
Uniform tax	-740,000	-1.9	9.6
Log-log:			
Efficient policy	-1,230,000	-3.6	26.5
Uniform standard	-980,000	-2.5	13.4
Uniform tax	-1,260,000	-3.2	20.8

NOTE.– the change in concentration is the population weighted average across grid squares

Table 3.1 summarizes the median outcomes resulting from the three pollution control policies considered in the model. The efficient policy leads to the lowest concentration of fine particulates and greatest net benefits. The uniform concentration standard results in the worst outcomes utilizing a command-and-control abatement policy. A uniform tax on emissions outperforms the uniform standard by achieving a cost-effective outcome, and results in the lowest total emissions of the three policies, but it provides insufficient incentives to the sources inflicting the greatest harm to limit emissions. Across the three policies, a log-log C-R function calls for lower emissions, lower concentrations of fine particulates, and greater net benefits of abatement compared to a log-linear functional form.

3.4 Conclusion

This paper contains a discussion of the effect of different shapes of a C-R function on environmental policy. An important distinction between the two functional forms considered is the impact on people facing the lowest concentration levels. If the log-log form reported in Krewski *et al.* (2009) is correct for fine particulates, then society may prefer substantially lower emissions. With the log-log functional form the benefits from a marginal unit of abatement are greater in clean locations than in dirty locations, all else equal. The log-log functional form leads to recommendations for significantly stricter pollution abatement policies, and correspondingly lower risks of mortality. Understanding the true shape of the C-R function between fine particulate concentrations and adult mortality is a worthy endeavor. Socially optimal policies are substantially different between the two functions.

The difference in policy outcomes between log-linear and log-log depend on the emissions within the area under the regulator's control. With greater unregulated emissions impacting concentrations within the area of interest, from emissions either outside the modeling domain or not subject to regulation within the region, the effectiveness of a policy achieving low concentrations is constrained, and the large potential risk reductions with log-log may not be atainable. Policies regulating all relevant emissions is crucial to realize the best possible outcomes.

Uniform pollution standards with command-and-control mechanisms to achieve the standards do not appear to be the economically preferred method of pollution control. We find that an efficient abatement policy leads to lower average concentrations while also achieving a better outcome for society. Environmental justice is a potential concern with an efficient abatement policy, because the focus is not on reducing risks for the most vulnerable populations. Contrary to our expectations, our results indicate that the level of inequality is similar or slightly less under an efficient abatement policy than under a uniform pollution standard, and both policies yield greater equality than conditions before abatement.

The application of Antweiler's (2012) idea for the iterative emissions tax to our computational problem appears to be interesting in its own right. He envisions an environmental regulator who actually adjusts the vector of source-specific taxes each period. That idea turns out to be powerful in an unexpected way: as the basis for a computationally efficient solution algorithm. The key to Antweiler's deep insight is that the regulator does not need to know abatement costs in order to guide the iterative policy to the optimum. Where his policy involves explicitly the passage of time, we use the same idea to solve an otherwise infeasible numerical problem that does not involve the passage of time. In both uses, Antweiler's and ours, regulated firms are assumed not to behave strategically in their response to each period's emission tax. If this assumption is not met, as in Moledina *et al.* (2003) and Kwerel (1977), and regulated sources anticipate the way their behavior in one periods feeds into the policy next period, one may expect the problem to become more difficult both computationally and in policy practice.

Our simulations identify the potential differences in outcomes and economically preferred policies between the two estimated Krewski *et al.* (2009) C-R functions. Application of this model to actual data, taken up in future work, will help to understand further the advantages and disadvantages of uniform environmental standards, such as the U.S. National Ambient Air Quality Standards. This model also highlights the importance of source-specific policies, suggesting that greater scrutiny is needed on cost-effective policies that do not account for spatial differences in damages by source.

Our analysis focused on the policy implications of a supralinear C-R function at low concentrations. Supralinearity may be important for places that face much higher concentrations. At the highest concentrations globally, a supralinear C-R function indicates that risk reductions would be comparatively small until substantially lower concentrations are achieved (Evans *et al.* 2013).

Chapter 4

Marginal Damages of Criteria Pollutant Air Emissions

4.1 Introduction

Analysis of criteria air pollution is complicated by many factors: the wide-spread dispersion of the pollutants, the hundreds of thousands of sources of emissions in the U.S., and the association between impacts to human health from exposure to pollution. This complexity makes it difficult to isolate the impact on affected areas from a single source of emissions, yet isolating this impact is at the heart of sound air pollution policy. An efficient approach to air pollution regulation requires that emissions from each source are set so that the marginal damages of emissions are equated to the marginal costs of emission abatement. Achieving an efficient outcome is not possible without isolating the impact of emissions from individual sources. Unlike emissions of carbon dioxide, where the location of the emission source does not affect the environmental impact, isolating the impact of emissions of criteria pollutants by source is especially important because marginal damages can have orders of magnitude difference depending on the source's location.

In this paper, the impact of emissions from the most heavily polluting sources in the U.S., measured by marginal damages, is estimated using a recently developed air pollution model. Previous estimates of marginal damages (or damages per ton) have generally been based on one of two types of models. One type is reduced form air dispersion models that lack a strong scientific foundation and have coarse geographic precision (Muller and Mendelsohn 2009; NRC 2010; Levy, Baxter and Schwartz 2009; EPA 2014a). The second type is complex chemical transport models (CTM) that cannot feasibly, given computational resources, isolate the impacts of each individual source, when examining emissions from many locations (EPA 2014b; EPA 2011; EPA 2008). The advantages of reduced form models are ease of use and replication, allowing for model runs that estimates the impacts from specific sources. Complex CTMs are precise and much more accurate than reduced form models, but require substantial computational resources, making it infeasible to preform the required model runs to analyze many sources individually. In an attempt to merge complex CTMs and reduced form models, the response surface model (RSM) was developed based on statistical estimation using model runs from complex CTMs (Fann, Fulcher and Hubbell 2009). The RSM was constrainted to estimating air pollution impacts to a limited number of sources in a few urban areas.

The Intervention Model for Air Pollution (InMAP), used for this analysis, is derived from a CTM but can be run much more quickly, using a simplified approach to chemical reactions and transport, allowing for a high volume of runs and fine geographic precision (Tessum 2014).[1] The estimates of marginal damages of emissions from many sources, provided in this paper, are thus based on a stronger foundation than those in the past.

This paper also examines the impact on marginal damage estimates of alternative shapes of the concentration-response (C-R) relationship between fine particulate (particles with diameter less than 2.5 microns, $PM_{2.5}$) air pollution and mortality. The C-R analyzed here come from the two most prominent studies in the U.S.: the Harvard-Six-City (H6C) study and the American Cancer Society (ACS) study. Estimates of the C-R from the ACS study (Pope *et al.* 1995; Pope *et al.* 2002; Krewski *et al.* 2009) has consistently found less severe, but still large and alarming, impacts of air pollution on mortality than the H6C study (Dockery *et al.* 1993; Laden *et al.* 2006; Lepeule *et al.* 2012).

The most recent estimates from the ACS study, by Krewski *et al.* (2009), utilized two functional forms of the C-R relationship: log-linear, a frequently used relationship, and log-log, and alternative relationship. The biggest difference between the log-linear and log-log C-R is the impact of $PM_{2.5}$ concentration reductions in areas with already low concentrations. The estimated log-log C-R in Krewski *et al.* indicates that the change in the mortality rate from a concentration reduction is largest in the lowest concentration areas. The log-log C-R implies that the marginal benefits of reducing emissions are increasing in abatement, an unusual phenomenon. If the true C-R relationship is log-log then more lives can be saved by reducing $PM_{2.5}$ concentrations to very low levels. Marginal damages of emissions of criteria air pollutants are greatly influenced both by the functional form of the C-R, and the magnitude of the effect of $PM_{2.5}$ concentrations on mortality from the ACS and H6C studies. These differences are analyzed analytically, and empirically using InMAP.

Section 4.2 analyzes the key factors that can influence marginal damage estimates. The marginal damage equation is derived and the implications of the different C-R functions are highlighted. Section 4.3 presents the results of marginal damage estimates

[1] The computational resources of a complex CTM are approximately 25,000 times greater than InMAP: complex CTMs require approximately a week of computing time on a super-computer, whereas InMAP can be run in a few hours on a personal computer.

for many sources of pollution using InMAP and compares them to previous estimates based on other methodology.

4.2 Methods

This paper examines the impact on human health from emissions of criteria pollutants. The damages from criteria pollutant emissions is associated primarily with their contribution to fine particulate concentrations. The focus is on fine particulates (as opposed to coarse particulates which have a wider diameter) because they go deep into the lungs when enhaled and possibly into the blood stream (EPA 2004). Fine particulates can also travel much greater distances and time in the atmosphere than coarse particulates, which settle more quickly and near the source of emission. Exposure to fine particulates, especially long term exposure, is associated with an increased risk of mortality. Fine particulates can be emitted directly from sources (primary $PM_{2.5}$) or can be formed in the atmosphere from chemical reactions (secondary $PM_{2.5}$). The total fine particulate concentration combines primary and secondary $PM_{2.5}$. Health impacts are linked with total fine particulate concentrations, and for this analysis we are agnostic to the chemical composition of the particle.[2]

Certain criteria pollutants can contribute to secondary $PM_{2.5}$ after emission. For instance, gas-to-liquid chemical reactions can convert SO_X into sulfate compounds that add to the total fine particulate concentration. The health impacts linked to the increased fine particulate concentration caused by emissions of a criteria pollutant is attributed to the source of emissions. By far the largest damages associated with the emission of the criteria pollutants examined here are due to their contribution to fine particulate concentrations.

In this analysis, the impacts of emissions are measured by the marginal damages from the emission of an additional ton of a pollutant from a source. Marginal damages from criteria pollutants are an important measure for policy makers, both to identify the most harmful emission sources, and to be compared with the marginal costs of pollution abatement strategies. Emission reductions at different sources are not equally valuable

[2] Fine particulates are comprised of several chemical components either as solid particles or liquid droplets. The major species are carbon, sulfate and nitrate compounds, and soil and ashe (EPA 2004).

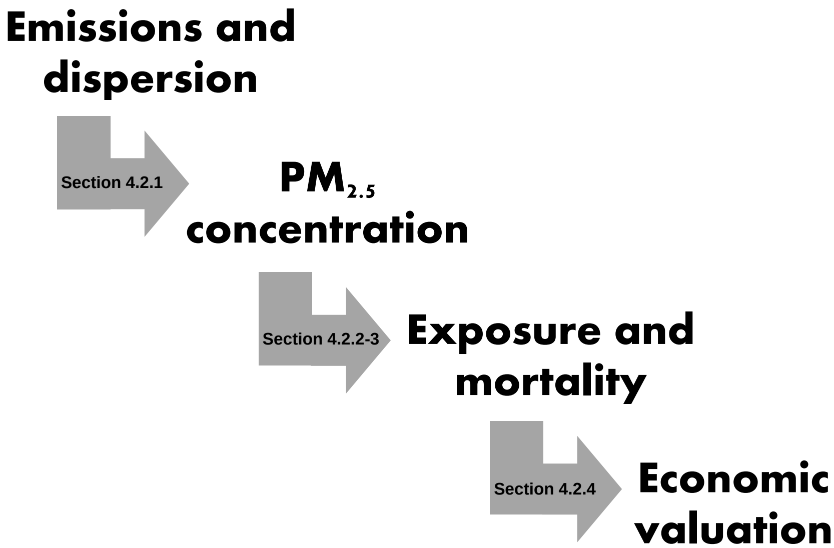


Figure 4.1: Diagram of link from emissions to damages.

to society and identifying the areas with the greatest opportunity to improve human health, and understanding the cause of the difference is necessary for efficient pollution control policies.

The link between emissions and marginal damages requires several steps as shown in Figure 4.1. First, the pollutant is emitted and disperses. Depending on the pollutant, certain chemical reactions can occur in the atmosphere, contributing to the ambient concentration of fine particulates in many receptors. The increased concentration of fine particulates raises the risk of mortality for the exposed population. Given the value of human life, the increased risk of mortality is translated into monetary damages.

Several factors influence the estimates of marginal damages. This paper seeks to highlight the differences in modeling that can influence estimates of the damages from emissions. We focus on three factors corresponding to the arrows in Figure 4.1. First, modeling the impact of criteria pollutant emissions on fine particulate concentrations in section 4.2.1. Dispersion of emissions and resulting pollution concentrations are exceedingly complex relationships to model, and the tradeoff between simplifying assumptions and computational requirements is key. The second factor examined, in sections 4.2.2

and 4.2.3, is the functional form and estimated magnitude of the effect of fine particulate concentrations on mortality. Exposure to fine particulates is believed to be a significant cause of premature mortality, however, the magnitude of the impact and the C-R relationship can greatly influence the estimated damages. The third factor is the appropriate value assigned for reductions in the risk of mortality (*i.e.* the value of a statistical life), and whether age should be considered, examined in section 4.2.4.

4.2.1 Air dispersion modeling

Most important for the validity of marginal damage estimates is the choice of modeling the change in concentrations of fine particulates resulting from a change in emissions of criteria pollutants. Usually the choice is between computationally intensive CTMs that produce state-of-the-science results, but are not specific to emissions at any single source, and reduced form models that can be run many times, but suffer from less accurate scientific representations of air dispersion and often lack desired geographic precision. This choice often determines the type of air pollution analysis that can be conducted.

Research with complex CTMs, which require substantial computational resources and run time, often use a limited number of runs of the air dispersion model, each as a comprehensive scenario of emission changes at many sources (Fann *et al.* 2012; EPA 2014b; EPA 2011). These model runs are valuable in assessing the impact and benefits of proposed pollution control policies at precise geographic locations. For instance, the CTM CAMx was used by the EPA when estimating the health co-benefits in the regulatory impact analysis of the proposed carbon pollution guidelines for existing power plants (EPA 2014b). Because each scenario involves emission changes at many sources, it is not possible to separate the benefits attributable to the emission reductions at any specific source. Without this information we cannot identify the most damaging sources of pollution, implement an economically efficient policy of pollution abatement, or select emission reductions to meet a specific pollution concentration target in a receptor of concern. Essentially, pollution abatement becomes an exercise of applying educated guesses as to the appropriate locations and quantities of emission reductions that"should" be made without sufficient economic justification. With state, regional or national policy initiatives based on this method, excessive and inefficient pollution reductions may result.

The reduced form Air Pollution Emission Experiments and Policy analysis (APEEP) model (Muller and Mendelsohn 2007) is designed to run thousands of scenarios of emission changes. Each scenario involves a change of one ton of emissions from a single source. The damages in all receptors in the U.S. are calculated after each scenario and compared with the damages from the baseline scenario (no emission change). The difference between the scenario with the emission change and the baseline is the marginal damages attributable to the source. This process is repeated thousands of times to obtain marginal damages of emissions from each county in the U.S.

The form of the results is ideal for economic analysis. However, the modeling of air pollution in APEEP is derived from the Climatological Regional Dispersion Model (CRDM) source-receptor (S-R) matrix. The CRDM S-R matrix represents the change in fine particulate concentration in any receptor in the U.S. from a change in emissions from any source in the U.S. The S-R matrix from the CRDM is calculated using a Gaussian plume model which requires many simplifying assumptions and may not produce accurate results. For instance, Gaussian plume models assume that the wind-speed is constant and the same at any elevation, and pollution travels in a straight line from the source (Masters and Ela 2008, page 452). Even more sophisticated Gaussian plume models are generally not recommended for predicting changes in pollution concentrations beyond 50 km from the emission source (EC/R 2014; Caputo, Gimenez and Schlamp 2003). Modeling secondary fine particulate concentrations from SO_X and NO_X emissions is complicated by the time required to travel to a receptor, and is difficult to model with Gaussian plume models that are steady state (Holmes and Morawska 2006).

APEEP estimates the marginal damages of emissions of several pollutants from each county in the Contiguous U.S. The geographic precision of APEEP is substantially less than the complex CTMs that are often run at 12 km grid cells across the U.S.[3]

In a recent report, the National Resource Council (NRC) conducted a detailed analysis of the externality costs of energy in the U.S. (NRC 2010). In order to estimate the damages per ton of emissions from each coal power plant in the U.S., the NRC used

[3] Compared to 144 km^2 area of 12 km grid cells, the median (and 5^{th} and 95^{th} percentile) area of counties in the Contiguous U.S. is 1687 km^2 (638 km^2, 7919 km^2).

APEEP. The authors chose APEEP rather than a complex CTM because the computational resources required to make these calculations would have been infeasible to implement (NRC 2010, page 83). Despite the valuable research produced using APEEP (Muller and Mendelsohn 2007, 2009; Henry *et al.* 2011; NRC 2010), there is room for improvement. Estimates of marginal damages derived from more complex scientific relationships, instead of Gaussian plume models, enhances confidence that atmospheric dispersion of emissions and changes in $PM_{2.5}$ concentrations are accurately represented. In addition, models with geographic precision smaller than county level for sources of emissions and receptors of air pollution increases the credibility and applicability of marginal damage estimates.

The analysis presented here applies InMAP which is designed to bridge the divide between complex CTMs and simple reduced form models. The results in this paper, which are based upon InMAP, provide estimates of marginal damages from many important sources of pollution in the U.S. The results lay out a pathway to an alternative model to APEEP based on more advanced air dispersion modeling and greater geographic precision.

InMAP is derived from WRF-Chem, a complex CTM, but InMAP is designed specifically for annual changes in $PM_{2.5}$ concentrations from emissions of criteria pollutants (Tessum 2014). The overwhelming majority of mortality related damages from criteria pollutants are estimated based on annual average $PM_{2.5}$ concentrations (EPA 2014b, page 4-21). By focusing on annual averages and ignoring hourly, daily, weekly and even monthly changes in concentrations, InMAP is able to calculate expected annual impacts from criteria pollutants at a fraction of the time and computing resources of WRF-Chem. It is therefore feasible to use InMAP to estimate the impacts from emissions from a single source. Through repeated model runs of InMAP, marginal damages of emissions from hundreds of sources are estimated in this paper. With substantially more model runs an S-R matrix of impacts between all model grid cells could be created for use in a reduced form model similar to APEEP. The creation of an S-R matrix is a worthwhile exercise that is in the process of being produced. Testing the reliability of InMAP's predicted concentration changes to WRF-Chem's is ongoing but incomplete; however, InMAP provides the potential for a stronger scientific basis to the modeling of the isolated impacts of emissions from many sources of pollution than has previously been produced from APEEP.

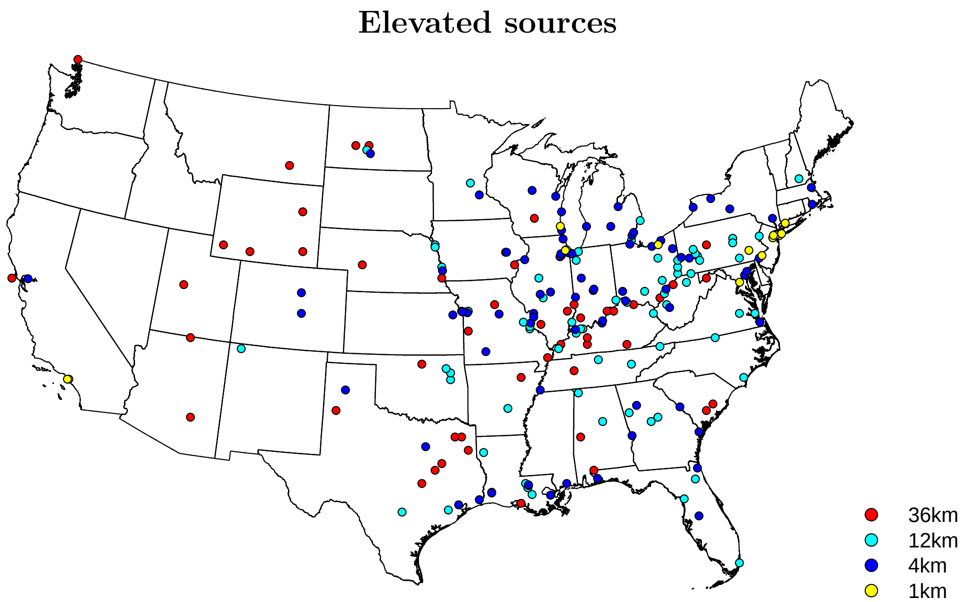
The geographic precision of InMAP also improves upon previous reduced form models. InMAP is designed with variable grid cell sizes based on population density. The primary grid cell unit is 36 km $\times$ 36 km which is used in sparsley populated regions of the U.S. For areas with progressively denser populations, the grid cells are squares with 12 km, 4 km and 1 km sides. This method allows for more precise estimates of impacts in areas with the highest populations, which, on average, have the largest damages from pollution. In addition, the geographic precision is based on population density rather than the more arbitrary delination between counties, as in APEEP. This difference in model precision is most important in large counties, such as Los Angeles County, CA, which receives one estimate in APEEP and assumes a uniform $PM_{2.5}$ concentration across the county, but is broken into over 1,000 grid cells in InMAP. With InMAP it is therefore possible to conduct environmental justice analyses for socioeconomic and racial group disparities given the small grid cells in highly populated areas.

InMAP is run by inputting emission changes, from a baseline, at any number of grid cells of eight primary pollutants: $PM_{2.5}$, PM_{10}, SO_X, NO_X, VOC, NH_3, CH_4 and CO. The baseline emissions and pollution concentrations are from a WRF-Chem model run representing 2005 conditions. InMAP then calculates the dispersion of the pollutants and the chemical reactions that occur in the atmosphere. The output of the model is the change in the fine particulate concentration at each receptor grid cell in the model resulting from the scenario of emission changes.

For the present analysis, InMAP was run by inputting a one-ton emission change from a single source grid cell. The results describe the isolated impact on total fine particulate concentrations at every receptor grid cell in the model from a one-ton emission change at the source. This process was repeated for 223 source grid cells for elevated emissions,[4] and 675 source grid cells for ground level emissions. Figure 4.2 displays the location of the elevated and ground source grid cells. For elevated sources InMAP was run for primary polluant changes of $PM_{2.5}$, SO_X, and NO_X. Ground level sources were run with emission changes of $PM_{2.5}$, SO_X, NO_X, and NH_3.[5]

[4] The effective stack height of elevated emissions sources is assumed to be 100 m.

[5] Almost all NH_3 emissions originate from ground level sources, and are therefore excluded from analysis of elevated sources. PM_{10}, CO and CH_4 emissions do not contribute to secondary $PM_{2.5}$ and thus excluded. VOC emissions contribute to secondary organic arosols that are included in the total fine



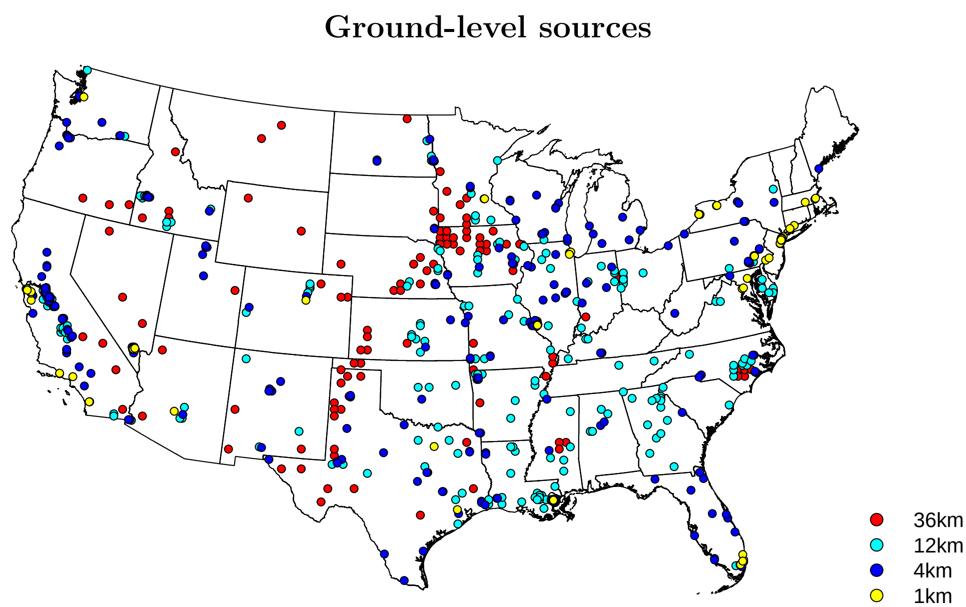


Figure 4.2: Locations of 223 elevated sources and 675 ground sources coded by source grid cell size.

---

particulate concentration. VOC emissions are modeled in InMAP, but the results are not presented here because testing of InMAP has shown the impact of VOC emissions on fine particulate concentrations

The goal of this analysis is to characterize the distribution of marginal damages from the most heavily polluting locations in the U.S., and from areas with various levels of population density. Source grid cells were chosen based on the quantity of criteria pollutant emissions in the cell and the grid cell size. Emissions from the 2011 National Emissions Inventory (EPA 2013a) from elevated and ground sources of the primary pollutants were estimated at each InMAP grid cell.[6] The proportion of total emissions, r_j, for a pollutant, p, were calculated for each grid cell j, and for elevated (l) and ground-level (g) emissions:

$$r_j^{p,l} = \frac{e_j^{p,l}}{\sum_{j=1}^{N} e_j^{p,l} + e_j^{p,g}} \quad \text{and} \quad r_j^{p,g} = \frac{e_j^{p,g}}{\sum_{j=1}^{N} e_j^{p,l} + e_j^{p,g}},$$

where p is an element of $P = \{\mathrm{PM}_{2.5}, \mathrm{SO}_X, \mathrm{NO}_X, \mathrm{NH}_3\}$, and N is the number of grid cells in InMAP. The proportions were summed across all pollutants for a grid cell, separately for elevated and ground level emissions, to create emission scores, S_j^l and S_j^g.

$$S_j^l = \sum_P r_j^{p,l} \quad \text{and} \quad S_j^g = \sum_P r_j^{p,g}.$$

The 223 elevated sources selected for analysis were the grid cells with the highest score with a mix from each of the four grid cell sizes: 36 km, 12 km, 4 km, 1 km squares. The fraction of sources from each grid cell size was based on the proportion of total emissions from each grid cell size.[7] Emissions from the selected elevated sources are 34% of U.S. total elevated $\mathrm{PM}_{2.5}$ emissions, 72% of total elevated SO_X emissions, and 41% of total elevated NO_X emissions. The 675 ground-level sources were chosen with the same method as the elevated sources. Emissions from the ground-level sources are 8% of U.S. total ground-level $\mathrm{PM}_{2.5}$ emissions, 7% of total ground-level SO_X emissions, 7% of total ground-level NO_X emissions, and 18% of total ground-level NH_3 emissions. Despite chosing sources with the greatest emissions, Figure 4.3 shows that the selected

---

are underestimated.

[6]Point sources were placed in the corresponding InMAP grid cell. Ground, mobile and area emissions were available at county level and were allocated to InMAP squares based on area and population of grid cell.

[7]Of the 223 elevated source grid cells selected 22% are from 36 km cells, 31% from 12 km cells, 37% from 4km cells, and 10% from 1km cells. Of the 675 ground level source grid cells selected 16% are from 36 km cells, 27% from 16 km cells, 38% from 4km cells, and 19% from 1 km cells.

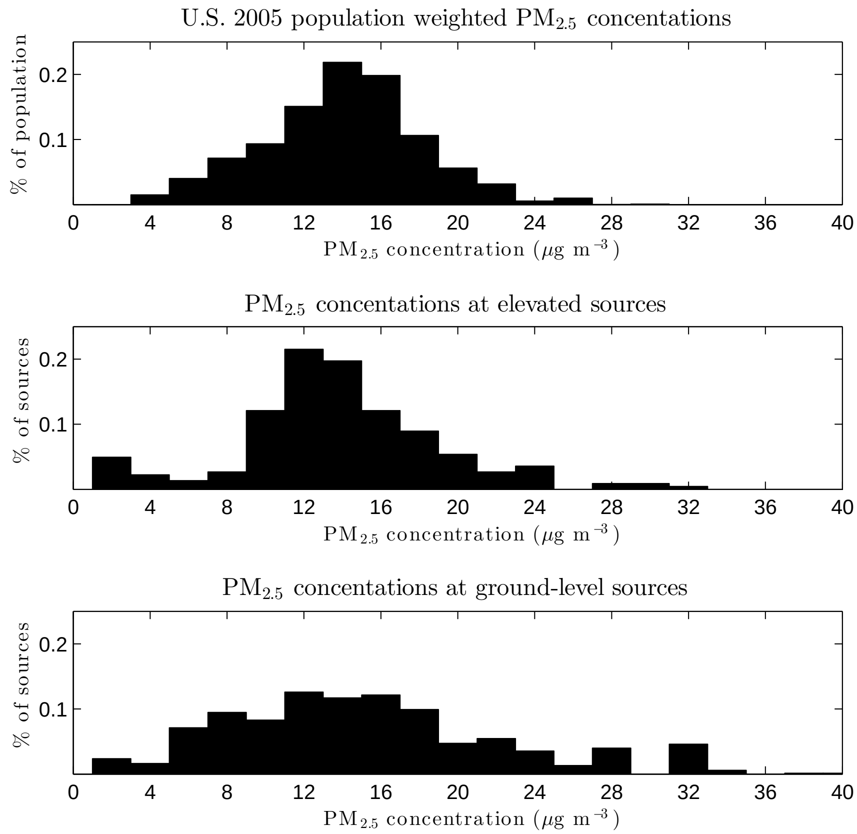


Figure 4.3: Distributions of 2005 $PM_{2.5}$ concentrations: U.S. (top panel), InMAP grid cells of 223 elevated sources (middle panel), and InMAP grid cells of 675 ground-level sources (bottom panel).

sources represent areas with a wide distribution of fine particulate concentrations. These distributions are similar to the U.S. population weighted $PM_{2.5}$ concentrations in 2005.

Different methods could be employed to select the source grid cells. Here, we are attempting to demonstrate and explain the difference in marginal damages of emissions from the locations that emit the greatest amount of criteria pollutants. The sources selected are therefore not a representative sample of locations across the U.S. The 223 grid cells of the elevated sources represent 0.8% of the total grid cells in InMAP, 1% of the total U.S. population, and 0.9% of the total area of the Contiguous U.S. The 675 grid cells of the ground-level sources represent 2.4% of the total grid cells in InMAP, 3.6% of the total U.S. population, and 2.0% of the total area of the Contiguous U.S.

Regardless of the method chosen to select the source grid cells, the marginal damages vary dramatically depending on the source of emission, as will be shown in the results.

The sources represent the grid cells with the greatest amount of emissions from the 2011 emissions inventory from the EPA, the most recently available version. However, the baseline concentrations in InMAP represents 2005 conditions, when average $PM_{2.5}$ concentrations were substantially higher (EPA 2012a). In terms of the sources selected, this discrepancy is not necessarily problematic. Whether the largest emitting sources changed from 2005 to 2011, the sources selected here still provide a wide range of sources across the U.S. based on location, $PM_{2.5}$ concentration and population, to examine the differences in marginal damages.

The change in fine particulate concentrations between 2005 and 2011 has a large effect on the estimates of marginal damages if calculated with a log-log C-R. The higher fine particulate concentrations, on average, in 2005 will underestimate the marginal damages with a log-log C-R for emissions in 2011. This is discussed in greater detail in section 4.2.2. To account for the magnitude of the difference with a log-log C-R, marginal damages are also calculated with estimated $PM_{2.5}$ concentrations for 2013. This marginal damage calculation with 2013 concentrations is only valid if the marginal change in concentrations in 2005 from a change in emissions, predicted by InMAP, is the same with the lower concentrations in 2013. From sensitivity testing of InMAP, the impact per ton of emissions on concentrations is nearly identical regardless of whether the change in emissions is marginal or large and discrete.[8] This feature is explained by the assumptions in InMAP, which as configured here, does not account for the change in chemical reaction rates when underlying atmospheric conditions (*e.g.* baseline pollution concentrations, chemical reaction rates) change. It is also possible that the marginal change in concentrations from a change in emissions at a source in 2005 is a reasonable estimate of the marginal change in concentration in 2013 at the same source. Importantly, this is not suggesting that the marginal damage of emissions are the same from 2005 to 2013, because the concentration level may alter the impact of $PM_{2.5}$ exposures on mortality. Future versions of InMAP will need to be created with a new baseline scenario to determine whether the marginal impact on concentrations

[8]Several test runs of InMAP showed that fractional or large changes in emissions from a single source or from many sources simultaneously resulted in the nearly identical impacts per ton of emission changes.

from a source of emissions is substantially different depending on baseline fine particulate concentration and other ambient air and meteorological conditions. This is an important question for the applicability of results from reduced-form air dispersion models to time periods and atmospheric conditions that differ from the baseline conditions. For now we use an estimate of 2013 baseline concentrations and InMAP emissions impacts for 2005 conditions to assess the marginal damages in 2013.[9]

4.2.2 Concentration-response relationship

We examine how the C-R relationships estimated from the H6C and ACS studies impacts marginal damages of emissions. Both studies indicate a strong association between mortality rates and ambient concentrations of fine particulates. Because everyone is exposed to ambient air and human life is sacrosanct, damages from mortality caused by fine particulate air pollution are enormous.[10] The question is how large are the damages and what should be done about it. In the most recent analyses of the H6C study (Lepuele *et al.* 2012) the magnitude of the estimated effect of $PM_{2.5}$ concentrations on mortality is approximately double the estimate in the ACS study (Krewski *et al.* 2009). This difference in estimated risk of mortality leads to a corresponding approximate doubling of the estimated marginal damages from emissions when using the H6C estimate compared with the ACS estimate.

In addition to the large discrepancy in the magnitude of the health impact between the two studies, there is also a question regarding the correct functional form of the relationship between fine particulate pollution and mortality, the concentration-response. It is typically assumed (EPA 2014b; EPA 2011; Lepeule *et al.* 2012; Muller and Mendelsohn 2009; NRC 2010) that the relative risk of mortality between any two pollution concentration levels follows a log-linear C-R. With log-linear there is a constant relative risk between any fixed concentration difference, regardless of the level of the concentration. For instance, in the analysis by Krewski *et al.* (2009, table 11) it is estimated that the risk of mortality at 15 μg m^{-3} is approximately 6% greater than at 5 μg m^{-3}; and similarly, the risk at 20 is 6% greater than at 10 μg m^{-3}. The Krewski *et*

[9]For comparision, the CRDM used by APEEP is based on 1990 meteorolgical data and 1996 emissions (Abt Associates Inc. 2000), yet is currently in use to calculate marginal damages.

[10]The Office of Management and Budget estimates total benefits of fine particulate regulations between \$19 and \$169 billion per year (OMB 2013).

al. (2009) analysis also estimated an alternative log-log C-R. The log-log C-R indicates that the relative risks of mortality are substantially lower at very low concentrations of fine particulates than at higher concentrations. Specifically, the risk is estimated to be approximately 6% greater at 20 μg m^{-3} compared to 10μg m^{-3}, but the risk is 9.5% greater at 15 μg m^{-3} compared to 5 μg m^{-3}. Krewski *et al.* (2009, page 27) state that the log-log C-R was a slightly better predictor of the difference in mortality risk than the log-linear function. The Krewski *et al.* (2009) analysis does not imply that the true relationship between mortality and fine particulate air pollution is log-log, but it does identify an uncertainty as to the true shape and a need for greater attention to this question.

The choice of a log-linear or a log-log C-R may have substantial impacts on the marginal damages from emissions and on the preferred policies to control pollution (Goodkind *et al.* 2014). If the true C-R is log-log there exists an opportunity to reduce air pollution related mortality at a greater rate per unit of air pollution reduction in low- compared to high-concentration areas. If the log-log C-R is correct, there is a possibility that in order to achieve the greatest mortality reductions, a larger share of resources should be directed at reducing air pollution in the cleanest locations. Prioritizing concentration reductions in the cleanest locations is at odds with current pollution abatement policies in the U.S., and contrary to objectives aimed at promoting environmental justice as minority and socioeconimcally vunerable populations are more likely to live in areas with greater environmental risks, including higher $PM_{2.5}$ concentrations (Marshall, Zwor and Nguyen 2014). With log-log, even though the risk reduction is greatest in areas where the concentration of $PM_{2.5}$ is low, the dirtiest locations often have larger populations.[11] Estimates of marginal damages of emissions are largely determined by the size of the population impacted as a greater share of each ton of the emitted pollutant is inhaled.[12] With a log-log C-R, high concentration areas with larger populations may have greater marginal damages per unit of concentration than low concentration areas with smaller populations, even though the rate of mortality risk is reduced more per unit of concentration reduction at low concentrations.

The present analysis estimates empirically and compares the marginal damages of

[11]The correlation between the log (base 10) population density and $PM_{2.5}$ concentrations based on the 2005 baseline model run in InMAP is 0.62.

[12]See Bennett *et al.* 2002 for discussion of intake fraction.

emissions for both a log-linear and a log-log C-R function. Unlike with a log-linear C-R, with log-log, baseline $PM_{2.5}$ concentrations greatly influence the marginal damage calculation. Because the InMAP baseline concentrations are modeled on 2005 conditions when fine particulate concentrations were significantly higher, the analysis here may underestimate the marginal damages with log-log based on the lower concentrations that currently exist. Therefore, an additional estimate of marginal damages is evalutated with estimated 2013 baseline $PM_{2.5}$ concentrations. An estimate of the 2013 $PM_{2.5}$ concentration at each InMAP grid cell (see Figure 4.4) is produced from the change in observed $PM_{2.5}$ concentrations at monitoring stations in 2005 and 2013.[13]

4.2.3 Marginal damage equation

The equation for damages from criteria pollutant emissions depends on the functional form of the chosen C-R. The derivation starts with the equation for mortality, $\mathcal{M}_i$, in receptor i as a function of the concentration of fine particulates, C_i, the population, Pop_i, and the mortality rate, λ_i: $\mathcal{M}_i(C_i) = \text{Pop}_i \cdot \lambda_i(C_i)$. We define the baseline mortality rate given the baseline fine particulate concentration C_i^0 as $\lambda_i^0 = \lambda_i(C_i^0)$. When the concentration level deviates from the baseline, the resulting change in mortality is

$$\begin{aligned} \Delta\mathcal{M}_i(C_i) &= \mathcal{M}_i(C_i) - \mathcal{M}_i(C_i^0) \\ &= \text{Pop}_i \cdot \lambda_i^0 \left[\frac{\lambda_i(C_i)}{\lambda_i^0} - 1 \right], \end{aligned} \tag{4.1}$$

where the ratio $\lambda_i(C_i)/\lambda_i^0$ is the relative risk of mortality, $\text{RR}_i(C_i)$, in receptor i between an alternative concentration and the baseline concentration. The equation for $\text{RR}_i(C_i)$

[13] 2013 concentrations for InMAP grid cells were estimated using observed concentration changes between 2005 and 2013. $PM_{2.5}$ concentrations from 1,074 monitoring locations in 2005 and 961 monitoring locations in 2013 were obtained from the EPA AirData database (EPA 2014c). These concentrations were applied to InMAP grid cells by averaging the nearest three monitors weighted by the inverse of the distance between the grid cell and the monitor. Define the InMAP concentrations from the nearest three monitoring station observations as $\hat{C}_i^{O,2005}$ and $\hat{C}_i^{O,2013}$, for 2005 and 2013, respectively. The ratio of the 2013 to 2005 observed concentrations at each grid cell was applied to the modeled 2005 InMAP concentrations to obtain estimates of the 2013 InMAP concentrations:

$$\bar{C}_i = \frac{\hat{C}_i^{O,2013}}{\hat{C}_i^{O,2005}} C_i,$$

where C_i are the 2005 baseline concentrations derived from WRF-Chem, and $\bar{C}_i$ are the estimated 2013 concentrations for the InMAP grid cells.

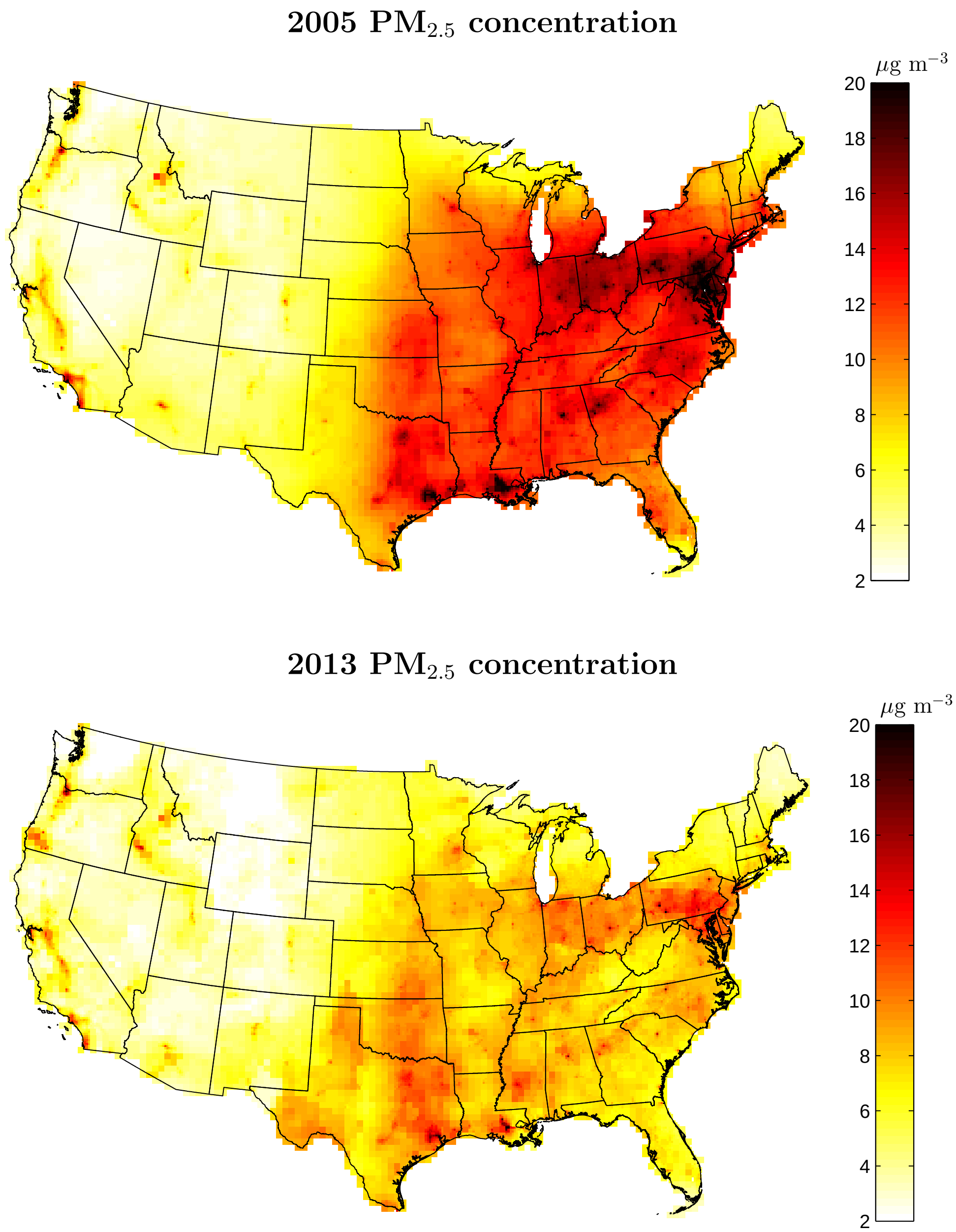


Figure 4.4: $PM_{2.5}$ concentrations (μg m^{-3}): top panel - InMAP 2005 modeled concentrations (derived from WRF-Chem), bottom panel - estimated 2013 concentrations for InMAP grid cells from observed $PM_{2.5}$ concentrations (see footnote 13 for explanation of 2013 InMAP concentrations calculations).

depends on the assumed C-R between adult mortality and fine particulate concentrations. Here we examine the log-linear (lin) and log-log (log) C-R forms. The log-linear relationship assumes that the natural log of mortality risk has a linear relationship with $PM_{2.5}$ concentrations and other controlling variables, $\mathbf{X}$:

$$\begin{aligned} \ln\left(\lambda_i^{\text{lin}}(C_i)\right) &= \beta^0 + \boldsymbol{\beta}^{\text{X}}\mathbf{X}_i + \gamma^{\text{lin}}C_i \quad \text{or} \\ \lambda_i^{\text{lin}}(C_i) &= \exp\left\{\beta^0 + \boldsymbol{\beta}^{\text{X}}\mathbf{X}_i\right\}\exp\left\{\gamma^{\text{lin}}C_i\right\}. \end{aligned} \tag{4.2}$$

Taking the ratio of (4.2) evaluated at C_i and C_i^0, the relative risk of mortality with the log-linear C-R is

$$\text{RR}_i^{\text{lin}}(C_i) = \frac{\lambda_i^{\text{lin}}(C_i)}{\lambda_i^{\text{lin},0}} = \exp\left\{\gamma^{\text{lin}}(C_i - C_i^0)\right\}. \tag{4.3}$$

The log-log C-R assumes that the natural log of mortality risk has a linear relationship with the natural log of fine particulate concentrations and other controlling variables, $\mathbf{X}$.

$$\begin{aligned} \ln\left(\lambda_i^{\log}(C_i)\right) &= \beta^0 + \boldsymbol{\beta}^{\text{X}}\mathbf{X}_i + \gamma^{\log}\ln(C_i) \quad \text{or} \\ \lambda_i^{\log}(C_i) &= \exp\left\{\beta^0 + \boldsymbol{\beta}^{\text{X}}\mathbf{X}_i\right\}(C_i)^{\gamma^{\log}}. \end{aligned} \tag{4.4}$$

The log-log relative-risk equation is

$$\text{RR}_i^{\log}(C_i) = \frac{\lambda_i^{\log}(C_i)}{\lambda_i^{\log,0}} = \left(\frac{C_i}{C_i^0}\right)^{\gamma^{\log}}. \tag{4.5}$$

The γ parameters can be estimated from the relative-risk results from the ACS study presented in Krewski *et al.* (2009) and from the H6C study in Lepeule *et al.* (2012). Solving equation (4.3) for γ^{lin} and applying the relative-risk estimates for a 10 μg m^{-3} change in $PM_{2.5}$ concentration, the log-linear parameters for the ACS and H6C studies are $\gamma_{\text{ACS}}^{\text{lin}} = 0.00583$ and $\gamma_{\text{H6C}}^{\text{lin}} = 0.0131$, respectively. Solving equation (4.5) for $\gamma^{\log}$ and applying the relative-risk estimates from Krewski *et al.* for a change in $PM_{2.5}$ concentration from 20 to 10 μg m^{-3}, the log-log parameter for the ACS study is $\gamma_{\text{ACS}}^{\log} = 0.0827$.

Applying the relative-risk equation to the change in mortality in (4.1) we can calculate the damages, D_i, in receptor i from a change in concentration.

$$\mathrm{D}_i(C_i) = \mathcal{V} \cdot \mathrm{Pop}_i \cdot \lambda_i^0 \left[\mathrm{RR}_i(C_i) - 1\right] \tag{4.6}$$

where $\mathcal{V}$ is the value of a statistical life (VSL), which we assume for the moment is a uniform value applied for all individuals. Equation (4.6) is the general form of damages for any relative-risk function. Applying (4.3) and (4.5) to (4.6) produces different estimates of the damages due to a change in concentration.

Next we examine the marginal change in damages in receptor i from a marginal change in concentration, separately for the log-linear and log-log C-R:

$$\mathrm{MD}_i^{\mathrm{lin}}(C_i) = \frac{\partial \mathrm{D}_i^{\mathrm{lin}}(C_i)}{\partial C_i} = \mathcal{V} \cdot \mathcal{M}_i(C_i^0) \cdot \gamma^{\mathrm{lin}} \cdot \mathrm{RR}_i^{\mathrm{lin}}(C_i) \tag{4.7}$$

$$\mathrm{MD}_i^{\mathrm{log}}(C_i) = \frac{\partial \mathrm{D}_i^{\mathrm{log}}(C_i)}{\partial C_i} = \mathcal{V} \cdot \mathcal{M}_i(C_i^0) \cdot \frac{\gamma^{\mathrm{log}}}{C_i} \cdot \mathrm{RR}_i^{\mathrm{log}}(C_i). \tag{4.8}$$

The marginal-damage equation explains how damages in a receptor change for any concentration reduction below the baseline. The log-linear marginal damage equation, (4.7), changes with different concentrations, C_i, only through the impact on the mortality rate in the relative-risk equation. This aspect causes the value of the marginal-damage equation to decrease with concentration reductions, but only by the degree to which $PM_{2.5}$ concentrations impact mortality rates. Figure 4.5 illustrates the marginal-damage equation of the log-linear C-R with the ACS and H6C parameter estimates for a single hypothetical receptor with an initial $PM_{2.5}$ concentration of 16 μg m^{-3}. The steeper slope of the H6C marginal damage curve reflects the larger estimated effect of $PM_{2.5}$ concentrations on mortality rates in the H6C study compared with the ACS study.

The log-log marginal damage equation, (4.8), is impacted by $PM_{2.5}$ concentrations, C_i, both directly in the denominator, and indirectly through the relative-risk equation. However, while concentration reductions lower the value of the relative-risk equation, the C_i term in the denominator causes marginal damages to increase with lower $PM_{2.5}$ concentrations. The net effect is that marginal damages increase as $PM_{2.5}$ concentrations fall. Put another way, with the log-log C-R more lives are saved from a unit of

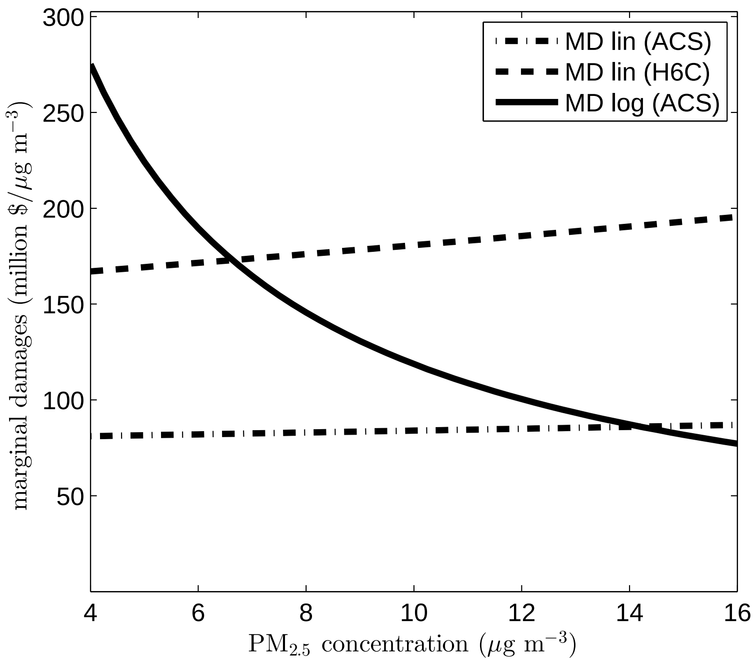


Figure 4.5: Illustration of change in marginal damages across receptor concentration. The hypothetical receptor has a population of 250,000, and initial mortality rate, λ_i^0, of 806.6 per 100,000 population.

$PM_{2.5}$ reduction at a receptor if there is a low concentration compared with the same reduction at the receptor if there is a high concentration.

Figure 4.5 illustrates the log-log C-R with the ACS parameter estimate. The figure indicates that for receptors with a relatively high fine particulate concentration (*e.g.* 12-16 μg m^{-3}), the greatest difference in estimated marginal damages is between the use of the ACS or the H6C log-linear parameter, whereas the functional form of the C-R is less critical. For low-concentration receptors (*e.g.* 4-6 μg m^{-3}), the distinction between log-linear and log-log becomes more important to the estimation of $PM_{2.5}$ health damages than the difference between the ACS and H6C log-linear estimates. In the 2005 InMAP baseline, low-concentrations receptors tended to contain lower populations, and therefore, less air pollution damages. In the future, if fine particulate concentrations continue to decline, the large difference in estimated damages between the C-R functional forms will apply to larger populations. If lower concentrations are achieved for a large portion of the population then the possibiliy exists that only limited pollution abatement would be economically justified with a log-linear C-R, but substantially greater abatement may be optimal if the C-R is log-log.

Equations (4.7) and (4.8) refer to changes in concentration in a single receptor, but abatement of pollution originates from the sources of emissions. Air pollution policies must focus on damages at all relevant sources of pollution even if the policy goal is to reduce $\mathrm{PM}_{2.5}$ concentrations in a single receptor. Therefore, we seek to calculate the marginal damages of *emissions* of the pollutants that contribute to total $\mathrm{PM}_{2.5}$ concentrations by differentiating the damages in receptor i (equation 4.6) with respect to a change in emissions from source j. We start with the marginal damages to a single receptor i of emission changes of pollutant p from source j at height h:[14]

$$\mathrm{MD}_{ji}^{p,h}(\Delta\mathbf{e}) = \frac{\partial \mathrm{D}_i\big(C_i(\Delta\mathbf{e})\big)}{\partial e_j^{p,h}} = \mathrm{MD}_i\big(C_i(\Delta\mathbf{e})\big)\frac{\partial C_i(\Delta\mathbf{e})}{\partial e_j^{p,h}}. \tag{4.9}$$

This is the marginal change in damages in receptor i from a one-ton change in emissions from source j, where $\mathrm{MD}_i\big(C_i(\Delta\mathbf{e})\big)$ is the marginal damage of a concentration change in receptor i, either equation (4.7) or (4.8) depending on the C-R. In equation (4.9), we see that the marginal damages in i from emissions at j are given by the marginal damages in receptor i from a concentration change ($\mathrm{MD}_i(C_i)$) weighted by the marginal impact of emissions from source j on the concentration in i ($\partial C_i(\Delta\mathbf{e})/\partial e_j^{p,h}$). The concentration of $\mathrm{PM}_{2.5}$, $C_i(\Delta\mathbf{e})$, is now a function of a three-dimensional matrix of emission *changes*, $\Delta\mathbf{e}$, from the baseline. The matrix $\Delta\mathbf{e}$ has one dimension for each grid cell source in InMAP, a second dimension for each pollutant, and a third dimension for the height of emission, $(N \times P \times H)$. The $\mathrm{PM}_{2.5}$ concentration in receptor i is affected by the emissions at all source grid cells of several pollutants. We define the concentration in receptor i based on its deviation from the baseline

$$C_i(\Delta\mathbf{e}) = C_i^0 + \sum_H \sum_P \sum_{j=1}^{N} \Delta e_j^{p,h} \cdot \pi_{ji}^{p,h}, \tag{4.10}$$

where p is the primary emitted pollutant that is an element of $P = \{\mathrm{PM}_{2.5}$, SO_X, NO_X,

[14]The height of emission, h, refers to either ground-level emissions (g) or the effective stack height of elevated emissions (l).

$\mathrm{NH_3}\}$; $\Delta e_j^{p,h}$ is the emission deviation from the baseline for source grid cell j, with positive values indicating an emission increase; and $\pi_{ji}^{p,h}$ is the marginal impact of emissions of primary pollutant p at height h from source j on the total $\mathrm{PM}_{2.5}$ concentration in receptor i. Each run of InMAP used in this analysis, with a one-ton change in emissions of a single pollutant, from one source produces a vector, $\boldsymbol{\pi}_j^{p,h}$, of impacts to all N receptors. The i^{th} element of this vector is $\pi_{ji}^{p,h}$. The final term in (4.9) is evaluated as $\partial C_i(\Delta\mathbf{e})/\partial e_j^{p,h} = \pi_{ji}^{p,h}$.

For this analysis we are most interested in the marginal damages of a change in emissions from the baseline conditions. Evaluating (4.9) at the baseline ($C_i = C_i^0$ or $\Delta\mathbf{e} = \mathbf{0}$) we can compare the marginal damages from source j to receptor i with the log-linear and log-log C-R[15]

$$\mathrm{MD}_{ji}^{\mathrm{lin},p,h}(\Delta\mathbf{e} = \mathbf{0}) = \pi_{ji}^{p,h} \cdot \mathrm{MD}_i^{\mathrm{lin}}(C_i^0) = \pi_{ji}^{p,h} \cdot \mathcal{V} \cdot \mathcal{M}_i(C_i^0)\gamma^{\mathrm{lin}} \quad \text{and} \tag{4.11}$$

$$\mathrm{MD}_{ji}^{\log,p,h}(\Delta\mathbf{e} = \mathbf{0}) = \pi_{ji}^{p,h} \cdot \mathrm{MD}_i^{\log}(C_i^0) = \pi_{ji}^{p,h} \cdot \mathcal{V} \cdot \mathcal{M}_i(C_i^0)\frac{\gamma^{\log}}{C_i^0}. \tag{4.12}$$

Comparing equations (4.11) and (4.12) and applying the parameter estimates from Krewski *et al.* (2009) for γ^{lin} and $\gamma^{\log}$, the marginal damages with a log-linear C-R are larger than with log-log if the baseline concentration is greater than 14.2 μg m^{-3}, and less otherwise. Marginal damages with log-log are more than double the marginal damages with log-linear for receptors with concentrations below 7.1 μg m^{-3}. In the 2005 InMAP baseline scenario, 39% of the U.S. population lived in receptors with a $\mathrm{PM}_{2.5}$ concentration greater than 14.2 μg m^{-3}, and only 9% lived in receptors where the concentration was below 7.1 μg m^{-3}. However, applying the estimate of $\mathrm{PM}_{2.5}$ concentrations in 2013 to InMAP, only 3% of the population is in an area with a concentration above 14.2 μg m^{-3}, and 20% below 7.1 μg m^{-3}. Therefore, the difference in marginal damages between log-linear and log-log is much larger in 2013 than in 2005.

For source j the combined marginal damages across all receptors, of its emissions of pollutant p from height h for any matrix of emission changes, $\Delta\mathbf{e}$ is

$$\mathrm{MD}_j^{p,h}(\Delta\mathbf{e}) = \sum_{i=1}^{N} \pi_{ji}^{p,h} \cdot \mathrm{MD}_i\big(C_i(\Delta\mathbf{e})\big), \tag{4.13}$$

[15]At the baseline concentrations, C_i^0, the relative-risk is unity, $\mathrm{RR}_i(C_i^0) \equiv 1$.

which is the sum of the marginal damages of increased concentrations in all downwind receptors weighted by the marginal impact from source j. It is important to distinguish between the marginal damages of increased $\mathrm{PM}_{2.5}$ concentrations in a receptor, $\mathrm{MD}_i(C_i(\Delta\mathbf{e}))$, and the marginal damages of increased emissions of a primary pollutant from a source, $\mathrm{MD}_j^{p,h}(\Delta\mathbf{e})$. The expression $\mathrm{MD}_i(C_i(\Delta\mathbf{e}))$ is useful in identifying locations where fine particulate concentrations are likely causing the greatest harm, but it does not provide guidance as to where emission reductions should be made to achieve desired concentration reductions. The marginal damage of emissions of pollutant p from source j, equation (4.13), is the externality value of emissions at the source, and is the primary focus of estimation in this analysis. This quantity provides more useful information for policy makers in determining the most effective locations to reduce (or increase) emissions to satisfy policy objectives. For instance, a Pigovian tax set equal to the value of equation (4.13) for each source and for each pollutant could potentially produce a socially optimal quantity of emissions of criteria pollutants (Goodkind *et al.* 2014).

4.2.4 Value of a statistical life

Perhaps the most consequential parameter contributing to the large marginal damages associated with criteria pollutant emissions is the value of a statistical life (VSL). Economic analyses of the wage-risk trade-off of working-age adults have consistently estimated large premiums to accept jobs that increase the risk of mortality (Viscusi and Aldy 2003). The VSL is derived using a hedonic-wage approach, estimating a wage equation that includes the risk of mortality for the worker's job, along with worker and job specific characteristics. The estimated value of the mortality risk parameter in the wage equation is the marginal wage premium required for a marginally riskier job. The average annual VSL for a sample of workers is derived from the value of this parameter after multiplying by the average wage and the annual hours worked (typically assumed to be 2000 hours).

Viscusi and Aldy (2003) compiled estimates of the VSL for workers in the U.S. and other countries. The median VSL was approximately \$7 million, with half of the studies ranging from \$5 to \$12 million. The studies with estimates below this range typically used data that could have suffered from self-selection bias, and the studies with estimates above this range typically used methods that did not directly measure the wage-risk tradeoff. The role of income, gender and age are all potentially important considerations in the estimation and application of the VSL.

The Environmental Protection Agency (EPA) recommends the use of a VSL of $7.4 million in 2006 U.S. Dollars (USD) for all benefit analysis involving risk of mortality for all age groups (EPA 2010). According to the estimated C-R in the ACS and H6C studies, mortality related to exposure to $PM_{2.5}$ pollution impacts older populations to a much larger degree than younger adults because baseline mortality rates are larger with older populations. This age dispariety raises the question of whether it is appropriate to assign the same VSL to people of all ages. If older people have lower VSLs than younger people then damages from criteria pollutant emissions may be overestimated through the use of a uniform VSL.

An alternative approach is the VSL-year method that assumes individuals have the same value of each additional year of life expectancy. The VSL-year method implies a VSL that is substantially different for younger and older populations. While there may be some intuitive appeal to the VSL-year methodology, it was rejected in empirical analysis (Aldy and Viscusi 2008). The findings in Aldy and Viscusi (2008) do not necessarily suggest a uniform VSL is the appropriate method, rather the VSL varies by age, but is not always declining with age, as is implied by the constant VSL-year approach. Estimates show that the VSL increases gradually until around age 46 and then slowly declines for older individuals. We apply an age-adjusted VSL along with a uniform VSL to understand the implications to the estimate of damages from criteria pollutant emissions under the two approaches.

The age-specific VSL used here is derived from the cohort-adjusted fitted VSL estimate from Aldy and Viscusi (2008) (see solid black line in Figure 4.6). The cohort-adjusted VSL is derived from the working age adult population and is therefore silent on the VSL for children and adults 63 years and older. For the construction of the age-specific VSL used in the present paper it is assumed that the VSL for all children is the same as for 18 year-olds. This assumption is inconsequential for the marginal damage calculation as the baseline mortality rate for young populations is much lower than for adults.

The age-specific VSL for the population above 62, however, greatly influences the

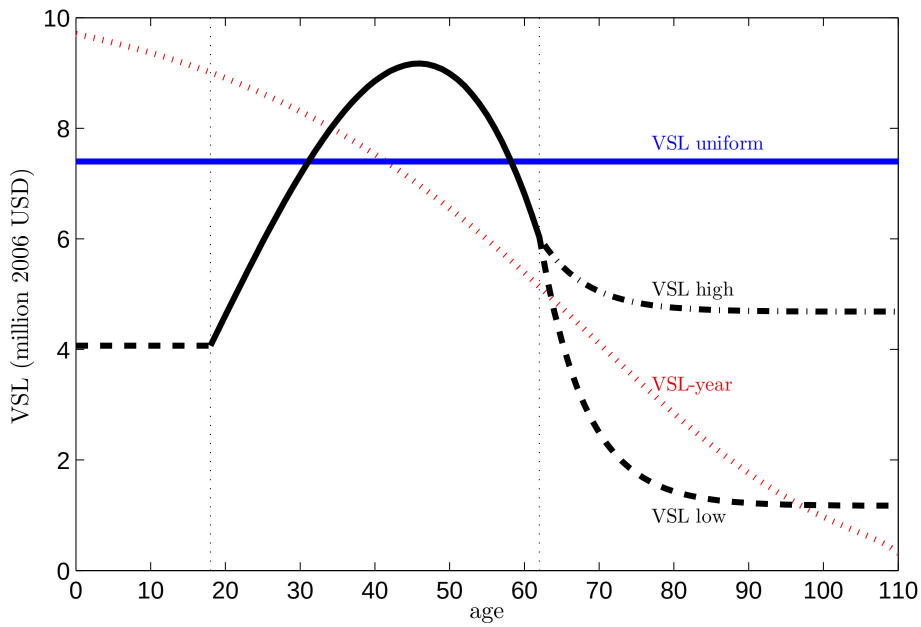


Figure 4.6: Age-specific estimates of VSL (million 2006 USD) used in marginal damage calculations: uniform across all ages (solid blue line) recommended by EPA (2010); cohort-adjusted VSL (solid black curve) from Aldy and Viscusi (2008) for ages 18 to 62; high and low VSL scenarios for older (>62) and younger (<18) populations (dashed black curves); implied VSL with constant VSL-year (dotted red curve).

overall marginal damage estimates, even though people 63 years and older constitute only 15% of the U.S. population. Without guidance from the VSL literature we apply two age-specific VSL scenarios for older individuals (see Figure 4.6). The first assumes a large drop in the VSL past age 62 and levels off around \$1 million for the oldest populations. This estimate is referred to as "VSL low". The second age-specific VSL scenario assumes a less dramatic drop-off of the VSL for older populations and levels of around \$4.6 million. This estimate is referred to as "VSL high". These VSL scenarios have no empirical backing, but are used to evaluate the impact of two distinct versions of the VSL for populations over 62 years old.

Finally, an additional estimate of marginal damages is conducted using a constant VSL-year approach to understand the implications to the impact of criteria pollutant emissions of this method. The constant VSL-year is calculated as \$312,000 of each additional year of life-expectancy. The VSL-year value is calculated so that the population weighted average of working age adults (age 18 to 62) of the implied VSL equals the uniform \$7.4 million VSL.[16] The VSL derived from the VSL-year method is shown in Figure 4.6 as the dotted red curve.

The age-specific VSLs ("high", "low", and constant VSL-year) are applied to the marginal damage equation by separating the population by age and using age-specific mortality rates. Equations (4.7) and (4.8) become

$$\mathrm{MD}_i^{\mathrm{lin}}(C_i) = \sum_{\mathrm{age}} \mathcal{V}^{\mathrm{age}} \cdot \mathcal{M}_i^{\mathrm{age}}(C_i^0) \cdot \gamma^{\mathrm{lin}} \cdot \mathrm{RR}_i^{\mathrm{lin}}(C_i) \tag{4.14}$$

$$\mathrm{MD}_i^{\mathrm{log}}(C_i) = \sum_{\mathrm{age}} \mathcal{V}^{\mathrm{age}} \cdot \mathcal{M}_i^{\mathrm{age}}(C_i^0) \cdot \frac{\gamma^{\mathrm{log}}}{C_i} \cdot \mathrm{RR}_i^{\mathrm{log}}(C_i). \tag{4.15}$$

where the age-specific mortality is $\mathcal{M}_i^{\mathrm{age}}(C_i^0) = \mathrm{Pop}_i^{\mathrm{age}} \cdot \lambda_i^{0,\mathrm{age}}$.

4.3 Results

We first describe and present the results of a recent benefit per ton estimate of emission reductions using a complex CTM without source-specificity, and a damage per ton estimate from a reduced form model with unique estimates from hundreds of sources. We follow with the marginal damage estimates from InMAP for the selected sources, and compare the results with past estimates.

4.3.1 Results from the literature

In the regulatory impact analysis for the proposed carbon pollution guidelines for existing power plants (EPA 2014b), benefits per ton of criteria pollutant emission reductions as a result of the regulation were estimated. The estimates were produced using the Comprehensive Air Quality Model with Extensions (CAMx), a complex CTM with 12 km × 12 km grid cells. Due to the computational requirements of running CAMx, the model was run once for electric-generating-unit (EGU) emissions of primary $\mathrm{PM}_{2.5}$, SO_X and NO_X from three regions of the Contiguous U.S.: East (all states east of Montana), West and California. Each run simulated a scenario of emission changes across the

[16] VSL calculation with constant VSL-year assumes a 3% discount rate. The curvature of the implied VSL across ages, in Figure 4.6, is the result of discounting and differences in expected mortality.

region and calculated the impacts to human health in each receptor across the country. Total benefits of improved health were divided by total emission changes of a pollutant for the region to calculate benefits per ton. We focus on the benefits per ton from the eastern U.S. as the East comprises 78% to 93% of the national EGU emissions of these pollutants. Estimated benefits per ton of emissions reductions in the East for year 2020 (in 2011 USD) are \$140,000 for $PM_{2.5}$, \$40,000 for SO_X and \$6,700 for NO_X (see light gray bars in Figure 4.7).[17]

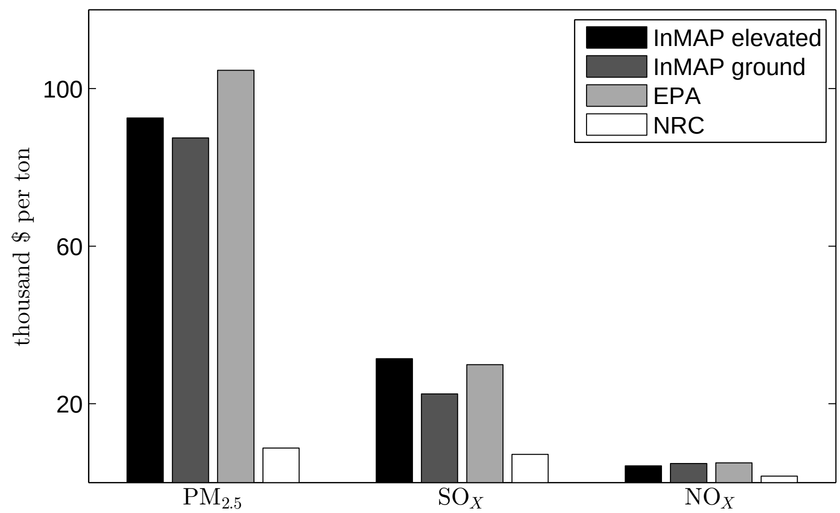


Figure 4.7: Comparison of estimates of median marginal damages with InMAP from elevated and ground level sources, benefits per ton of emission reductions in East region of U.S. (EPA 2014b), and median damages per ton from coal power plants (NRC). Values from EPA and the NRC are adjusted to have an equivalent VSL with InMAP.

The benefits per ton estimated in the EPA regulatory impact analysis (2014b) are much larger than the damages per ton estimated in a report from the National Resource Council (2010) on the externality costs of energy. The NRC report used the reduced form APEEP model to calculate the damages per ton from coal power plants in the U.S. Because APEEP can be run easily many times, damages per ton were estimated for each individual coal power plant. The median (and 5^{th} and 95^{th} percentile) damage per ton estimates of coal power plant emissions were \$7,100 (\$2,600, \$26,000) for primary $PM_{2.5}$,

[17]Estimates were calculated with a \$9.9 million (2011 USD) VSL for year 2020 income, and the Krewski *et al.* (2009) log-linear C-R function for mortality reductions. The benefits include impacts to mortality and morbidity, but mortality constitute 95% of the benefits.

\$5,800 (\$1,800, \$11,000) for SO_X, and \$1,300 (\$700, \$2,800) for NO_X (see white bars in Figure 4.7).[18]

The regulatory impact analysis (EPA 2014b) benefit per ton estimates are not directly comparable to the NRC report (2010) damage per ton estimates because the sources and emission changes are not exactly the same, but they are examining emissions from similar types of sources and locations and are seeking to answer a similar question: how damaging are emissions of criteria pollutants? Figure 4.7 shows that even after adjusting the two estimates for differences in VSL and dollar-year equivalency, the EPA regulatory impact analysis estimates are 3 to 11 times larger than the median estimates from the NRC. Without comparing identical locations and emission reductions it is not possible to know with certainty the reason for the divergence in estimates, but the choice of the air dispersion modeling is likely a key factor. Estimates in section 4.3.2 from InMAP are more similar in value with the EPA regulatory impact analysis (2014b) than with the estimates in the NRC report (2010). Because InMAP and the EPA RIA estimates are based on a stronger scientific representation of air dispersion modeling, and yield similar estimates of the magnitude of damages from emissions, the marginal damage (or damages per ton) estimates derived from APEEP may underestimate the true values.

4.3.2 Marginal damage estimates for selected sources using InMAP

For the 223 source grid cells for elevated emissions of primary $PM_{2.5}$ the median and mean marginal damage, across sources, is \$92,500 and \$201,300 per ton, respectively.[19] Emissions of SO_X and NO_X from elevated sources cause substantial damages, but are not as large per ton as primary $PM_{2.5}$. The median (and mean) marginal damages are \$31,400 (\$38,700) for SO_X emissions, and \$4,200 (\$5,400) for NO_X emissions from the selected elevated sources. Table 4.1 and Figure 4.8 displays the distribution of marginal damage estimates across sources for each pollutant (**A** - elevated sources, **B** - ground level sources). For emissions from ground-level sources, median (and mean) marginal

[18] Damages per ton were calculated with a \$6 million (2000 USD) VSL, and for mortality impacts used the Pope *et al.* (2002) log-linear C-R function from the ACS study (the Krewski *et al.* (2009) estimates are a subsequent estimate from the ACS study, and show approximately the same impact of fine particulate concentrations on mortality).

[19] If not otherwise stated marginal damages are calculated using equation (4.13) with a log-linear C-R and a uniform VSL of \$7.4 million (2006 USD).

damages per ton are \$87,500 (\$477,600) for $PM_{2.5}$, \$22,500 (\$34,400) for SO_X, \$4,800 (\$8,400) for NO_X, and \$48,200 (\$347,600) for NH_3. Mean marginal damages are much larger than median marginal damages for $PM_{2.5}$ and NH_3 emissions because of very large impacts that are experienced very near the emission source in the most densly populated areas.

Figure 4.8 suggests that the difference in marginal damages across pollutants and across source locations is larger than the difference between elevated and ground-level emissions. For a given pollutant, emissions cause dramatically different damages depending on the location of the source. For elevated emission sources the 95^{th} percentile marginal damage source is 38 times more damaging than the 5^{th} percentile source for $PM_{2.5}$. For both SO_X and NO_X emissions, marginal damages from the 95^{th} percentile elevated source are approximately 11 times greater than the 5^{th} percentile elevated source. Even larger ratios exist between the most and least damaging ground-level sources.

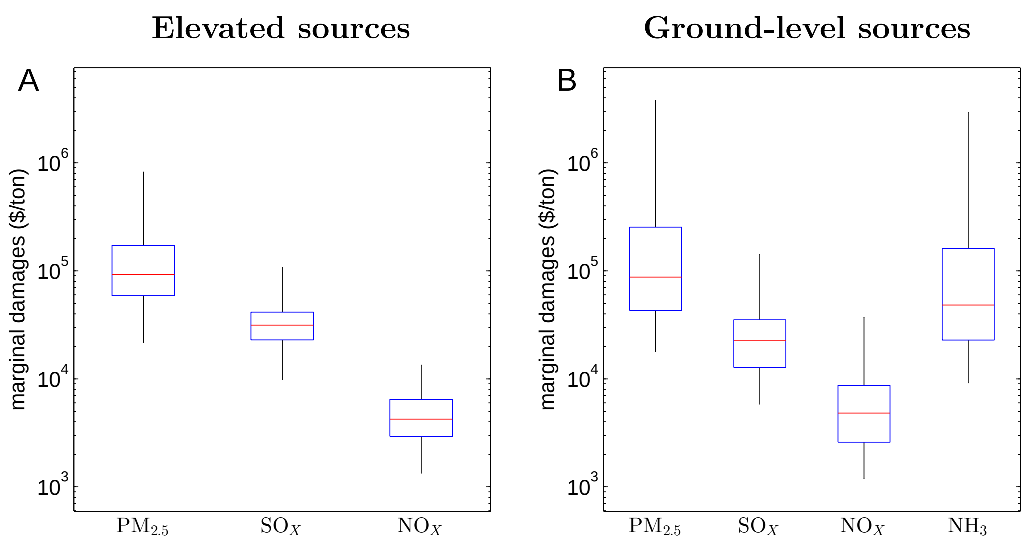


Figure 4.8: Distribution of estimated marginal damages of emissions from elevated sources (**A**) and ground level sources (**B**). Center line is median, top and bottom of box is the 25^{th} and 75^{th} percentile, and the vertical lines extend to the 5^{th} and 95^{th} percentiles.

Table 4.1: InMAP marginal damages from selected sources (thousand $ per ton)

		Percentile						
Pollutant	mean	1st	5th	25th	50th	75th	95th	99th
Elevated sources:								
$PM_{2.5}$	201.3	11.3	21.6	58.9	92.5	172.1	823.0	2418.1
SO_X	38.7	6.5	9.9	23.0	31.4	41.4	107.5	181.4
NO_X	5.4	0.6	1.3	2.9	4.2	6.5	13.4	23.8
Ground sources:								
$PM_{2.5}$	477.6	8.6	17.9	42.9	87.5	253.5	3795.5	4407.6
SO_X	34.4	3.4	5.8	12.7	22.5	35.2	142.6	153.8
NO_X	8.4	0.7	1.2	2.6	4.8	8.7	37.2	43.8
NH_3	347.6	5.3	9.2	22.8	48.2	161.3	2928.1	3390.3

Table 4.2: Marginal damages (thousand $ per ton) and population (millions) by distance of receptors from source (km)

	0 - 100		100 - 250		250 - 500		500 - 1000		>1000	
	mean	med	mean	med	mean	med	mean	med	mean	med
Elevated sources:										
$PM_{2.5}$	164.5	50.1	12.7	9.9	11.9	8.6	8.2	6.1	4.0	1.4
SO_X	16.1	5.3	6.5	4.7	7.6	5.5	5.6	3.9	2.8	0.9
NO_X	3.4	1.9	0.7	0.7	0.6	0.6	0.5	0.5	0.3	0.1
Pop	*3.8*	*1.8*	*8.4*	*7.9*	*25.7*	*26.9*	*64.4*	*67.5*	*177.2*	*168.4*
Ground sources:										
$PM_{2.5}$	452.4	55.6	9.5	5.7	6.3	4.4	5.0	3.2	4.4	3.5
SO_X	19.4	3.5	4.7	2.4	3.8	2.5	3.3	2.0	3.1	2.6
NO_X	6.9	2.7	0.5	0.4	0.4	0.2	0.3	0.2	0.4	0.4
NH_3	330.4	23.4	6.0	2.5	4.0	2.3	3.6	1.8	3.5	2.6
Pop	*3.7*	*1.5*	*6.9*	*4.5*	*17.3*	*16.7*	*47.7*	*41.5*	*204.0*	*209.9*

The results in Table 4.2 show that criteria pollutants cause substantial damages far from the source of emission. For the elevated sources selected in this analysis the mean marginal damages per ton to receptors more than 100 km from the respective source are \$32,800 for $PM_{2.5}$ emissions, \$19,800 for SO_X emissions, and \$1,700 for NO_X emissions. Table 4.2 shows that at greater distances from the source, marginal damages remain large because there is a substantially larger population exposed. For ground-level sources the mean marginal damages per ton to receptors more than 100 km from the source are \$20,800 for $PM_{2.5}$ emissions, \$11,800 for SO_X emissions, \$1,200 for NO_X emissions, and \$13,600 for NH_3 emisisons.

Table 4.3: Percent of marginal damages attributed to impacts in receptors less than 100 km from emission source and more than 100 km from emission source, averaged across sources of each grid cell size.

distance from source:	0 - 100 km				>100 km			
grid cell size:	1 km	4 km	12 km	36 km	1 km	4 km	12 km	36 km
Elevated sources:								
$PM_{2.5}$	97.0	71.9	49.4	24.4	3.0	28.1	50.6	75.6
SO_X	81.8	30.2	26.0	12.2	18.2	70.0	74.0	87.9
NO_X	93.5	63.3	45.5	25.0	6.5	36.7	54.5	75.0
Ground sources:								
$PM_{2.5}$	97.9	77.5	48.0	16.7	2.1	22.5	52.0	83.3
SO_X	79.2	26.0	22.1	7.4	20.8	74.0	77.9	92.6
NO_X	93.3	69.2	51.2	22.8	6.7	30.8	48.8	77.3
NH_3	97.7	71.3	40.2	12.2	2.3	28.7	59.8	87.8

The marginal damages of emissions from a source are strongly associated with the population density of the source grid cell. Table 4.3 shows the percentage of the marginal damages for a source that are experienced in receptors less than 100 km from the source. For sources in the smallest grid cells, generally in areas with higher population density, nearly all of the damages of emissions are suffered by people living near the source. In sparcely populated areas, a larger share of the marginal damages impact those living far

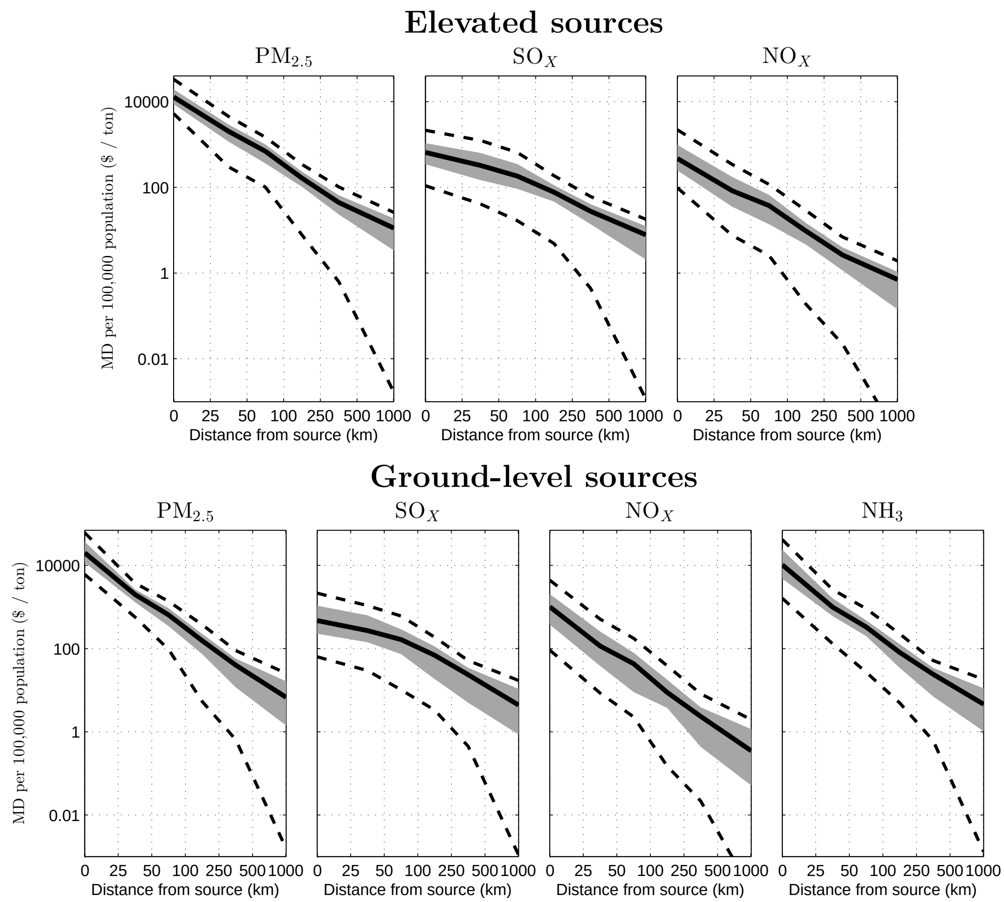


Figure 4.9: Distribution of marginal damages based on distance of receptors from the source of emission: elevated sources (top panel), ground-level sources (bottom panel). Values are $ per ton per 100,000 population. Dashed lines represent the 5$^{\mathrm{th}}$ and 95$^{\mathrm{th}}$ percentiles. The gray area represents the range from the 25$^{\mathrm{th}}$ to 75$^{\mathrm{th}}$ percentiles. The black line is the median.

from the source. Table 4.3 does not imply that marginal damages to far away receptors are larger from 36 km grid cells than from 1 km grid cells, rather the total marginal damages are much larger from 1 km grid cells and the share of the total impacting distant receptors is smaller.

The relationship between marginal damages and distance of receptors from the source, adjusted for the size of the population exposed, is demonstrated in Figure 4.9. In this figure the distribution of marginal damages per 100,000 population in receptors at various distances from the sources are displayed for the selected sources. The

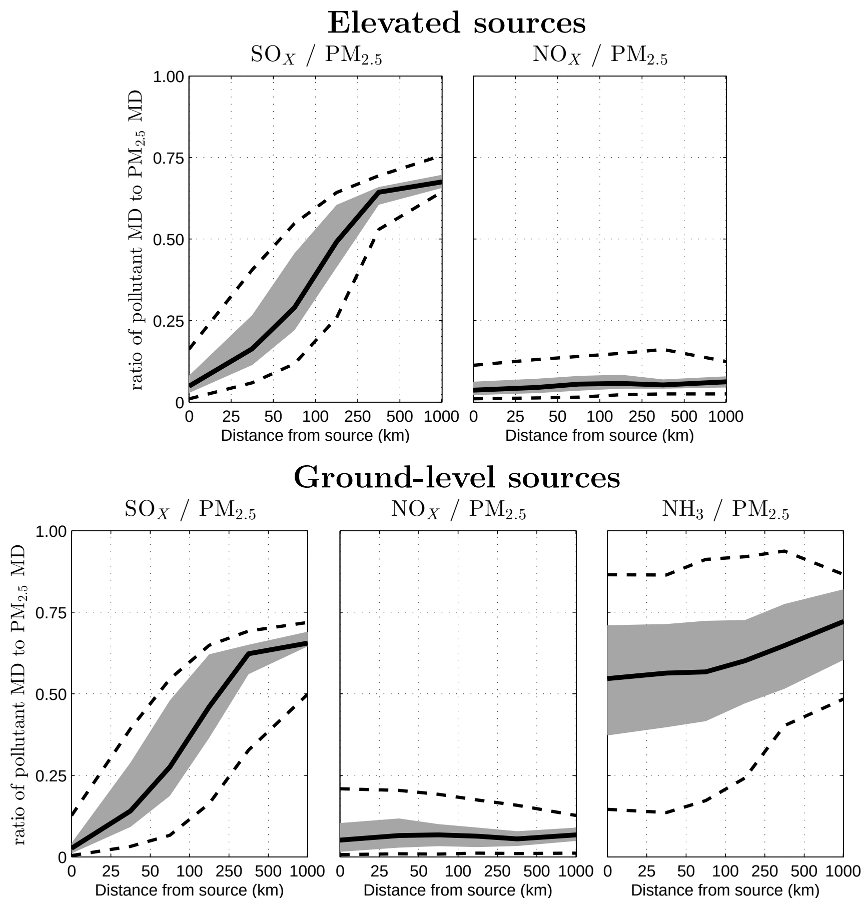


Figure 4.10: Distribution of the ratio of marginal damages between each pollutant and $PM_{2.5}$, based on distance of receptors from the source of emission: elevated sources (top panel), ground-level sources (bottom panel). Dashed lines represent the 5^{th} and 95^{th} percentiles. The gray area represents the range from the 25^{th} to 75^{th} percentiles. The black line is the median.

marginal damages per person decrease dramatically with distance, however, there are much larger populations exposed at greater distances.

For a given location, the rate at which primary $PM_{2.5}$ and SO_X emissions contribute to total fine particulate concentrations near the source helps explain the large difference in estimated marginal damages between these two pollutants. SO_X emissions require chemical reactions to form sulfate which can then form into fine particulate chemical compounds. These reactions take time, and so limit the impact of SO_X emissions on total $PM_{2.5}$ concentrations near the source. This relationship is shown in Figure 4.10

where the ratio of marginal damages between each pollutant and $PM_{2.5}$ is plotted for receptors various distances from the source of emission. For receptors less than 50 km from the elevated sources of emissions, marginal damages from SO_X are, on average, 10% of the marginal damages from $PM_{2.5}$. This ratio increases sharply at greater distances from the source. The ratio of marginal damages between SO_X and $PM_{2.5}$ is 0.42 for receptors 50 to 250 km from the elevated sources, and 0.65 for receptors 250 to 1000 km from the elevated sources. Figure 4.10 shows that a similar relationship between SO_X and $PM_{2.5}$ marginal damages is found for ground-level sources. Compared to sources that emit primary $PM_{2.5}$, it is less important to locate sources of SO_X emissions far from population centers.

Figure 4.10 also shows that the ratio of marginal damages between NO_X and $PM_{2.5}$ remains approximately the same for any distance between receptors and sources. For NH_3 emissions, the marginal damages are more similar in magnitude to $PM_{2.5}$, and the ratio does not change much based on distance from the source.

The damages resulting from long range transport of the emissions is based largely on the size of the population east of the source. Figure 4.11 shows that at greater distances from the source, receptors east of the source are impacted to a much larger degree than populations north, south or west of the source. In Figure 4.11 the area of each circle repsents the per capita marginal damages of people living various distances and directions from a source. The values are the average across all elevated sources. Between 250 and 500 km from the source of emission, marginal damages per person are large for populations east of the source, and less, but still substantial, for populations north, south and west of the source. For populations 500 to 1000 km from the emission source, marginal damages per person remain relatively large east of the source, but much smaller in the other directions. Note that while the marginal damages per person are less in the outer ring, there are, on average, larger populations in the outer ring than in the inner ring.

4.3.3 C-R functional form

The form of the C-R function can have large impacts on the marginal damages of emissions, but the high baseline $PM_{2.5}$ concentrations in 2005 limits large differences to mostly low population/low marginal damage areas. Figure 4.12 displays the distribution

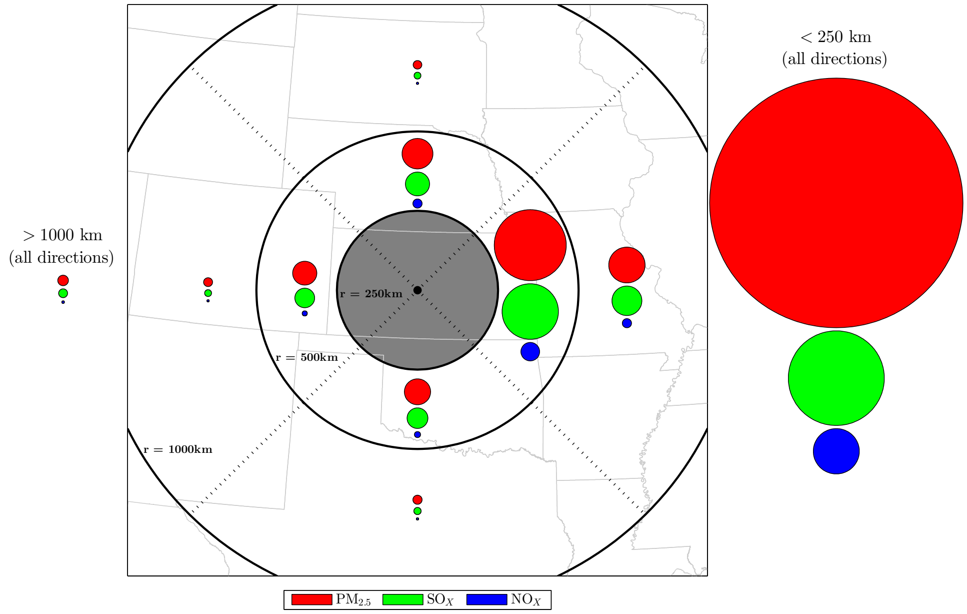


Figure 4.11: Marginal damages from elevated sources based on distance and direction of receptors from the source of emission. Area of circles represents the average, across elevated sources, of marginal damages per person living in the respective region. Black dot in center represents the source of emissions, but the data is not specific to this location. Gray state boundaries are included to provide scale for the distances from a source. Inner white ring represents people living 250 to 500 km from a source, and the direction (north, east, south or west) of the source. Outer white ring represents people living 500 to 1000 km from the source.

of the ratio of marginal damages with a log-log and a log-linear C-R for primary $PM_{2.5}$ emissions. For sources with a ratio above one, marginal damages are greater if calculated with a log-log C-R compared with the calculation made with a log-linear C-R. For the elevated emissions sources (panel **A**) in 2005 the median ratio is 0.96, suggesting that at the median, marginal damages are higher with log-linear. For ground level sources (panel **B**) in 2005 the median ratio is 1.10. For both types of sources the median differences between log-linear and log-log are unremarkable. The sources at the top of the distribution (where log-log marginal damages are substantially higher) tend to be the least damaging emissions because they are in the low population areas, suggesting that in 2005 the largest differences between log-linear and log-log occurred in locations whose emissions have the least impacts to human health. As Figure 4.5 showed, the dispariety between log-log and log-linear grows larger as $PM_{2.5}$ concentrations get smaller. Because

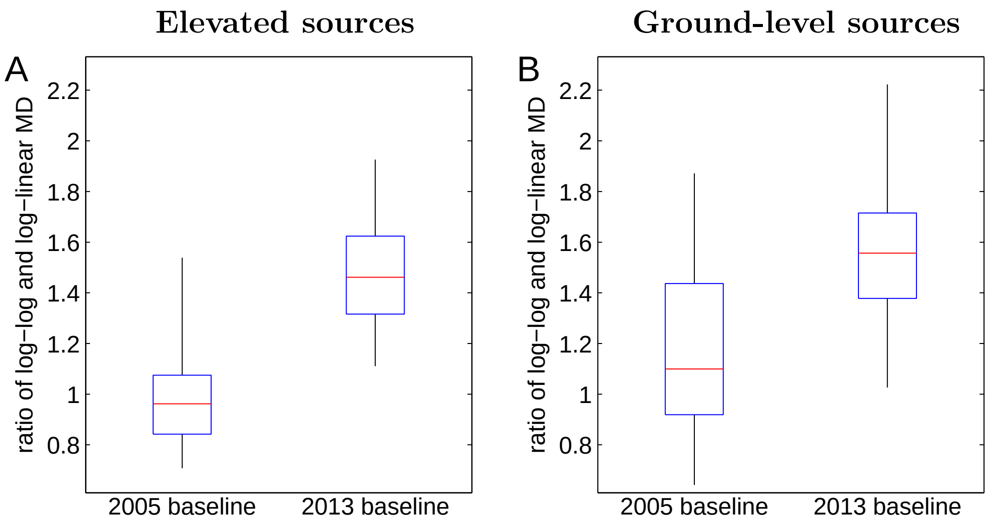


Figure 4.12: Distribution of ratio of marginal damages calculated with the log-log C-R and the log-linear C-R (with either the 2005 or the 2013 baseline $PM_{2.5}$ concentrations). Elevated sources (**A**) and ground level sources (**B**). Center line is median, top and bottom of box is the 25^{th} and 75^{th} percentile, and the vertical lines extend to the 5^{th} and 95^{th} percentiles.

the lowest concentrations tend to be found in the least populated areas the sources that contribute greatest to the low concentration receptors also tend to have lower marginal damages of emissions.

With 2013 baseline concentrations the difference between log-log and log-linear is greater and is relevant to higher population receptors. We assume that the marginal impacts on concentrations from emissions at the 2005 baseline concentrations are the same with 2013 $PM_{2.5}$ concentrations.[20] The median ratio of log-log to log-linear marginal damages is 1.46 for elevated sources of primary $PM_{2.5}$ emissions, and 1.56 for ground level sources (see Figure 4.12). With the lower fine particulate concentrations in 2013 only 1 (0.5%) of the elevated sources and 8 (1.1%) of the ground level sources have larger marginal damages with log-linear compared with log-log, while the marginal damages are 50% greater with log-log than with log-linear for 95 (42.6%) of the elevated sources and 414 (61.3%) of the ground level sources. Marginal damages with log-log are larger than log-linear for the other primary pollutants as well. The median ratio of marginal

[20]In the 2013 marginal damage equation the baseline mortality rates are derived from the 2005 mortality rates, the $PM_{2.5}$ concentration change between 2005 and 2013, and the log-linear or log-log C-R equation from Krewski *et al.* (2009).

damages between log-log and log-linear for 2013 baseline concentrations are 1.48 for SO_X and 1.51 for NO_X for elevated sources, and 1.58 for SO_X, 1.61 for NO_X, and 1.56 for NH_3 for ground-level sources.

The marginal damages with log-log are much greater for most sources at 2013 concentrations than at 2005 concentrations because lower concentrations cause the marginal damage curve to shift up for each source. This is illustrated from examining the largest emitting source of SO_X emissions from the 2011 NEI. The coal fired Duke Energy Beckjord power plant in New Richmond, OH emitted over 90,000 tons of SO_X in 2011. Figure 4.13A shows the marginal damage curve calculated with the log-linear C-R, for any level of emissions for the Beckjord plant from zero emissions to its current emissions, both with 2005 and 2013 baseline concentrations. The curves are nearly identical, therefore, the total damages of SO_X emissions from this plant (if it maintained the same quantity of emissions in both 2005 and 2011) are approximately the same. In 2005 total damages from SO_X emissions with a log-linear C-R are \$3.17 billion, and in 2013 total damages are \$3.06 billion. Figure 4.13B displays the marginal damage curves calculated with the log-log C-R from the same plant. The solid curve is with the higher 2005 baseline concentrations, and the dashed curve is with the lower 2013 baseline concentrations. The 2013 marginal damages have shifted up dramatically with the log-log C-R because the value of risk reductions is greater given the lower concentrations of $PM_{2.5}$ in 2013. If this plant maintained the same level of emissions between 2005 and 2013, and the C-R is log-log, the damages are much greater in 2013. In 2005 the estimated damages due to SO_X emissions for this plant are \$2.74 billion (less than with a log-linear C-R), but in 2013 the damages from this plant emitting the same quantity of SO_X are \$4.28 billion, much greater than with log-linear. For comparison, the total estimated revenue from electricity sales by the Beckjord plant in 2011 is \$0.45 billion (EIA 2014). The Beckjord plant is slated for retirement in 2015.

Figure 4.13 also shows that with log-log the marginal damage curve is decreasing in emissions, or increasing in abatement, but the slope is very flat. The flat slope is explained by the relatively small amount that any one plant can contribute to the total $PM_{2.5}$ concentration in a receptor, even when abating all emissions from the largest source of SO_X in the U.S. This is especially true with SO_X emissions that do not have as disproportionate an impact on local $PM_{2.5}$ conditions due to the time required to form

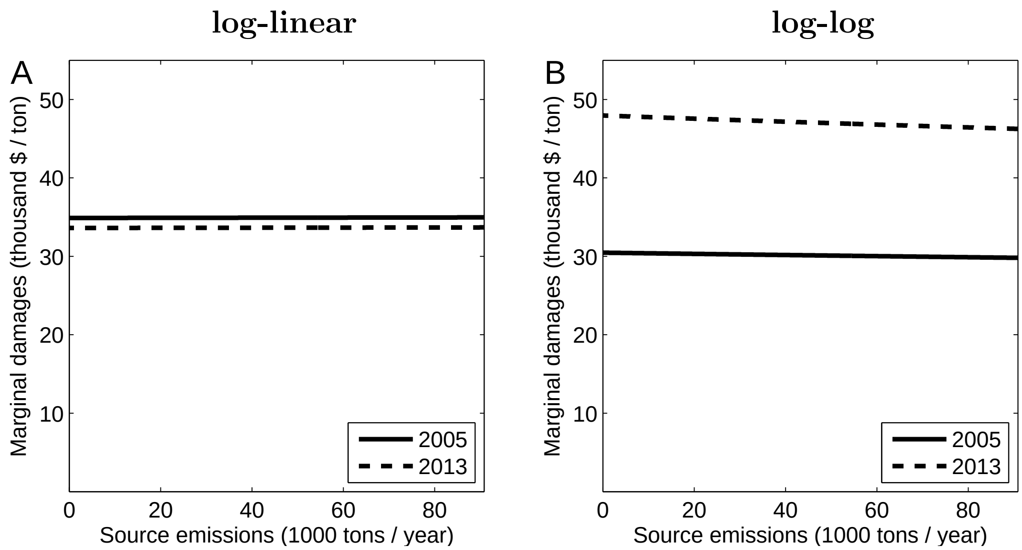


Figure 4.13: Marginal damage function for SO_X emissions from Duke Energy Beckjord power plant in New Richmond, OH, with 2005 and 2013 baseline $PM_{2.5}$ concentrations. **A** - log-linear C-R. **B** - log-log C-R.

sulfate particles. The large shift in the marginal damages in Figure 4.13B shows that with log-log, lower $PM_{2.5}$ concentrations provide the opportunity to make greater risk reductions from additional abatement from any source. In other words, concentration reductions, or emission abatement, from one location or source, are complements with emission reductions at other sources (Goodkind *et al.* 2014). Emission reductions at one source increase the marginal damages (or marginal benefits) of emissions (abatement) at nearby sources.

With a log-log C-R, emission reductions that impact low concentration receptors are associated with a relatively large reduction in the mortality rate compared to high concentration receptors. As $PM_{2.5}$ concentrations continue to decline, the importance of identifying the true shape of the C-R function gains prominence because of the large difference in the estimated mortality resulting from emissions between the use of a log-linear and log-log C-R.

4.3.4 Value of a statistical life

All estimates of the marginal damages of emissions thusfar have assumed a uniform VSL across all ages of \$7.4 million. For the analysis of damages from critieria pollutants a disproportional share of impacts are experienced by older people. Applying an age-specific VSL to the marginal damage calculation can substantially reduce the estimated burden of these air pollutants. Figure 4.14 shows the distribution of the ratio between marginal damages calculated with an age-specific VSL and marginal damages calculated with a uniform VSL. Using the "low" age-specific VSL scenario described in section 4.2.4 that assumes a substantially lower VSL for older populations, the marginal damages are 41 to 54% of the marginal damages with a uniform VSL. A similar relationship is true for all of the primary pollutants, at ground-level and elevated sources. The marginal damages with the "high" age-specific VSL scenario, which assumes the VSL for older populations decreases only slightly after age 62, are 73 to 80% of the marginal damages with a uniform VSL. The difference in the estimates between the "low" VSL and "high" VSL is attributed exclusively to the value assigned for mortality of individuals over age 62, as the two age-specific VSL scenarios are the same for everyone 62 and younger. The average value of the age-specific VSLs are slightly lower than the uniform VSL. For ages 18 and above the population-weighted average VSLs are \$6.6 million and \$7.1 million for the "low" and "high" age-specific VSL, respectively. The difference in average VSL explains only a small part of the large differences in marginal damages when calculated with a uniform VSL and an age-specific VSL. Using a constant value of an additional year of life expectancy, the marginal damages with the VSL-year method are 47 to 58% of the marginal damages with a uniform VSL.

With only 15% of the population over age 62, the differences between the marginal damages with the uniform VSL, the "high" age-specific VSL, the "low" age-specific VSL, and the constant VSL-year, highlights the relative weight that the mortality of older individuals has on the overall marginal damage calculation. These differences demonstrate the importance of identifying the appropriate VSL to apply to the non-working age population if age varying methods are to be used.

The current policy of the EPA is to apply a uniform VSL for individuals of any age when evaluating the benefits associated with mortality risk reductions (EPA 2010). This policy is not necessarily justified and may be driven by politics as much as science. In

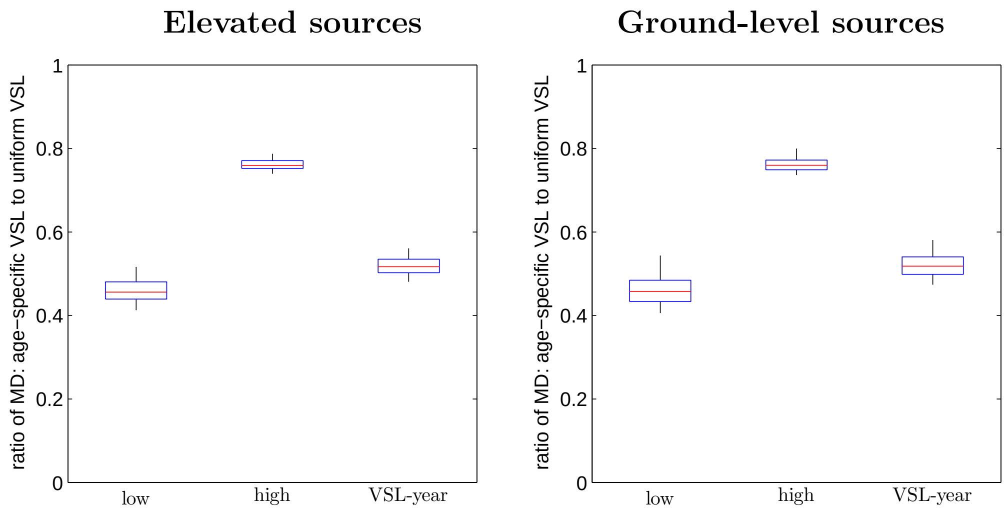


Figure 4.14: Distribution of ratio of marginal damages using an alternative age-specific VSL method and marginal damages using a uniform VSL. Center line is median, top and bottom of box is the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentile, and the vertical lines extend to the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles.

2003 the EPA set a lower VSL for older populations, but reversed the policy after public and Congressional outcry over the "senior death discount" (Banzhaf 2014). Given the large impacts of criteria air pollution on mortality and the disproportionate burden on older individuals, it is important to question if age should be a consideration in society's valuation of human life. An efficient air pollution policy, matching marginal damages of emissions and marginal costs of abatement at each source, would likely recommend substantially different emission reductions if derived from a uniform or age-specific VSL.

4.4 Conclusion

This paper attempts to identify and explain the key factors that affect the calculation of damages associated with emission of criteria pollutants. The choice of modeling of the impact of emissions on mortality in receptors is key in providing valid estimates of damages of air pollution. Estimates of damages using complex CTMs are unable to separate the impact from individual sources if many sources are of emissions are modeled. Reduced form models, such as APEEP, are not based on the state-of-the-science understanding of air dispersion, and may underestimate the damages from air pollution. InMAP, derived from a complex CTM, can be run to identify the impacts from individual sources, an important factor given the wide distribution of marginal damages depending on the location of the source.

The estimates from InMAP from the selected elevated and ground level sources presented here are a first step in developing a complete S-R matrix that can identify the impact of emissions from any source to any receptor in the U.S. With a complete S-R matrix source-specific pollution abatement policies can be evaluated and designed with prinicles of economic efficiency.

The form of the C-R function between mortality and $PM_{2.5}$ concentrations is an area that deserves more attention especially because the lower fine particulate concentrations in the U.S. in recent years amplifies the difference in outcomes between the log-log and log-linear C-R.

Chapter 5

Conclusion

The study of fine particulate air pollution is important for several reasons: the impacts are severe, everyone is exposed, and there are no safe levels. Understanding the effects of fine particulates and how to manage emissions is difficult because of the complicated relationships involved. Connecting emissions from a particular source with damages to any individual involves modeling the dispersion of emissions, chemical reactions in the atmosphere, human exposure to pollution and the resulting health impacts, and finally the valuation of a change in risks of human mortality. This dissertation sought to identify the key aspects of these steps that influence how we measure air pollution impacts and regulate emissions.

Within the U.S., damages from criteria pollutant emissions vary dramatically depending on the location of the source. The reliance on concentration standards for fine particulates limits the efficacy of air pollution regulations to be protective of human health, a key directive of the Clean Air Act. Our results suggest that better outcomes are possible by regulating emissions, rather than concentrations. In addition, policy instruments that apply source-specific prices equal to their marginal damages are preferred. Source-specific policies help incentivize eliminating the most damaging emissions, and prices encourage cost-effective abatement controls and the development of new abatement technologies that can decrease the burden of emission reductions for polluters.

Creating source-specific policies, such as emissions taxes set equal to the marginal damage of emissions, are only possible if the isolated impact of emissions on concentrations at each receptor can be identified. A source-receptor (S-R) matrix, identifying the impact of emissions at each location to every possible receptor, makes it possible to create a set of source-specific emission taxes. Currently available S-R matrices, such as the one utilized in APEEP (Muller and Mendelsohn 2007), are valuable but may underestimate the impact of emissions and lack the preferred geographic precision for sources of emissions and receptors of pollution.

The recently developed Intervention Model for Air Pollution (InMAP) provides an important advance in isolating the impact of emissions from specific sources. Even the most advanced air pollution models cannot perfectly represent the dispersion of air pollution and the chemical reactions governing the development of fine particulates. A trade-off exists between accuracy in representing the physical world and the time and computational resources required to run an air dispersion model. It becomes impractical to run models to isolate the impact of emissions from many individual sources with too much complexity. Using overly simplified models introduces doubt regarding the accuracy of the results. InMAP bridges the divide between complex chemical transport models, that require substantial computational resources, and reduced form models, that can be run very quickly. A goal for future research is the creation of an S-R matrix derived from InMAP. This would mark a substantial improvement over existing matrices. Not only would the S-R matrix make the analysis and development of source-specific regulations possible, it could be used identify the optimal location to build new emission producing plants, given various other cost and location constraints.

The economic theory and practical implementation of cost-effective source-specific regulations requires further analysis. The practical difficulties of source-specific policies include eliciting private information on abatement costs from emitters of pollution. One possible solution is to shift focus exclusively to describing the external damages from air pollution and setting source-specific taxes. Polluters, given appropriate incentives to internalize the external damages, could achieve a substantially improved outcome compared to the outcomes of command-and-control policies. This policy would allow regulators to ignore abatement costs, something they may not be adept at describing, and induce the creativity of industry to find innovative abatement solutions. One of the difficulties with a command-and-control policy is designing abatement strategies for technologies that do not yet exist, and have no financial incentive to be developed. Analysis of the potential advantages, drawbacks, and obstacles of implementing source-specific taxes to a real-world problem of air pollution is necessary.

Economists also need to understand the strategic incentives that would be provided to polluters from source-specific taxes. If taxes are designed to equal the marginal damages of emissions, a source could under- or over-abate emissions to influence the size of the tax. Analysis should focus on the potential inefficiencies that could result from strategic behavior compared with the advantages of providing polluters with better incentives to make cost-effective abatement decisions. For several air pollutants, it is possible that strategic behavior by emitters would not be problematic. Any individual source of pollution may have limited impact on the tax they face. If the marginal damage function for a source is relatively flat, then abatement decisions will have little influence on the tax. A flat marginal damage function is likely when the C-R is log-linear or when there are many sources of emissions and substantial dispersion of the pollutant.

This dissertation has also identified the shape of the concentration-response (C-R) as consequential for the type of regulation that should be pursued, the magnitude of the damages from fine particulate exposure, and which populations can benefit most from concentration reductions. If the C-R between fine particulates and mortality is log-log, then the benefits of reducing pollution to very low levels is much greater than has generally been believed. Pursuing regulation of air pollution across the entire range of exposure becomes essential with a log-log C-R.

Here we demonstrated the differences in policy outcomes and marginal damages between the log-linear and log-log C-R estimated from the American Cancer Society study by Krewski *et al.* (2009). This does not, however, provide guidance as to the true shape of the C-R. Future epidemiological studies should be designed to better evaluate how the C-R differs across concentration levels.

Our focus has been devoted to ambient concentrations of fine particulates in the U.S. A large share of the world's population is exposed to much higher concentrations. The question of the shape of the C-R is relevant here as well. A log-log C-R at high fine particulate concentrations suggests that large reductions in pollution may lead to only limited mortality risk reductions (Evans *et al.* 2013).

The most comprehensive longitudinal studies of the link between fine particulate exposure and mortality have been conducted in the U.S. and at relatively low concentrations. The shape of the C-R for high concentrations does not necessarily correspond to the shape at low concentrations, and extrapolating the results from these studies to concentrations outside the range observed introduces large uncertainty for the C-R in areas with high concentrations. Studies designed to identify the C-R for densely populated developing nations are difficult to create, but will provide valuable information for our understanding of the C-R response of fine particulates, and the risk reductions that are possible from cleaning the air in the dirtiest locations in the world.

There are positive and negative implications if the C-R is log-log over the range of high ambient concentrations. The good news is that, compared to low concentration areas, the mortality attributable to fine particulate air pollution in high concentration areas is not dramatically higher. The bad news is that even large reductions in concentrations would have relatively small reductions in risk of mortality. Therefore, it may be difficult to justify, economically, the emission reductions required to achieve low enough concentrations where substanial gains in public health are realized.

The magnitude of the damages from $PM_{2.5}$, both globally and in the U.S., are suggestive that the fine particulate concentrations are far too high. Achieving an optimal level of fine particulates could potentially involve dramatic shifts in the global economy. Estimation of sectoral and industry level damages from fine particulates along with input-output analysis could provide guidance to the likely changes in the economy. A full accounting of the external costs of air pollution across the economy would give government officials helpful information regarding policy priorities and for making long-term investments. Emission taxes, if used to implement pollution control policies, would reap substantial revenues. The distributional impacts of taxes, and the appropriate redistribution of tax revenues are important questions for further study.

A central component of any analysis of fine particulate air pollution is the valuation of changes in mortality risk. It is insufficient to only model the change in mortality from a change in emissions, because we need to compare this with the value of products and activities that created the emissions. Identifying an appropriate value is difficult and controversial. The EPA currently uses a value of a statistical life (VSL) of $7.4 million, derived by combining estimates from 26 studies that individually ranged from $0.85 million to $19.8 million (EPA 2010). While the range of studies used by the EPA demonstrates the uncertainty in the appropriate value of the VSL, it is a necessary parameter to conduct cost-benefit analyses. Applying a uniform value of the VSL to people of all ages, as is the policy of the EPA, may also be unsatisfactory. It is important to ask difficult questions about how mortality risks should be valued and if age should be a consideration.

The valuation of air pollution reductions could alternatively be framed in terms of improved health for the entire population and an increase in life expectancy. Studies could potentially be designed to evaluate the value that people assign to additional years of life expectancy, years that would be realized at different points in the future depending on the age of the person. In addition, less air pollution could result in healthier lives for all remaining years; healthier lives that could increase the value of each year lived. Ideally, from either perspective (reduced mortality versus extended life and improved heath) one would hope to arrive at similar conclusions. A re-framing of the issue could help evaluate our current valuation methods.

Concentrations of fine particulates have decreased dramatically in the U.S. in the last several decades, yet significant risks remain. In parts of the developing world, fine particulate pollution is an alarming problem. Additional research is required in the science, epidemiology and economics to further our understanding so that appropriate controls can be implemented.

Bibliography

[1] Abt Associates Inc. **2000** The particulate-related health benefits of reducing power plant emissions.

[2] Adar, Zvi, James M. Griffin. **1976**. Uncertainty and the choice of pollution control instruments. *Journal of Environmental Economics and Management*, 3, 178–188.

[3] Aldy, Joseph E. and W. Kip Viscusi. **2008**. Adjusting the value of a statistical life for age and cohort effects. *The Review of Economics and Statistics*, 90 (3), 573–581.

[4] Ambrose, John A., and Rajat S. Barua. **2004**. The pathophysiology of cigarette smoking and cardiovascular disease. *Journal of the American College of Cardiology*, 43 (10), 1731–1737.

[5] Anderson, Soren T., Ramanan Laxminarayan, and Stephen W. Salant. **2012**. Diversify or focus? Spending to combat infectious diseases when budgets are tight. *Journal of Health Economics*, 31 (4), 658–675.

[6] Andreoni, James and Arik Levinson. **2001**. The simple analytics of the environmental Kuznets curve. *Journal of Public Economics*, 80 (2), 269–286.

[7] Antweiler, Werner. **2012**. Emission permit trading for air pollution hot spots. Unpublished manuscript, Sauder School of Business, University of British Columbia.

[8] Banzhaf, H. Spencer. **2014**. The Cold-war origins of the value of statistical life. *Journal of Economic Perspectives*, 28 (4), 213–226.

[9] Baumol, William J. and Wallace E. Oates. **1988**. The theory of environmental policy. Second edition, New York: Cambridge University Press.

[10] Bennett, Deborah H., Thomas E. McKone, John S. Evans, William W. Nazaroff, Manuele D. Margni, Olivier Jolliet, and Kirk R. Smith. **2002**. Defining intake fraction. *Environmental Science & Technology*, 36 (9), 206A–211A.

[11] Biglaiser, Gary, John K. Horowitz, and John Quiggin. **1995**. Dynamic pollution regulation. *Journal of Regulatory Economics*, 8 (1), 33-44.

[12] Birnbaum, Linda S. **2012**. Environmental chemicals: Evaluating low-dose effects. *Environmental Health Perspectives*, 120 (4), A143–A144.

[13] Boulier, Bryan L., Tejwant S. Datta, and Robert S. Goldfarb. **2007**. Vaccination externalities. *The B.E. Journal of Economic Analysis and Policy*, 7 (1), Article 23.

[14] Brulle, Robert J. and David N. Pellow. **2006**. Environmental justice: Human health and environmental inequality. *Annual Review of Public Health*, 27, 103–124.

[15] Caputo, Marcelo, Marcelo Gimenez and Miguel Schlamp. **2003**. Intercomparison of atmospheric dispersion models. *Atmospheric Environment*, 37 (18), 2435–2449.

[16] Clark, Lara P., Dylan B. Millet, and Julian D. Marshall. **2014**. National patterns in environmental injustice and inequality: Outdoor NO2 air pollution in the United States. *PLOS ONE*, 9 (4), e94431.

[17] Crouse, Dan L., Paul A. Peters, Aaron van Donkelaar, Mark S. Goldberg, Paul J. Villeneuve, Orly Brion, Saeeda Khan, Dominic Odwa Atari, Michael Jerrett, C. Arden Pope III, Michael Brauer, Jeffrey R. Brook, Randall V. Martin, David Stieb, and Richard T. Burnett. **2012**. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: A Canadian national-level cohort study. *Environmental Health Perspectives*, 120 (5), 708–714.

[18] Dockery, Douglas W., C. Arden Pope III, Xiping Xu, John D. Spengler, James H. Ware, Martha E. Fay, Benjamin G. Ferris, and Frank E. Speizer. **1993**. An association between air pollution and mortality in six U. S. cities. *New England Journal of Medicine*, 329 (24), 1753–1759.

[19] EC/R Incorporated. **2014**. The HEM-3 user's guide.

[20] Evans, Jessica, Aaron van Donkelaar, Randall V. Martin, Richard Burnett, Daniel G. Rainham, Nicholas J. Birkett, and Daniel Krewski. **2013**. Estimates of global mortality attributable to particulate air pollution using satellite imagery. *Environmental Research* 120, 33–42.

[21] Fann, Neal, Charles M. Fulcher, and Bryan J. Hubbell. **2009**. The influence of location, source, and emission type in estimates of the human health benefits of reducing a ton of air pollution. *Air Quality, Atmosphere & Health* 2 (3), 169–176.

[22] Fann, Neal, Amy D. Lamson, Susan C. Anenberg, Karen Wesson, David Risley, and Bryan J. Hubbell. **2012**. Estimating the national public health burden associated with exposure to ambient $PM_{2.5}$ and ozone. *Risk Analysis*, 32 (1), 81–95.

[23] Fischer, Carolyn, Ian W.H. Parry, and William A. Pizer. **2003**. Instrument choice for environmental protection when technological innovation is endogenous. *Journal of Environmental Economics and Management*, 45 (3), 523-545.

[24] Gersbach, Hans and Amihai Glazer. **1999**. Markets and regulatory hold-up problems. *Journal of Environmental Economics and Management*, 37 (2), 151–164.

[25] Goodkind, Andrew L., Jay S. Coggins, and Julian D. Marshall. **2014**. A spatial model of air pollution: The impact of the concentration-response function *The Journal of the Association of Environmental and Resource Economists*, 1 (4), 451–479.

[26] Hawley, William A., Penelope A. Phillips-Howard, Feiko O. Ter Kuile, Dianne J. Terlow, John M. Vulule, Maurice Ombok, Bernard L. Nahlen, John E. Gimnig, Simon K. Kariuki, Margarette S. Kolczak, and Allen W. Hightower. **2003**. Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western Kenya. *American Journal of Tropical Medicine and Hygiene*, 68 (4 suppl), 121–127.

[27] Henry, David D. III, Nicholas Z. Muller, and Robert O. Mendelsohn. **2011**. The social cost of trading: Measuring the increased damages from sulfur dioxide trading in the United States. *Journal of Policy Analysis and Management*, 30 (3), 598–612.

[28] Holmes, N.S. and L. Morawska. **2006**. A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available. *Atmospheric Environment*, 40 (30), 5902–5928.

[29] Hoyert, Donna L. and Jiaquan Xu. **2012**. Deaths: Preliminary data for 2011. *National Vital Statistics Reports*, 61 (6), 1–51.

[30] Institute for Health Metrics and Evaluation. **2014**. Data Visualizations.

Available: http://www.healthdata.org/results/data-visualizations

[31] Jaffe, Adam B., Richard G. Newell, and Robert N. Stavins. **2003**. Technological change and the environment. *In* K.-G. Maler and J. Vincent, eds., *Handbook of Environmental Economics, vol. 1*, Amsterdam: North-Holland.

[32] Karp, Larry and Jiangfeng Zhang. **2012**. Taxes versus quantities for a stock pollutant with endogenous abatement costs and asymmetric information. *Economic Theory*, 49 (2), 371–409.

[33] Krewski, Daniel, Michael Jerrett, Richard T. Burnett, Renjun Ma, Edward Hughes, Yuanli Shi, Michelle C. Turner, C. Arden Pope III, George Thurston, Eugenia E. Calle, and Michael J. Thun. **2009**. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. Boston: Health Effects Institute.

[34] Kwerel, Evan. **1977**. To tell the truth: Imperfect information and optimal pollution control. *The Review of Economic Studies*, 44 (3), 595–601.

[35] Laden, Francine, Joel Schwartz, Frank E. Speizer, and Douglas W. Dockery. **2006**. Reduction in fine particulate air pollution and mortality: extended follow-up of the Harvard Six Cities study. *American Journal of Respiratory and Critical Care Medicine*, 173 (6), 667–672.

[36] Lepeule, Johanna, Francine Laden, Douglas Dockery, and Joel Schwartz. **2012**. Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard Six Cities study from 1974 to 2009. *Environmental Health Perspectives*, 120 (7), 965–970.

[37] Levy, Jonathan I., Lisa K. Baxter, and Joel Schwartz. **2009**. Uncertainty and variability in health-related damages from coal-fired power plants in the United States. *Risk Analysis*, 29 (7), 1000–1014.

[38] Lim, Stephen S., Abraham D. Flaxman, Goodarz Danaei, Kenji Shibuya, Heather Adair-Rohani, Mohammad A. AlMazroa, Markus Amann, H. Ross Anderson, Kathryn G. Andrews, Martin Aryee, *et al.* **2012**. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters

in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380 (9859), 2224–2260.

[39] Marshall, Julian D. **2008**. Environmental equality: Air pollution exposures in California's south coast air basin. *Atmospheric Environment*, 42 (21), 5499–5503.

[40] Marshall, Julian D., Kathryn R. Zwor, and Nam P. Nguyen. **2014**. Prioritizing environmental justice and equality: Diesel emissions in southern California. *Environmental Science & Technology*, 48 (7), 4063–4068.

[41] Masters, Gilbert M. and Wendell P. Ela. **2008**. Introduction to environmental engineering and science. Upper Saddle River: Prentice Hall.

[42] Milliman, Scott R. and Raymond Prince. **1989**. Firm incentives to promote technological change in pollution control. *Journal of Environmental Economics and Managment*, 17 (3), 247-265.

[43] Mohai, Paul, David Pellow, and J. Timmons Roberts. **2009**. Environmental justice. *Annual Review of Environment and Resources*, 34, 405–430.

[44] Moledina, Amyaz A., Jay S. Coggins, Stephen Polasky, and Christopher Costello. **2003**. Dynamic environmental policy with strategic firms: Prices *vs.* quantities. *Journal of Environmental Economics and Management*, 45 (2 suppl), 356-376.

[45] Montero, Juan-Pablo. **2002**. Permits, standards, and technology innovation. *Journal of Environmental Economics and Management*, 44 (1), 23-44.

[46] Muller, Nicholas Z. and Robert Mendelsohn. **2007**. Measuring the damages of air pollution in the United States. *Journal of Environmental Economics and Management*, 54 (1), 1–14.

[47] Muller, Nicholas Z. and Robert Mendelsohn. **2009**. Efficient pollution regulation: Getting the prices right. *American Economic Review*, 99 (5), 1714–1739.

[48] Muller, Nicholas Z. and Robert Mendelsohn. **2012**. Efficient pollution regulation: Getting the prices right: reply. *American Economic Review*, 102 (1), 608–612.

[49] National Research Council. **2010**. Hidden costs of energy: Unpriced consequences of energy production and use. Washington, D.C.: The National Academies Press.

[50] Office of Management and Budget. **2013**. 2012 report to congress on the benefits and costs of federal regulations and unfunded mandates on state, local, and tribal entities.

Available: http://www.whitehouse.gov/omb/inforeg_regpol_reports_congress.

[51] Ostro, Bart. **2004**. Outdoor air pollution: Assessing the environmental burden of disease at national and local levels. Geneva: World Health Organization.

[52] Pope, C. Arden III, Michael J. Thun, Mohan M. Namboodiri, Douglas W. Dockery, John S. Evans, Frank E. Speizer, and Clark W. Heath Jr. **1995**. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *American Journal of Respiratory and Critical Care Medicine*, 151 (3), 669–674.

[53] Pope, C. Arden III, Richard T. Burnett, Michael J. Thun, Eugenia E. Calle, Daniel Krewski, Kazuhiko Ito, George D. Thurston. **2002**. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *The Journal of the American Medican Association* 287 (9), 1132–1141.

[54] Pope, C. Arden III, and Douglas W. Dockery. **2006**. Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association*, 56 (6), 709–742.

[55] Pope, C. Arden III, Richard T. Burnett, Daniel Krewski, Michael Jerrett, Yuanli Shi, Eugenia E. Calle, Michael J. Thun. **2009a**. Cardiovascular mortality and exposure to airborne fine particulate matter and cigarette smoke: Shape of the exposure-response relationship. *Circulation*, 120 (11), 941-948.

[56] Pope, C. Arden III, Majid Ezzati and Douglas W. Dockery. **2009b**. Fine-particulate air pollution and life expectancy in the United States. *New England Journal of Medicine*, 360 (4), 376–386.

[57] Pope, C. Arden III, Richard T. Burnett, Michelle C. Turner, Aaron Cohen, Daniel Krewski, Michael Jerrett, Susan M. Gapstur, and Michael J. Thun. **2011**. Lung

cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: Shape of the exposure-response relationship. *Environmental Health Perspectives*, 119 (11), 1616–1621.

[58] Repetto, Robert. **1987**. The policy implications of non-convex environmental damages: A smog control case study. *Journal of Environmental Economics and Management*, 14 (1), 13–29.

[59] Requate, Till and Wolfram Unold. **2003**. Environmental policy incentives to adopt advanced abatement technology: Will the true ranking please stand up? *European Economic Review*, 47 (1), 125–146.

[60] Rothenberg, Stephen J. and Jesse C. Rothenberg. **2005**. Testing the dose-response specification in epidemiology: Public health and policy consequences for lead. *Environmental Health Perspectives*, 113 (9), 1190–1195.

[61] Schwartz, Joel. **2011**. A spline for the time. *Thorax*, 66 (10), 841–842.

[62] Smith, Kirk R. and Jennifer L. Peel. **2010**. Mind the gap. *Environmental Health Perspectives*, 118 (12), 1643–1645.

[63] Stayner Leslie, Kyle Steenland, Mustafa Dosemeci, and Irva Hertz-Picciotto. **2003**. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. *Scandanavian Journal of Work, Environment, and Health*, 29 (4), 317–324.

[64] Steenland, Kyle, Ryan Seals, Mitch Klein, Jennifer Jinot, and Henry D. Kahn. **2011**. Risk estimation with epidemiologic data when response attenuates at high-exposure levels. *Environmental Health Perspectives*, 119 (6), 831–837.

[65] Su, Jason G., Rachel Morello-Frosch, Bill M. Jesdale, Amy D. Kyle, Bhavna Shamasunder, and Michael Jerrett. **2009**. An index for assessing demographic inequalities in cumulative environmental hazards with applications to Los Angeles, California. *Environmental Science & Technology*, 43 (20), 7626–7634.

[66] Tarui, Nori and Stephen Polasky. **2005**. Environmental regulation with technology adoption, learning and strategic behavior. *Journal of Environmental Economics and Management*, 50 (3), 447–467.

[67] Tessum, Chris. **2014**. InMAP: A new model for air pollution interventions. *Chapter 1: Ph.D. Dissertation* (unpublished).

[68] U.S. Energy Information Administration. **2014**. Electricity data browser.

Available: http://www.eia.gov/electricity/data/browser/. Obtained: Nov. 6 2014.

[69] U.S. Environmental Protection Agency. **2004**. The particle pollution report: Current understanding of air quality and emissions through 2003. EPA-454-R-04-002, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

[70] U.S. Environmental Protection Agency. **2008**. Final ozone NAAQS regulatory impact analysis. EPA-HQ-OAR-2007-0225-0284, Office of Air and Radiation, Research Triangle Park, NC.

[71] U.S. Environmental Protection Agency. **2010**. Guidelines for preparing economic analyses. National Center for Environmenal Economics, Office of Policy.

[72] U.S. Environmental Protection Agency. **2011**. Regulatory impact analysis for the final mercury and air toxics standards. EPA-452/R-11-011, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

[73] U.S. Environmental Protection Agency. **2012a**. Our nation's air: Status and trends through 2010. EPA-454/R-12-001, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

[74] U.S. Environmental Protection Agency. **2012b**. Regulatory impact analysis for the final revisions to the National Ambient Air Quality Standards for particulate matter. EPA-452/R-12-003, Office of Air Quality Planning and Standards, Triangle Research Park, NC.

[75] U.S. Environmental Protection Agency. **2013a**. The 2011 national emissions inventory. Office of Air Quality Planning and Standards, Research Triangle Park, NC.

[76] U.S. Environmental Protection Agency. **2013b**. National trends in particulate matter levels.

Available: http://www.epa.gov/airtrends/pm.html.

[77] U.S. Environmental Protection Agency. **2013c**. National Emissions Inventory air pollutant emissions trends data.

Available: http://www.epa.gov/ttn/chief/trends/index.html.

[78] U.S. Environmental Protection Agency. **2014a**. User's manual for the Co-Benefits Risk Assessment (COBRA) screening model. Climate Protection Partnerships Division, State and Local Climate and Energy Programs.

[79] U.S. Environmental Protection Agency. **2014b**. Regulatory impact analysis for the proposed carbon pollution guidelines for existing power plants and emission standards for modified and reconstructed power plants. EPA-542/R-14-002, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

[80] U.S. Enviromental Protection Agency. **2014c**. AirData, Annual Summary Data.

Available: http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html.

[81] Viscusi, W. Kip and Joseph E. Aldy. **2003**. The value of a statistical life: A critical review of market estimates throughout the world. *The Journal of Risk and Uncertainty* 27 (1), 5–76.

[82] Weitzman, Martin L. **1974**. Prices *vs.* quantities. *Review of Economic Studies*, 41 (4), 477–491.

Appendix A

Chapter 2 Proofs

Appendix A: Chapter 2 Proofs

Proof of Proposition 1. For a given realization u, social welfare at the optimal quantity policy $q^* = (\alpha - \eta)/(\delta - \beta)$

$$\mathrm{SW}(q^*, u) = \frac{(\alpha-\eta)^2}{2(\delta-\beta)} - \frac{\alpha-\eta}{\delta-\beta}\, u.$$

Also for a given u, the social welfare achieved at the optimal abatement response $a^*(t,u)$ to any tax t is found in equation (2.9). Let $\tilde{t}$ be the price at which marginal benefits equal expected marginal costs. This is the optimal price policy for small ν, but not necessarily for large ν. We establish the proposition for $t = \tilde{t}$, noting that the advantage to the price policy grows larger still when the optimal tax in (2.23) is used. Let $\Delta(u)$ denote the advantage $\mathrm{SW}(\tilde{t}, u) - \mathrm{SW}(q^*, u)$ of this (possibly nonoptimal) price policy over the optimal quantity policy.

The support of u can be partitioned into three intervals as follows:

$$\begin{aligned}
U_1 &= \left\{ u : u > \frac{\delta(\alpha-\eta)}{\delta-\beta} \right\} \\
U_2 &= \left\{ u : \frac{\delta(\alpha-\eta-e^0(\delta-\beta))}{\delta-\beta} \le u \le \frac{\delta(\alpha-\eta)}{\delta-\beta} \right\} \\
U_3 &= \left\{ u : u < \frac{\delta(\alpha-\eta-e^0(\delta-\beta))}{\delta-\beta} \right\}.
\end{aligned}$$

The welfare advantage of the price policy for u in these three intervals is

$$\Delta_1(u) = \frac{\alpha-\eta}{2(\delta-\beta)}(\eta-\alpha+2u) > 0 \qquad \text{for } u \in U_1 \tag{A.1a}$$

$$\Delta_2(u) = \frac{u^2(\beta+\delta)}{2\delta^2} > 0 \qquad \text{for } u \in U_2 \text{ and } u \neq 0 \tag{A.1b}$$

$$\Delta_3(u) = \frac{\alpha-\eta-e^0(\delta-\beta)}{2(\delta-\beta)}\big(\eta-\alpha+e^0(\delta-\beta)+2u\big) > 0 \qquad \text{for } u \in U_3. \tag{A.1c}$$

At the endpoints of the U_j, the corresponding $\Delta_j(u)$ values are equal. If $\nu = 0$, clearly $E[\mathrm{SW}(\tilde{t}, u)] = E[\mathrm{SW}(q^*, u)]$. If, on the other hand, $\nu > 0$, equations (A.1a)–(A.1c) ensure that

$$E[\mathrm{SW}(\tilde{t},u)] - E[\mathrm{SW}(q^*,u)] = \int_{U_1} \Delta_1(u)\,du + \int_{U_2} \Delta_2(u)\,du + \int_{U_3} \Delta_3(u)\,du > 0.$$

To see that the difference is increasing in ν, note that for $u > 0$ the $\Delta_j(u)$ that is relevant at ν is increasing in u, and for $u < 0$ the $\Delta_j(u)$ that is relevant at $-\nu$ is increasing in $-u$. In either case we have that the advantage to the price policy increases in ν. This completes the proof. □

Proof of Proposition 2. Note that, from (2.6),

$$\tilde{t} = \eta + \delta\tilde{a} = \eta + \delta\left(\frac{\eta-\alpha}{\beta-\delta}\right).$$

To see that (*i.*) is true, consider (2.23) for the case with $\eta > \eta^*$. Either $\nu \in (\nu_{4A}, \nu_{1A})$ or $\nu \geq \nu_{1A}$. In the former case, $\hat{t}_{2b}$ is optimal and we must show that

$$\hat{t}_{2b} = \frac{\beta(\eta-\nu) - 2\alpha\delta}{\beta - 2\delta} > \eta + \delta\left(\frac{\eta-\alpha}{\beta-\delta}\right).$$

The inequality can be rearranged to yield

$$\nu > \frac{\delta(\eta-\alpha)}{\beta-\delta} = \nu_{4A},$$

which is true by assumption. In the latter case, $\hat{t}_1$ is optimal and we must show that

$$\hat{t}_1 = \alpha + \frac{\beta e^0}{2} > \eta + \delta\left(\frac{\eta-\alpha}{\beta-\delta}\right).$$

The inequality can be rearranged to yield $\eta > \eta^*$, which is again true by assumption.

To see that (*ii.*) is true, consider (2.23) for the case with $\eta < \eta^*$. Either $\nu \in (\nu_{4B}, \nu_{1B})$ or $\nu \geq \nu_{1B}$. In the former case, $\hat{t}_{3b}$ is optimal and we must show that

$$\hat{t}_{3b} = \frac{\beta(\eta+\nu) - \beta\delta e^0 - 2\alpha\delta}{\beta - 2\delta} < \eta + \delta\left(\frac{\eta-\alpha}{\beta-\delta}\right).$$

The inequality can be rearranged to yield

$$\nu > \frac{\delta(e^0(\beta - \delta) + \alpha - \eta)}{\beta - 2\delta} = \nu_{4B},$$

which is true by assumption. In the latter case, $\hat{t}_1$ is optimal and we must show that

$$\hat{t}_1 = \alpha + \frac{\beta e^0}{2} < \eta + \delta\left(\frac{\eta - \alpha}{\beta - \delta}\right).$$

The inequality can be rearranged to yield $\eta < \eta^*$, which is true by assumption. This completes the proof. $\square$

Proof of Proposition 3. Suppose that $\eta > \eta^*$, which means that $q^* = 0$ and that $E[\mathrm{SW}(q^*, u)] = 0$. This outcome can be acheived by the price-settting regulator who sets any $t \leq t_{\min}$, which yields $E[\mathrm{SW}(t_{\min}, u)] = 0$. If, instead, $\eta < \eta^*$, we know that $q^* = e^0$ and the welfare outcome is $E[\mathrm{SW}(q^*, u)]$. This outcome can be achieved by the price-setting regulator who sets $t \geq t_{\max}$, which ensures complete abatement. In both cases, a price policy is available at which $E[\mathrm{SW}(t^*(\nu;\theta), u)] \geq E[\mathrm{SW}(q^*(\nu;\theta), u)]$. This completes the proof. $\square$

Proof of Proposition 4. The proof of ($i.$) follows from the definition in (2.24) and the associated derivation.

To see ($ii.$), suppose that $\nu > E[\mathrm{MC}(0, u)] - \mathrm{MB}(0)$. Define

$$\nu_{1A} = \eta - \alpha - \frac{e^0}{2}(\beta - 2\delta)$$

which appears as equation (2.19). Note that $t^*(\nu, \theta) = \hat{t}_{2b} > t_{\min}$ when $\nu \in \big(E[\mathrm{MC}(o, u)] - \mathrm{MB}(0), \nu_{1A}\big)$ and that $t^*(\nu;\theta) = \hat{t}_1 > t_{\min}$ when $\nu \geq \nu_{1A}$. From (2.15) and (2.16) we know that $\hat{t}_1$ and $\hat{t}_{2b}$ are single-valued and continuous in ν on their feasible intervals. The only possible points of discontinuity are at $\nu = E[\mathrm{MC}(0, u)] - \mathrm{MB}(0)$ and at ν_{1A}. But $\lim_{\nu \nearrow (\eta - \alpha)} t_{\min} = \alpha$ and $\lim_{\nu \searrow (\eta - \alpha)} \hat{t}_{2b} = \alpha$, and so $t^*(\nu, \theta)$ is continuous at $\nu = E[\mathrm{MC}(0, u)] - \mathrm{MB}(0)$. Also,

$$\lim_{\nu \nearrow \nu_{1A}} \hat{t}_{2b} = \alpha + \frac{\beta e^0}{2} = \hat{t}_1.$$

We conclude that $t^*(\nu;\theta)$ is continuous and that $t^*(\nu,\theta) > t_{\min}$ whenever $\nu > E[\mathrm{MC}(0,u)] - \mathrm{MB}(0)$. What is more, we have shown that for $\nu \geq \nu_{1A}$, $t^* = \hat{t}_1$.

To see (*iii.*), note that $\eta > \eta^*$ implies that $q^*(\nu;\theta) = 0$ and so $E[\mathrm{SW}(q^*(\nu;\theta),u)] = 0$. Either $\nu \in (E[\mathrm{MC}(0,u)] - \mathrm{MB}(0), \nu_{1A})$, in which case $t^* = \hat{t}_{2b}$, or else $\nu \geq \nu_{1A}$, in which case $t^* = \hat{t}_1$. In the former case, the difference $E[\mathrm{SW}(t^*(\nu;\theta),u)] - E[\mathrm{SW}(q^*(\nu;\theta),u)]$ can be simplified to

$$E[\mathrm{SW}(t^*(\nu;\theta),u)] - 0 = \frac{\delta\big(\nu - (E[\mathrm{MC}(0,u)] - \mathrm{MB}(0))\big)^3}{3\nu(\beta - 2\delta)^2}, \tag{A.2}$$

which is strictly positive whenever $\nu > E[\mathrm{MC}(0,u)] - \mathrm{MB}(0)$. Moreover, the derivative of (A.2) with respect to ν is

$$\frac{\partial}{\partial \nu}\left[\frac{\delta\big(\nu - (E[\mathrm{MC}(0,u)] - \mathrm{MB}(0))\big)^3}{3\nu(\beta - 2\delta)^2}\right] = \frac{\delta\big(\nu - (E[\mathrm{MC}(0,u)] - \mathrm{MB}(0))\big)^2}{3\nu^2(\beta - 2\delta)^2}(2\nu - \alpha + \eta),$$

which is strictly positive because, under Assumption 2, we have $\eta > \alpha$.

In the latter case with $\nu \geq \nu_{1A}$,

$$E[\mathrm{SW}(t^*(\nu;\theta),u)] - 0 = \frac{e^0}{4\nu}\left[(\nu - \nu^*_{\min})\left(\nu - \nu^*_{\min} + \frac{e^0}{3}\sqrt{3\delta(2\beta - \delta)}\right)\right]. \tag{A.3}$$

We know that the right side of (A.3) is strictly positive whenever $\nu > \nu^*_{\min}$. To see that it is strictly positive whenever $\nu > \nu_{1A}$, note that $\nu_{1A} > \nu^*_{\min}$ so long as $\beta < 2\delta$, which is assumed to be true. The derivative of (A.3) with respect to ν is

$$\begin{aligned}\frac{\partial}{\partial \nu}\left(\frac{e^0}{4\nu}\left[(\nu - \nu^*_{\min})\left(\nu - \nu^*_{\min} + \frac{e^0}{3}\sqrt{3\delta(2\beta - \delta)}\right)\right]\right) &= \frac{\partial}{\partial \nu}\left[\frac{e^0}{4\nu}\left((\nu - \eta - \eta^*)^2 - \frac{\delta(e^0)^2}{12}(2\beta - \delta)\right)\right] \\ &= -\frac{e^0}{4\nu^2}\left((\nu - \eta + \eta^*)^2 - \frac{\delta(e^0)^2}{12}(2\beta - \delta)\right) + \frac{e^0}{2\nu}(\nu - \eta + \eta^*).\end{aligned}$$

The last expression is strictly positive whenever

$$(\nu - \eta + \eta^*)(\nu + \eta - \eta^*) > -\frac{\delta(e^0)^2}{12}(2\beta - \delta),$$

which is true for $\nu > \nu^*_{\min}$. This completes the proof. $\square$

Proof of Proposition 5. The proof of $(i.)$ follows from the definition in (2.24) and the associated derivation.

To see $(ii.)$, suppose that $\nu > \mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)]$. Define

$$\nu_{1B} = \frac{\beta e^0}{2} + \alpha - \eta$$

which appears as equation (2.20). Note that $t^*(\nu, \theta) = \hat{t}_{3b} < t_{\max}$ when $\nu \in \big(\mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)], \nu_{1B}\big)$ and that $t^*(\nu; \theta) = \hat{t}_1 < t_{\max}$ when $\nu \geq \nu_{1B}$. From (2.15) and (2.17) we know that $\hat{t}_1$ and $\hat{t}_{3b}$ are single-valued and continuous in ν on their feasible intervals. The only possible points of discontinuity are at $\nu = \mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)]$ and at ν_{1B}. But $\lim_{\nu \nearrow (\mathrm{MB}(e^0) - E[\mathrm{MC}(e^0,u)])} t_{\max} = \alpha + \beta e^0$ and $\lim_{\nu \searrow (\mathrm{MB}(e^0) - E[\mathrm{MC}(e^0,u)])} \hat{t}_{3b} = \alpha + \beta e^0$, and so $t^*(\nu, \theta)$ is continuous at $\nu = \mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)]$. Also,

$$\lim_{\nu \nearrow \nu_{1B}} \hat{t}_{3b} = \alpha + \frac{\beta e^0}{2} = \hat{t}_1.$$

We conclude that $t^*(\nu; \theta)$ is continuous and that $t^*(\nu, \theta) < t_{\max}$ whenever $\nu > \mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)]$. What is more, we have shown that for $\nu \geq \nu_{1B}$, $t^* = \hat{t}_1$.

To see $(iii.)$, consider two cases: either $\nu \in \big(\mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)], \nu_{1B}\big)$, in which case $t^* = \hat{t}_{3b}$, or else $\nu \geq \nu_{1B}$, in which case $t^* = \hat{t}_1$. In the former case, we can write

$$E[\mathrm{SW}(t^*(\nu; \theta), u)] - E[\mathrm{SW}(q^*(\nu, \theta), u)] = \frac{\delta\big(\nu - (\mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)])\big)^3}{3\nu(\beta - 2\delta)^2}, \tag{A.4}$$

which is positive whenever $\nu > \mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)]$. The derivative of (A.4) with respect to ν is

$$\frac{\partial}{\partial \nu}\left[\frac{\delta\big(\nu - (\mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)])\big)^3}{3\nu(\beta - 2\delta)^2}\right] = \frac{\delta\big(\nu - (\mathrm{MB}(e^0) - E[\mathrm{MC}(e^0, u)])\big)^2}{3\nu^2(\beta - 2\delta)^2}\,(2\nu + e^0(\beta - \delta) + \alpha - \eta),$$

which is strictly positive under Assumption 2.

In the latter case with $\nu \geq \nu_{1B}$, the argument follows that found in the proof of Proposition 4. Replace $\nu^*_{\min}$ with $\nu^*_{\max}$ in (A.3) and the same argument establishes the claim made here. This completes the proof. $\square$

Proof of Proposition 6. The proof of $(i.)$ follows from the definition in (2.27) and the associated derivation.

To see (*ii.*), note that the derivation of ν^*_{min} in (B.9) was achieved by setting $E[\text{SW}(t^*(\nu;\theta),u)] = E[\text{SW}(q^*(\nu;\theta),u)]$ and solving for ν. From (2.15) we know that $\hat{t}_1 = \alpha + (\beta e^0/2)$ and from (2.16) we know that $t_{\text{min}} = \text{MC}(0,-\nu)$. Insert ν^*_{min} into the latter and rearrange to find the difference

$$\hat{t}_1 - t_{\text{min}} = \frac{e^0\delta}{2} + \frac{e^0}{6}\sqrt{3\delta(2\beta-\delta)} > 0.$$

We show that $t^*(\nu;\theta)$ is not lower hemicontinuous. Define the sequence (ν^m) with element $\nu^*_{\text{min}} + (1/m)$ and specify $t_{\text{min}} \in t^*(\nu^*_{\text{min}};\theta)$. No sequence $(t^m) \in \mathbb{R}_+$ exists with $t^m \to t_{\text{min}}$ and $t^m \in t^*(\nu;\theta)$ for each m.

To see (*iii.*), note that $E_u[a^* \mid \nu < \nu^*_{\text{min}}] = 0$. For $\nu > \nu^*_{\text{min}}$, we have

$$\begin{aligned} E_u[a^*] &= \int_{-\nu}^{\hat{t}_1 - E[\text{MC}(e^0,u)]} e^0 f(u)\,du + \int_{\hat{t}_1 - E[\text{MC}(e^0,u)]}^{\hat{t}_1 - E[\text{MC}(o,u)]} \left(\frac{\hat{t}_1 - \eta - u}{\delta}\right) f(u)\,du \\ &= \frac{e^0}{2\nu}\left(\hat{t}_1 - \frac{\delta e^0}{2} - \eta + \nu\right). \end{aligned}$$

Insert the expression for $\hat{t}_1$ to get $E_u[a^*] = e^0(\eta^* - \eta + \nu)/2\nu$, which is strictly positive whenever $\nu > \eta - \eta^*$. But this inequality must be satisfied for $\nu > \nu^*_{min}$.

The proof of (*iv.*) follows closely that of Proposition 4(*iii.*) and so is omitted. This completes the proof. □

Proof of Proposition 7. The proof of (*i.*) follows from the definition in (2.27) and the associated derivation.

To see (*ii.*), note that the derivation of ν^*_{max} in (B.10) was achieved by setting $E[\text{SW}(t^*,u)] = E[\text{SW}(q^*,u)]$ and solving for ν. From (2.15) we know that $\hat{t}_1 = \alpha + (\beta e^0/2)$ and from (2.17) we know that $t_{\text{max}} = \text{MC}(e^0,\nu)$. Insert ν^*_{max} into the latter and rearrange to find the difference

$$t_{\text{max}} - \hat{t}_1 = \frac{e^0\delta}{2} + \frac{e^0}{6}\sqrt{3\delta(2\beta-\delta)} > 0.$$

We show that $t^*(\nu;\theta)$ is not lower hemicontinuous. Define the sequence (ν^m) with element $\nu^*_{\text{max}} + (1/m)$ and specify $t_{\text{max}} \in t^*(\nu^*_{\text{max}};\theta)$. No sequence $(t^m) \in \mathbb{R}_+$ exists with $t^m \to t_{\max}$ and $t^m \in t^*(\nu;\theta)$ for each m.

To see (*iii.*), note that $E_u[a^* \mid \nu < \nu^*_{\max}] = e^0$. For $\nu > \nu^*_{\max}$, we have

$$\begin{aligned} E_u[a^*] &= \int_{-\nu}^{\hat{t}_1 - E[\mathrm{MC}(e^0,u)]} e^0 f(u)\,du + \int_{\hat{t}_1 - E[\mathrm{MC}(e^0,u)]}^{\hat{t}_1 - E[\mathrm{MC}(0,u)]} \left(\frac{\hat{t}_1 - \eta - u}{\delta}\right) f(u)\,du \\ &= \frac{e^0}{2\nu}\left(\hat{t}_1 - \frac{\delta e^0}{2} - \eta + \nu\right). \end{aligned}$$

Insert the expression for $\hat{t}_1$ to get $E_u[a^*] = e^0(\eta^* - \eta + \nu)/2\nu$, which is strictly less than e^0 whenever $\nu < \eta - \eta^*$. But this last inequality must be satisfied for $\nu > \nu^*_{\max}$.

The proof of (*iv.*) follows closely that of Proposition 5(*iii.*) and so is omitted. This completes the proof. □

Appendix B

Chapter 2 Optimal Tax Rule Derivations

Appendix B: Chapter 2 Optimal Tax Rule Derivations

We start with the equations (2.11)–(2.14), with the interval of relevant taxes given by $T = [t_{\min}, t_{\max}]$ and with $t_{\min} = \eta - \nu$ and $t_{\max} = \eta + \nu + \delta e^0$. Equations (2.11)–(2.14) can be differentiated using Liebniz's rule, and the resulting first derivatives set equal to zero. (The derivatives themselves are quite lengthy and so are omitted.) Solving the first-order necessary conditions for the $\hat{t}_j$ yields equations (2.15)–(2.18).

The second-order sufficient conditions associated with the $\hat{t}_j$ are as follows. (Their derivation is somewhat tedious and is available upon request.)

$\hat{t}_1$: The second-order sufficient condition for $\hat{t}_1$ to be a local maximum is

$$\frac{\partial^2 \Gamma_1(\hat{t}_1)}{\partial t^2} = -\frac{e^0}{2\nu} < 0. \tag{B.1}$$

When $T_1 \neq \emptyset$, we can be sure that $\hat{t}_1$ is a local maximum.

$\hat{t}_4$: The second-order sufficient condition for $\hat{t}_4$ to be a local maximum is

$$\frac{\partial^2 \Gamma_4(\hat{t}_4)}{\partial t^2} = \frac{(\beta - \delta)}{\delta^2} < 0. \tag{B.2}$$

This condition is satisfied, and thus $\hat{t}_4$ is a local maximum, if and only if $\beta < \delta$. Whenever Assumption 1 is satisfied, if $T_4 \neq \emptyset$ we know that $\hat{t}_4$ is a local maximum. Otherwise (under Assumption 2), $\hat{t}_4$ is a local minimum.

$\hat{t}_{2a}$: Because $\hat{t}_{2a} = t_{\min}$, we know that $\hat{t}_{2a}$ is an endpoint tax level that must be checked against all possible local maxima. The second-order sufficient condition for $\hat{t}_{2a}$ to be a local maximum is

$$\frac{\partial^2 \Gamma_2(\hat{t}_{2a})}{\partial t^2} = \frac{-\eta + \alpha + \nu}{2\delta\nu} < 0. \tag{B.3}$$

This condition is satisfied, and thus $\hat{t}_{2a}$ is a local maximum, if and only if $\nu < \eta - \alpha$.

$\hat{t}_{2b}$: The second-order sufficient condition for a local maximum at $\hat{t}_{2b}$ is

$$\frac{\partial^2 \Gamma_2(\hat{t}_{2b})}{\partial t^2} = \frac{\eta - \alpha - \nu}{2\delta\nu} < 0. \tag{B.4}$$

This condition is satisfied, and thus $\hat{t}_{2b}$ is a local maximum, if and only if $\nu > \eta - \alpha$.

$\hat{t}_{3a}$: Note that $\hat{t}_{3a} = t_{\max}$, so that $\hat{t}_{3a}$ is another endpoint tax level that must be checked against all possible local maxima. The second-order sufficient condition for $\hat{t}_{3a}$ to be a local maximum is

$$\frac{\partial^2 \Gamma_3(\hat{t}_{3a})}{\partial t^2} = \frac{-\alpha - e^0(\beta - \delta) + (\eta + \nu)}{2\delta^2 \nu} < 0. \tag{B.5}$$

This condition is satisfied, and thus $\hat{t}_{3a}$ is a local maximum, if and only if $\nu < e^0(\beta - \delta) + \alpha - \eta$.

$\hat{t}_{3b}$: The second-order sufficient condition for $\hat{t}_{3b}$ to be a a local maximum is

$$\frac{\partial^2 \Gamma_3(\hat{t}_{3b})}{\partial t^2} = \frac{\alpha + e^0(\beta - \delta) - (\eta + \nu)}{2\delta^2 \nu} < 0. \tag{B.6}$$

This condition is satisfied, and thus $\hat{t}_{3b}$ is a local maximum, if and only if $\nu > e^0(\beta - \delta) + \alpha - \eta$.

Our purpose in the remainder of the appendix is to demonstrate that the optimal tax rules found in (reftstar1), (2.24), and (2.27) are correct. This requires checking all six possible maxima, ruling out those that are minima or infeasible in each circumstance, and then comparing the rest to each other to show which is the global maximum of the expected social welfare function. Remember that $E[\mathrm{SW}(t)]$ combines the four Γ_j functions, and these four functions are defined only over the domain of values that are part of the combined function. Therefore, it is possible for a potential local optimum to be outside the domain of its corresponding Γ_j function. This is referred to in the remainder of the Appendix as the local optimum being infeasible and therefore it cannot be the global maximum. It is also possible for the domain of a Γ_j function to be empty. In this case the local optimum is infeasible. Thus, for a local optimum to be feasible the corresponding Γ_j function must be non-empty and the local optimum must be an element of the Γ_j domain defined in (2.7a) - (2.7d).

B.1 Optimal tax rule for $\beta < \delta$

This is the situation in which Assumption 1 is satisfied. Under Assumption 1 we have

$$\eta < \alpha \quad \text{and} \quad \eta > e^0(\beta - \delta) + \alpha.$$

The six possible local optima are considered in turn. We show that one and only one of the first-order necessary conditions can simultaneously yield both (*i.*) a feasible outcome, and (*ii.*) a local maximum. The corresponding tax level must be the global maximum.

The first step for each is to determine whether it is a local maximum or not. If it is, we then derive the conditions for its feasibility. These are restrictions on the parameter vector that must be satisfied in order for the corresponding $\hat{t}_j$ to be the globally optimal tax. It might be helpful to refer to Figure 2.3 while reading this section.

$\hat{t}_1$: By equation (B.1) the second derivative of $\Gamma_1(t)$ with respect to t is strictly negative. Therefore, $\hat{t}_1$ is a local maximum. For it to be feasible $\hat{t}_1$ must be an element of T_1. From (2.7a) this implies that $\hat{t}_1 \in [\underline{t}_1, \bar{t}_1]$ which is equivalent to

$$\eta + \delta e^0 - \nu \leq \alpha + \frac{\beta e^0}{2} \leq \eta + \nu.$$

Notice that if $\hat{t}_1$ satisfies these conditions then it must also be true that T_1 is non-empty. The left inequality is equivalent to $\nu \geq \eta - \alpha - (e^0/2)(\beta - 2\delta)$ and the right inequality is equivalent to $\nu \geq \alpha - \eta + \beta e^0/2$. Combining these two expressions yields the sufficient condition for $\hat{t}_1$ feasibility:

$$\nu \geq \max\left\{\eta - \alpha - \frac{e^0}{2}(\beta - 2\delta), \alpha - \eta + \frac{\beta e^0}{2}\right\}.$$

Define

$$\nu_{1A} = \eta - \alpha - \frac{e^0}{2}(\beta - 2\delta) \quad \text{and} \quad \nu_{1B} = \alpha - \eta + \frac{\beta e^0}{2},$$

which appear as equations (2.19) and (2.20). We have that

$$[\nu_{1A} > \nu_{1B}] \text{ if and only if } [\eta > \eta^*].$$

We conclude that $\hat{t}_1$ is the optimum in each of the following cases:

1. $[\eta > \eta^*]$ and $[\nu \geq \nu_{1A}]$ or
2. $[\eta < \eta^*]$ and $[\nu \geq \nu_{1B}]$,

where η^* is given in (1).

$\hat{t}_4$: Under Assumption 1, where $\beta < \delta$, the inequality in (B.2) is always satisfied. Thus, $\hat{t}_4$ is a local maximum. For it to be feasible $\hat{t}_4$ must be an element of T_4. From (7d) this implies that $\hat{t}_4 \in (\underline{t}_4, \overline{t}_4)$ which is equivalent to

$$\eta + \nu < \frac{\beta\eta - \alpha\delta}{\beta - \delta} < \eta + \delta e^0 - \nu.$$

The left inequality is equivalent to $\nu < \delta(\eta - \alpha)/(\beta - \delta)$ and the right inequality is equivalent to $\nu < \delta[e^0(\beta - \delta) + \alpha - \eta]/(\beta - \delta)$. Combining these two expressions yields the sufficient condition for $\hat{t}_4$ feasibility:

$$\nu < \min\left\{\frac{\delta(\eta - \alpha)}{\beta - \delta}, \frac{\delta[e^0(\beta - \delta) + \alpha - \eta]}{\beta - \delta}\right\}.$$

Define

$$\nu_{4A} = \frac{\delta(\eta - \alpha)}{\beta - \delta} \quad \text{and} \quad \nu_{4B} = \frac{\delta[e^0(\beta - \delta) + \alpha - \eta]}{\beta - \delta}.$$

which appear as equations (2.21) and (2.22). We have that

$$[\nu_{4A} < \nu_{4B}] \text{ if and only if } [\eta > \eta^*].$$

We conclude that $\hat{t}_4$ is the optimum in each of the following cases:

1. $[\eta > \eta^*]$ and $[\nu < \nu_{4A}]$ or
2. $[\eta < \eta^*]$ and $[\nu < \nu_{4B}]$.

$\hat{t}_{2a}$: By definition, $\hat{t}_{2a} \equiv t_{\min}$. From equation (B.3) the condition for $\hat{t}_{2a}$ to be a maximum is

$$\frac{\nu + \alpha - \eta}{2\nu\delta} < 0,$$

which implies that $\nu < \eta - \alpha$. But this cannot hold because by Assumption 1

$\eta < \alpha$, and $\nu \geq 0$. We conclude that $\hat{t}_{2a}$ is a local minimum and cannot be the optimal tax level.

$\hat{t}_{2b}$: Because $\eta < \alpha$ and v is non-negative, by (B.4) the second-order condition is always satisfied at $\hat{t}_{2b}$, which is therefore a local maximum. For feasibility $\hat{t}_{2b}$ must be an element of T_2, but due to the $\min\{\cdot\}$ function in the expression for $\bar{t}_2$ we must consider separately the cases when $T_1 = \emptyset$ and $T_4 = \emptyset$. Suppose first that $T_1 = \emptyset$, which is equivalent to $\nu < \delta e^0/2$. Then $\hat{t}_{2b} \in T_2$ implies from (2.7b) that $\hat{t}_{2b} \in [\underline{t}_2, \bar{t}_2]$ which is equivalent to

$$\eta - \nu \leq \frac{\beta(\eta - \nu) - 2\alpha\delta}{\beta - 2\delta} \leq \eta + \nu.$$

The left inequality is equivalent to $\nu \geq \eta - \alpha$, which is always true because $\eta \leq \alpha$ and $\nu \geq 0$. The right inequality is equivalent to

$$\nu \geq \frac{\delta(\eta - \alpha)}{\beta - \delta}.$$

Combining this expression with the condition for $T_1 = \emptyset$ shows the sufficient condition required to hold for $\hat{t}_{2b}$ to be feasible when $T_1 = \emptyset$:

$$\frac{\delta(\eta - \alpha)}{\beta - \delta} \leq \nu < \frac{\delta e^0}{2}.$$

This requires $\delta(\eta - \alpha)/(\beta - \delta) < \delta e^0/2$ which is true if and only if $\eta > \eta^*$.

Now suppose that $T_4 = \emptyset$, which is equivalent to $\nu \geq \delta e^0/2$. In this case $\hat{t}_{2b} \in T_2$ implies from (2.7b) that $\hat{t}_{2b} \in [\underline{t}_2, \bar{t}_2)$ which is equivalent to

$$\eta - \nu \leq \frac{\beta(\eta - \nu) - 2\alpha\delta}{\beta - 2\delta} < \eta + \delta e^0 - \nu.$$

As before the left inequality will always be satisfied. The right inequality is equivalent to

$$\nu < \eta - \alpha - \frac{e^0}{2}(\beta - 2\delta).$$

Combining this inequality with the condition for $T_4 = \emptyset$ shows the sufficient conditions required to hold for $\hat{t}_{2b}$ to be feasible when $T_4 = \emptyset$:

$$\frac{\delta e^0}{2} \leq \nu < \eta - \alpha - \frac{e^0}{2}(\beta - 2\delta).$$

This requires that $\delta e^0/2 < \eta - \alpha - \frac{e^0}{2}(\beta - 2\delta)$, which is true if and only if $\eta > \eta^*$

Combining the sufficient conditions for $\hat{t}_{2b}$ feasibility when $T_1 = \emptyset$ and $T_4 = \emptyset$ we conclude that $\hat{t}_{2b}$ is the optimal tax if and only if:

$$[\eta > \eta^*] \text{ and } [\nu_{4A} \leq \nu < \nu_{1A}],$$

where ν_{1A} and ν_{4A} are defined in equations (2.19) and (2.20).

$\hat{t}_{3a}$: By (2.17), $\hat{t}_{3a} \equiv t_{\max}$. From equation (B.5) we know that $\hat{t}_{3a}$ is a local maximum if and only if

$$\frac{\nu - \alpha - e^0(\beta - \delta) + \eta}{2\nu\delta} < 0,$$

which implies that $\nu < e^0(\beta - \delta) + \alpha - \eta$. But this cannot hold because by Assumption 1, $\eta > e^0(\beta - \delta) + \alpha$, and $\nu \geq 0$. We conclude that $\hat{t}_{3a}$ is a local minimum and cannot be the optimal tax.

$\hat{t}_{3b}$: Because $\eta > e^0(\beta - \delta) + \alpha$ and v is non-negative, the second-order condition in (B.6) is always satisfied at $\hat{t}_{3b}$. It is therefore a local maximum when feasible. For feasibility of $\hat{t}_{3b}$ it must be an element of T_3, but due to the max $\{\cdot\}$ function in the expression for $\underline{t}_3$ we must consider separately the cases when $T_1 = \emptyset$ and $T_4 = \emptyset$. Suppose first that $T_1 = \emptyset$, which is equivalent to $\nu < \delta e^0/2$. Then $\hat{t}_{3b} \in T_3$ implies from (2.7c) that $\hat{t}_{3b} \in [\underline{t}_3, \overline{t}_3]$ which is equivalent to

$$\eta + \delta e^0 - \nu \leq \frac{\beta(\eta + \nu) - \beta\delta e^0 - 2\alpha\delta}{\beta - 2\delta} \leq \eta + \nu\delta e^0.$$

The right inequality is equivalent to $\nu \geq e^0(\beta - \delta) + \alpha - \eta$, which is always true because $\eta > e^0(\beta - \delta) + \alpha$ and $\nu \geq 0$. The left inequality is equivalent to

$$\nu \geq \frac{\delta\left[e^0(\beta - \delta) + \alpha - \eta\right]}{\beta - \delta}.$$

Combining this expression with the condition for $T_1 = \emptyset$ shows the sufficient condition required to hold for $\hat{t}_{3b}$ to be feasible when $T_1 = \emptyset$:

$$\frac{\delta\left[e^0(\beta-\delta)+\alpha-\eta\right]}{\beta-\delta} \leq \nu < \frac{\delta e^0}{2}.$$

This requires that $\delta\left[e^0(\beta-\delta)+\alpha-\eta\right]/(\beta-\delta) < \delta e^0/2$ which is true if and only if $\eta < \eta^*$.

Now suppose that $T_4 = \emptyset$, which is equivalent to $\nu \geq \delta e^0/2$. In this case $\hat{t}_{3b} \in T_3$ implies from (2.7c) that $\hat{t}_{3b} \in (\underline{t}_3, \bar{t}_3]$ which is equivalent to

$$\eta+\nu < \frac{\beta(\eta+\nu)-\beta\delta e^0-2\alpha\delta}{\beta-2\delta} \leq \eta+\nu+\delta e^0.$$

As before the right inequality will always be satisfied. The left inequality is equivalent to

$$\nu < \alpha-\eta+\frac{\beta e^0}{2}.$$

Combining this inequality with the condition for $T_4 = \emptyset$ shows the sufficient condition required to hold for $\hat{t}_{3b}$ to be feasible when $T_4 = \emptyset$:

$$\frac{\delta e^0}{2} \leq \nu < \alpha-\eta+\frac{\beta e^0}{2}.$$

This requires that $\delta e^0/2 < \alpha-\eta+\beta e^0/2$ which is true if and only if $\eta < \eta^*$

Combining the sufficient conditions for $\hat{t}_{3b}$ feasibility when $T_1 = \emptyset$ and $T_4 = \emptyset$ we conclude that $\hat{t}_{3b}$ is the optimal tax if and only if:

$$[\eta < \eta^*] \text{ and } [\nu_{4B} \leq \nu < \nu_{1B}],$$

where ν_{1B} and ν_{4B} are defined in equations (2.20) and (2.22)

Note that the ranges in which these four taxes are optimal do not overlap. They also form a partition over the entire interval $[0,\infty)$ of possible values of ν. Therefore, the optimal tax policy is simply the one local maximum that, for any set of parameters, satisfies the corresponding conditions.

One final consideration is when $\eta = \eta^*$. In this instance $\nu_{1A} = \nu_{1B} = \nu_{4A} = \nu_{4B}$. Thus, when $\eta = \eta^*$, $\hat{t}_{2b}$ and $\hat{t}_{3b}$ are never feasible and the optimal tax is at $\hat{t}_1 = \hat{t}_4$.

The optimal tax is given in equation (2.23).

B.2 Optimal tax rule for $\boldsymbol{\beta \in (\delta, 2\delta)}$

This is the first situation in which Assumption 2 is satisfied. Under Assumption 2 we have

$$\eta > \alpha \quad \text{and} \quad \eta < e^0(\beta - \delta) + \alpha$$

As in the previous section the six possible local optima are considered in turn to find the local maxima and regions of feasibility. Unlike in section B.1 it is possible for the feasible regions of the local maxima to overlap. Thus, the global maximum is found by comparing the value of the relevant Γ_j functions at the local optima for the overlapping regions.

$\hat{t}_1$: The analysis of $\hat{t}_1$ is identical to that in section B.1. Thus, $\hat{t}_1$ is a local maximum and $\hat{t}_1 \in T_1$ in each of the following cases:

1. $[\eta > \eta^*]$ and $[\nu \geq \nu_{1A}]$ or
2. $[\eta < \eta^*]$ and $[\nu \geq \nu_{1B}]$,

where ν_{1A} and ν_{1B} are defined in equations (2.19) and (2.20).

$\hat{t}_4$: Under Assumption 2, where $\beta > \delta$, the inequality in (B.2) is never satisfied. Thus, $\hat{t}_4$ is a local minimum and we conclude that it cannot be the optimal tax level.

$\hat{t}_{2a}$: By (2.16), $\hat{t}_{2a} \equiv t_{\min}$ and from (2.7b), $t_{\min} \in T_2$, therefore, $\hat{t}_{2a}$ is always feasible. Equation (B.3) implies that $\hat{t}_{2a}$ is a local maximum when $\nu < \eta - \alpha$. Under Assumption 2, $\eta > \alpha$ so there exists a region of the non-negative ν-space such that $\hat{t}_{2a}$ is a feasible local maximum.

$\hat{t}_{2b}$: By (B.4) the second-order condition is satisfied at $\hat{t}_{2b}$ when $\nu > \eta - \alpha$ and is therefore a local maximum over this region of ν-space. For feasibility of $\hat{t}_{2b}$ it must be an element of T_2, but due to the $\min\{\cdot\}$ function in the expression for $\bar{t}_2$ we must consider separately the cases when $T_1 = \emptyset$ and $T_4 = \emptyset$. Suppose first that $T_1 = \emptyset$, which is equivalent to $\nu < \delta e^0/2$. In this case $\hat{t}_{2b} \in T_2$ implies from (7b) that $\hat{t}_{2b} \in [\underline{t}_2, \bar{t}_2]$ which is equivalent to

$$\eta - \nu \leq \frac{\beta(\eta - \nu) - 2\alpha\delta}{\beta - 2\delta} \leq \eta + \nu.$$

The left inequality is equivalent to $\nu \geq \eta - \alpha$, which is always true when $\hat{t}_{2b}$ is a local maximum. The right inequality is equivalent to

$$\nu \leq \frac{\delta(\eta - \alpha)}{\beta - \delta}.$$

Combining this expression with the condition for $T_1 = \emptyset$ along with the feasibility condition that $\nu \geq \eta - \alpha$ shows the sufficient conditions required to hold for $\hat{t}_{2b}$ to be feasible when $T_1 = \emptyset$:

$$\eta - \alpha \leq \nu < \min\left\{\frac{\delta(\eta - \alpha)}{\beta - \delta}, \frac{\delta e^0}{2}\right\}.$$

The first term in the min$\{\cdot\}$ function is relavent ($\delta(\eta - \alpha)/(\beta - \delta) < \delta e^0/2$) if and only if $\eta < \eta^*$. Thus, when $T_1 = \emptyset$ and $\eta < \eta^*$, $\hat{t}_{2b}$ is feasible and a local maximum when $\eta - \alpha < \nu \leq \delta(\eta - \alpha)/(\beta - \delta)$. When $T_1 = \emptyset$ but $\eta > \eta^*$, the second term in the min$\{\cdot\}$ function is relavent, and $\hat{t}_{2b}$ is feasible and a local maximum when $\eta - \alpha < \nu < \delta e^0/2$.

Now suppose that $T_4 = \emptyset$, which is equivalent to $\nu \geq \delta e^0/2$. In this case $\hat{t}_{2b} \in T_2$ implies from (7b) that $\hat{t}_{2b} \in [\underline{t}_2, \bar{t}_2)$ which is equivalent to

$$\eta - \nu \leq \frac{\beta(\eta - \nu) - 2\alpha\delta}{\beta - 2\delta} < \eta + \delta e^0 - \nu.$$

As before, the left inequality will always be satisfied. The right inequality is equivalent to

$$\nu < \eta - \alpha - \frac{e^0}{2}(\beta - 2\delta).$$

Combining this inequality with the condition for $T_4 = \emptyset$ shows the sufficient conditions that must hold in order for $\hat{t}_{2b}$ to be feasible when $T_4 = \emptyset$:

$$\frac{\delta e^0}{2} \leq \nu < \eta - \alpha - \frac{e^0}{2}(\beta - 2\delta).$$

This requires $\delta e^0/2 < \eta - \alpha - (e^0/2)(\beta - 2\delta)$ which is true if and only if $\eta > \eta^*$.

Combining the sufficient conditions for $\hat{t}_{2b}$ feasibility when $T_1 = \emptyset$ and $T_4 = \emptyset$ we conclude that $\hat{t}_{2b}$ is a local maximum and feasible in each of the following cases:

1. $[\eta > \eta^*]$ and $[\eta - \alpha < \nu < \nu_{1A}]$ or
2. $[\eta < \eta^*]$ and $[\eta - \alpha < \nu \leq \nu_{4A}]$,

where ν_{1A} and ν_{4A} are defined in equations (2.19) and (2.20).

$\hat{t}_{3a}$: By (2.17), $\hat{t}_{3a} \equiv t_{\max}$ and from (2.7c), $t_{\max} \in T_3$, therefore, $\hat{t}_{3a}$ is always feasible. Equation (B.5) implies that $\hat{t}_{3a}$ is a local maximum when $\nu < e^0(\beta - \delta) + \alpha - \eta$. Under Assumption 2, $\eta < e^0(\beta - \delta) + \alpha$ so there exists a region of the non-negative ν-space such that $\hat{t}_{3a}$ is a feasible local maximum.

$\hat{t}_{3b}$: By (B.6) the second-order condition is satisfied at $\hat{t}_{3b}$ when $\nu > e^0(\beta - \delta) + \alpha - \eta$ and is therefore a local maximum over this region of ν-space. For feasibility of $\hat{t}_{3b}$ it must be an element of T_3, but due to the $\max\{\cdot\}$ function in the expression for $\underline{t}_3$ we must consider separately the cases when $T_1 = \emptyset$ and $T_4 = \emptyset$. Suppose first that $T_1 = \emptyset$, which is equivalent to $\nu < \delta e^0/2$. Then $\hat{t}_{3b} \in T_3$ implies from (2.7c) that $\hat{t}_{3b} \in [\underline{t}_3, \bar{t}_3]$ which is equivalent to

$$\eta + \delta e^0 - \nu \leq \frac{\beta(\eta + \nu) - \beta\delta e^0 - 2\alpha\delta}{\beta - 2\delta} \leq \eta + \nu + \delta e^0.$$

The right inequality is equivalent to $\nu \geq e^0(\beta - \delta) + \alpha - \eta$, which is always true when $\hat{t}_{3b}$ is a local maximum. The left inequality is equivalent to

$$\nu \leq \frac{\delta[e^0(\beta - \delta) + \alpha - \eta]}{\beta - \delta}.$$

Combining this expression with the condition for $T_1 = \emptyset$ along with the feasibility condition that $\nu \geq e^0(\beta - \delta) + \alpha - \eta$ shows the sufficient conditions that must hold in order for $\hat{t}_{3b}$ to be feasible when $T_1 = \emptyset$:

$$e^0(\beta - \delta) + \alpha - \eta \leq \nu < \min\left\{ \frac{\delta[e^0(\beta-\delta)+\alpha-\eta]}{\beta-\delta}, \frac{\delta e^0}{2} \right\}.$$

The first term in the min $\{\cdot\}$ function is relavent $(\delta[e^0(\beta-\delta)+\alpha-\eta]/(\beta-\delta) < \delta e^0/2)$ if and only if $\eta > \eta^*$. Thus, when $T_1 = \emptyset$ and $\eta > \eta^*$, $\hat{t}_{3b}$ is feasible and a local maximum when $e^0(\beta-\delta)+\alpha-\eta < \nu \leq \delta[e^0(\beta-\delta)+\alpha-\eta]/(\beta-\delta)$. When $T_1 = \emptyset$ but $\eta < \eta^*$, the second term in the min$\{\cdot\}$ function is relavent, and $\hat{t}_{3b}$ is feasible and a local maximum when $e^0(\beta-\delta)+\alpha-\eta < \nu < \delta e^0/2$.

Now suppose that $T_4 = \emptyset$, which is equivalent to $\nu \geq \delta e^0/2$. In this case $\hat{t}_{3b} \in T_3$ implies from (2.7c) that $\hat{t}_{3b} \in (\underline{t}_3, \bar{t}_3]$ which is equivalent to

$$\eta + \nu < \frac{\beta(\eta+\nu) - \beta\delta e^0 - 2\alpha\delta}{\beta - 2\delta} \leq \eta + \nu + \delta e^0.$$

As before the right inequality will always be satisfied. The left inequality is equivalent to

$$\nu < \alpha - \eta + \frac{\beta e^0}{2}.$$

Combining this inequality with the condition for $T_4 = \emptyset$ shows the sufficient conditions that must hold in order for $\hat{t}_{3b}$ to be feasible when $T_4 = \emptyset$:

$$\frac{\delta e^0}{2} \leq \nu < \alpha - \eta + \frac{\beta e^0}{2}.$$

This requires $\delta e^0/2 < \alpha - \eta + \beta e^0/2$ which is true if and only if $\eta < \eta^*$.

Combining the sufficient conditions for $\hat{t}_{3b}$ feasibility when $T_1 = \emptyset$ and $T_4 = \emptyset$ we conclude that $\hat{t}_{3b}$ is a local maximum and feasible in each of the following cases:

1. $[\eta < \eta^*]$ and $\left[e^0(\beta-\delta)+\alpha-\eta < \nu < \nu_{1B}\right]$ or
2. $[\eta < \eta^*]$ and $\left[e^0(\beta-\delta)+\alpha-\eta < \nu \leq \nu_{4B}\right]$,

where ν_{1B} and ν_{4B} are defined in equations (2.20) and (2.22).

To find the global maximum we must combine the feasibility conditions and compare Γ_j values for each of the local maxima: $\hat{t}_1, \hat{t}_{2a}, \hat{t}_{3a}, \hat{t}_{2b}, \hat{t}_{3b}$.

For low values of ν the global maximum will either be at $\hat{t}_{2a}$ or $\hat{t}_{3a}$. The Γ_j values for these two local maxima are

$$E[SW(t_{\min})] = \Gamma_2(\hat{t}_{2a}) = 0 \quad \text{and} \tag{B.7}$$

$$\begin{aligned} E[SW(t_{\max})] = \Gamma_3(\hat{t}_{3a}) &= \int_{-\nu}^{\nu} \left(\frac{\beta - \delta}{2}(e^0)^2 + e^0(\alpha - \eta - u) \right) du \\ &= e^0(\eta^* - \eta). \end{aligned} \tag{B.8}$$

Thus, $\hat{t}_{3a}$ is strictly preferred to $\hat{t}_{2a}$ if and only if $\Gamma_3(\hat{t}_{3a}) > \Gamma_2(\hat{t}_{2a}) = 0$. This is true precisely when $\eta < \eta^*$. Conversely, $\hat{t}_{2a}$ is strictly preferred to $\hat{t}_{3a}$ if and only if $\eta > \eta^*$.

Consider first the case where $\eta > \eta^*$. The sufficient conditions for each of the local maxima to be feasible are summarized below[1]:

$$\begin{aligned} \hat{t}_{2a} &: \nu \in [0, \eta - \alpha] \\ \hat{t}_{3a} &: \nu \in [0, e^0(\beta - \delta) + \alpha - \eta] \\ \hat{t}_{2b} &: \nu \in (\eta - \alpha, \nu_{1A}) \\ \hat{t}_{3b} &: \nu \in (e^0(\beta - \delta) + \alpha - \eta, \nu_{4B}] \\ \hat{t}_{1} &: \nu \in [\nu_{1A}, \infty) \end{aligned}$$

The inequality $\eta > \eta^*$ implies both that $\eta - \alpha > e^0(\beta - \delta) + \alpha - \eta$ and that $\nu_{1A} > \nu_{4B}$. This first condition demonstrates that $\hat{t}_{3a}$ cannot be the global maximum because its feasible range is contained in the feasible range of $\hat{t}_{2a}$, and when $\eta > \eta^*$, $\hat{t}_{2a}$ strictly dominates $\hat{t}_{3a}$. We state without proof that on the interval $(e^0(\beta - \delta) + \alpha - \eta, \eta - \alpha]$, $\Gamma_2(\hat{t}_{2a}) > \Gamma_3(\hat{t}_{3b})$; and on the interval $(\eta - \alpha, \nu_{4B}]$, $\Gamma_2(\hat{t}_{2b}) > \Gamma_3(\hat{t}_{3b})$. These conditions show that the feasible local maximum $\hat{t}_{3b}$ cannot be the global maximum when $\eta > \eta^*$. With $\hat{t}_{3a}$ and $\hat{t}_{3b}$ eliminated from consideration as the global maximum, the optimal tax policy is simply the single remaining feasible local maximum in each region of ν-space. Notice that the feasible ranges for $\hat{t}_{2a}$, $\hat{t}_{2b}$ and $\hat{t}_1$ form a partition over the entire interval $[0, \infty)$ of possible values of ν.

[1] When $\nu = \eta - \alpha$, $\hat{t}_{2a} = \hat{t}_{2b}$ and neither are local maxima, rather this is an inflection point of Γ_2. Similarly, when $\nu = e^0(\beta - \delta) + \alpha - \eta$, $\hat{t}_{3a} = \hat{t}_{3b}$ and it is an inflection point on Γ_3. $\nu = \eta - \alpha$ is included in the feasible range for $\hat{t}_{2a}$ but not $\hat{t}_{2b}$; and $\nu = e^0(\beta - \delta) + \alpha - \eta$ is included in the feasible range for $\hat{t}_{3a}$ but not for $\hat{t}_{3b}$.

Now consider the case where $\eta < \eta^*$. The sufficient conditions for each of the local maxima to be feasible are summarized below:

$$\begin{aligned}
\hat{t}_{2a} &: \nu \in [0, \eta - \alpha] \\
\hat{t}_{3a} &: \nu \in [0, e^0(\beta - \delta) + \alpha - \eta] \\
\hat{t}_{2b} &: \nu \in (\eta - \alpha, \nu_{4A}] \\
\hat{t}_{3b} &: \nu \in (e^0(\beta - \delta) + \alpha - \eta, \nu_{1B}) \\
\hat{t}_1 &: \nu \in [\nu_{1B}, \infty)
\end{aligned}$$

The inequality $\eta < \eta^*$ implies both that $e^0(\beta-\delta)+\alpha-\eta > \eta-\alpha$ and that $\nu_{1B} > \nu_{4A}$. The first of these conditions demonstrates that $\hat{t}_{2a}$ cannot be the global maximum because its feasible range is contained in the feasible range of $\hat{t}_{3a}$, and when $\eta < \eta^*$, $\hat{t}_{3a}$ strictly dominates $\hat{t}_{2a}$. We state without proof that on the interval $(\eta-\alpha, e^0(\beta-\delta)+\alpha-\eta]$, its true that $\Gamma_3(\hat{t}_{3a}) > \Gamma_2(\hat{t}_{2b})$; and on the interval $(e^0(\beta-\delta)+\alpha-\eta, \nu_{4A}]$, its true that $\Gamma_3(\hat{t}_{3b}) > \Gamma_2(\hat{t}_{2b})$. These conditions show that the feasible local maximum $\hat{t}_{2b}$ cannot be the global maximum when $\eta < \eta^*$. With $\hat{t}_{2a}$ and $\hat{t}_{2b}$ eliminated from consideration as the global maximum, the optimal tax policy is simply the single remaining feasible local maximum in each region of ν-space. Notice that the feasible ranges for $\hat{t}_{3a}$, $\hat{t}_{3b}$ and $\hat{t}_1$ form a partition over the interval $[0, \infty)$ of possible values of ν.

One final consideration is the case in which $\eta = \eta^*$. In this instance $\eta - \alpha = e^0(\beta - \delta) + \alpha - \eta$, $\nu_{1A} = \nu_{1B}$, $\Gamma_2(\hat{t}_{2a}) = \Gamma_3(\hat{t}_{3a})$ and $\Gamma_2(\hat{t}_{2b}) = \Gamma_3(\hat{t}_{3b})$. Thus, when $\eta = \eta^*$, $\hat{t}_{2a}$ and $\hat{t}_{3a}$ are both optimal tax levels for $\nu \leq \eta - \alpha = e^0(\beta - \delta) + \alpha - \eta$; $\hat{t}_{2b}$ and $\hat{t}_{3b}$ are both optimal tax levels for $\nu \in (\eta - \alpha = e^0(\beta - \delta) + \alpha - \eta, \nu_{1A} = \nu_{1B})$; and $\hat{t}_1$ is the optimal tax for $\nu \geq \nu_{1A} = \nu_{1B}$.

The optimal tax is given in equation (2.24)

B.3 Optimal tax rule for $\beta \geq 2\delta$

In this situation Assumption 2 is again satisfied which implies that

$$\eta > \alpha \quad \text{and} \quad \eta < e^0(\beta - \delta) + \alpha.$$

The six possible local optima are considered again to find the local maxima and regions of feasbility. The global maximum is found by comparing the value of the relevant Γ_j functions at the local optimas for the entire region of possible values of ν.

$\hat{t}_1$: The analysis of $\hat{t}_1$ is identical to that in section B.1. Thus, $\hat{t}_1$ is always a local maximum when feasible and $\hat{t}_1 \in T_1$ in each of the following cases:

1. $[\eta > \eta^*]$ and $[\nu \geq \nu_{1A}]$ or
2. $[\eta < \eta^*]$ and $[\nu \geq \nu_{1B}]$,

where ν_{1A} and ν_{1B} are defined in equations (2.19) and (2.20).

$\hat{t}_4$: Under Assumption 2, where $\beta > \delta$, the inequality in (B.2) is never satisfied. Thus, $\hat{t}_4$ is a local minimum and we conclude that it cannot be the optimal tax.

$\hat{t}_{2a}$: By (2.16), $\hat{t}_{2a} \equiv t_{\min}$ and from (7b), $t_{\min} \in T_2$, therefore, $\hat{t}_{2a}$ is always feasible. Equation (B.3) implies that $\hat{t}_{2a}$ is a local maximum when $\nu < \eta - \alpha$. Under Assumption 2, $\eta > \alpha$ so there exists a region of the non-negative ν-space such that $\hat{t}_{2a}$ is a feasible local maximum.

$\hat{t}_{2b}$: Equation (B.4) implies that $\hat{t}_{2b}$ is a local maximum when $\nu > \eta - \alpha$. A necessary condition for feasibility of $\hat{t}_{2b}$ is that $\hat{t}_{2b} \geq \underline{t}_2$ which implies

$$\eta - \nu \leq \frac{\beta(\eta - \nu) - 2\alpha\delta}{\beta - 2\delta}.$$

This inequality is equivalent to $\nu \leq \eta - \alpha$, but this violates the second-order condition for $\hat{t}_{2b}$ to be a local maximum. Therefore, if $\hat{t}_{2b}$ is feasible it cannot be a local maximum, so we conclude it cannot be the optimal tax.

$\hat{t}_{3a}$: By (2.17), $\hat{t}_{3a} \equiv t_{\max}$ and from (7c), $t_{\max} \in T_3$, therefore, $\hat{t}_{3a}$ is always feasible. Equation (B.5) implies that $\hat{t}_{3a}$ is a local maximum when $\nu < e^0(\beta - \delta) + \alpha - \eta$. Under Assumption 2, $\eta < e^0(\beta - \delta) + \alpha$ so there exists a region of the non-negative ν-space such that $\hat{t}_{3a}$ is a feasible local maximum.

$\hat{t}_{3b}$: Equation (B.6) implies that $\hat{t}_{3b}$ is a local maximum when $\nu > e^0(\beta - \delta) + \alpha - \eta$.

A necessary condition for feasibility of $\hat{t}_{3b}$ is that $\hat{t}_{3b} \leq \bar{t}_3$ which implies

$$\frac{\beta(\eta+\nu) - \beta\delta e^0 - 2\alpha\delta}{\beta - 2\delta} \leq \eta + \nu + \delta e^0.$$

This inequality is equivalent to $\nu \leq e^0(\beta-\delta)+\alpha-\eta$, but this violates the second-order condition for $\hat{t}_{3b}$ to be a local maximum. Therefore, if $\hat{t}_{3b}$ is feasible it cannot be a local maximum, so we conclude it cannot be the optimal tax.

To find the global maximum we must compare the feasibility conditions and Γ_j values for each of the local maxima: $\hat{t}_1, \hat{t}_{2a}$, and $\hat{t}_{3a}$. As was demonstrated in section B.2 a tax at $\hat{t}_{2a}$ is strictly preferred to $\hat{t}_{3a}$ if $\eta > \eta^*$ and vice versa when $\eta < \eta^*$.

Consider first the situation where $\eta > \eta^*$. The tax $\hat{t}_{2a}$ is always feasible and is a maximum if $\nu < \eta - \alpha$. The only other possible optimal tax is $\hat{t}_1$ which is always a local maximum and is feasible if $\nu \geq \nu_{1A}$. First, notice that $\eta - \alpha \geq \nu_{1A}$ is equivalent to $\beta \geq 2\delta$, which is given by assumption. Next, we want to find the critical value of ν at which $\Gamma_1(\hat{t}_1) \geq \Gamma_2(\hat{t}_{2a})$. Because $\Gamma_2(\hat{t}_{2a}) = 0$ we need to find the threshold value of ν at which $\Gamma_1(\hat{t}_1) \geq 0$. Solving this expression for ν:

$$\nu \geq \eta - \eta^* + \frac{e^0}{6}\sqrt{3\delta(2\beta-\delta)}.$$

Define

$$\nu^*_{\min} = \eta - \eta^* + \frac{e^0}{6}\sqrt{3\delta(2\beta-\delta)} \tag{B.9}$$

as the value of ν that equates $\Gamma_2(\hat{t}_{2a})$ and $\Gamma_1(\hat{t}_1)$. Further,

$$\Gamma_1(\hat{t}_1) - \Gamma_2(\hat{t}_{2a}) = \frac{e^0}{4\nu}\Big(\nu - \nu^*_{\min}\Big)\Big(\nu - \nu^*_{\min} + \frac{e^0}{3}\sqrt{3\delta(2\beta-\delta)}\Big),$$

which is strictly positive for $\nu > \nu^*_{\min}$. Two more conditions must be considered. First, $\nu^*_{\min} \geq \nu_{1A}$ is equivalent to $\beta \geq 2\delta$, which is given by assumption. Second, $\nu^*_{\min} \leq \eta - \alpha$ holds whenever $\beta \geq 2\delta$. Therefore, when $\nu < \nu^*_{\min}$, $\hat{t}_{2a}$ is always a maximum and is strictly preferred to $\hat{t}_1$; when $\nu > \nu^*_{\min}$, $\hat{t}_1$ is feasible and is strictly preferred to $\hat{t}_{2a}$; and when $\nu = \nu^*_{\min}$, $\Gamma_2(\hat{t}_{2a}) = \Gamma_1(\hat{t}_1)$. Thus, $\hat{t}_{2a}$ is the optimal tax level for $\nu \in [0, \nu^*_{\min}]$ and $\hat{t}_1$ is the optimal tax level for $\nu \in [\nu^*_{\min}, \infty)$.

Next consider the case with $\eta < \eta^*$. The value $\hat{t}_{3a}$ is always feasible and is a local maximum if $\nu < e^0(\beta - \delta) + \alpha - \eta$. The only other possible optimal tax is $\hat{t}_1$ which is always a maximum and is feasible if $\nu \geq \nu_{1B}$. First, notice that $e^0(\beta - \delta) + \alpha - \eta \geq \nu_{1B}$ is equivalent to $\beta \geq 2\delta$, which is given by assumption. Next, we want to find the critical value of ν at which $\Gamma_1(\hat{t}_1) \geq \Gamma_3(\hat{t}_{3a})$. Solving this expression for ν:

$$\nu \geq \eta^* - \eta + \frac{e^0}{6}\sqrt{3\delta(2\beta - \delta)}.$$

Define

$$\nu^*_{\max} = \eta^* - \eta + \frac{e^0}{6}\sqrt{3\delta(2\beta - \delta)} \tag{B.10}$$

as the value of ν that equates $\Gamma_3(\hat{t}_{3a})$ and $\Gamma_1(\hat{t}_1)$. Further,

$$\Gamma_1(\hat{t}_1) - \Gamma_3(\hat{t}_{3a}) = \frac{e^0}{4\nu}\left(\nu - \nu^*_{\max}\right)\left(\nu - \nu^*_{\max} + \frac{e^0}{3}\sqrt{3\delta(2\beta - \delta)}\right).$$

which is strictly positive for $\nu > \nu^*_{\max}$. Two more conditions must be considered. First, $\nu^*_{\max} \geq \nu_{1B}$ is equivalent to $\beta \geq 2\delta$, which is given by assumption. Second, $\nu^*_{\max} \leq e^0(\beta - \delta) + \alpha - \eta$ holds whenever $\beta \geq 2\delta$. Therefore, when $\nu < \nu^*_{\max}$, $\hat{t}_{3a}$ is always a maximum and is strictly preferred to $\hat{t}_1$; when $\nu > \nu^*_{\max}$, $\hat{t}_1$ is feasible and is strictly preferred to $\hat{t}_{3a}$; and when $\nu = \nu^*_{\max}$, $\Gamma_3(\hat{t}_{3a}) = \Gamma_1(\hat{t}_1)$. Thus, $\hat{t}_{3a}$ is the optimal tax level for $\nu \in [0, \nu^*_{\max}]$, and $\hat{t}_1$ is the optimal tax level for $\nu \in [\nu^*_{\max}, \infty)$.

One final consideration is the case in which $\eta = \eta^*$. In this instance $\Gamma_2(\hat{t}_{2a}) = \Gamma_3(\hat{t}_{3a})$ and $\nu^*_{\min} = \nu^*_{\max}$. Thus, when $\eta = \eta^*$, $\hat{t}_{2a}$ and $\hat{t}_{3a}$ are both optimal tax levels for $\nu \leq \nu^*_{\min} = \nu^*_{\max}$; and $\hat{t}_1$ is the optimal tax for $\nu \geq \nu^*_{\min} = \nu^*_{\max}$.

The optimal tax is given in equation (2.27).

B.4 Restricting marginal costs to be nonnegative

We have been silent on the possibility that, for large ν, the marginal cost of abatement might be negative for small a. Negative marginal costs might not be considered unreasonable, if one is persuaded by the argument of Porter and van der Linde (1995). Requiring nonnegative marginal costs complicates our analysis, but not irretrievably so. Here we sketch the changes one would need to make, under Assumption 2, in order to account for a restriction that $\nu \leq \eta$.

For $\beta < 2\delta$

In order for $\hat{t}_1$ to be the optimal tax, either (2.20) or (2.22) must be violated. Because we are restricting the lower bound on ν and not the upper bound, and because (2.22) is relevant when $\eta < \eta^*$, we need only determine conditions guaranteeing that η is greater than or equal to the right side of (2.20). That is:

$$\eta \geq \frac{\beta e^0}{4} + \frac{\alpha}{2}. \tag{B.11}$$

Therefore, if negative marginal costs are ruled out $\hat{t}_1$ can be the optimal tax level if (B.11) is satisfied. Because $\hat{t}_{2b}$ can be optimal only if $\eta > \eta^*$, requiring marginal costs to be non-negative does not further restrict the feasibility of $\hat{t}_{2b}$ being the optimal tax level.

In order for $\hat{t}_{3b}$ to be the optimal tax level, we must have $\nu > e^0 (\beta - \delta) + \alpha - \eta$, which is true only if

$$\eta > \eta^* - \frac{\alpha}{2}. \tag{B.12}$$

If negative marginal costs are ruled out, an interior tax level that strictly dominates a quantity policy can exist if (B.12) holds. Without restrictions on marginal cost an interior tax level that dominates the quantity policy can always exist with a sufficiently high ν.

For $\beta \geq 2\delta$

In order for $\hat{t}_1$ to be the optimal tax, either $\nu \geq \nu^*_{\text{min}}$ or $\nu \geq \nu^*_{\text{max}}$ must hold. Because we must restrict the lower bound on ν and not the upper bound, and because ν^*_{max} is relevant when $\eta < \eta^*$, we need only determine conditions under which $\nu^*_{\text{max}} \leq \eta$. This is true whenever

$$\eta \geq \frac{\eta^*}{2} + \frac{e^0}{12}\sqrt{3\delta\,(2\beta - \delta)}. \tag{B.13}$$

Therefore, if negative marginal costs are ruled out the interior tax level that strictly dominates a quantity policy can exist if (B.13) is satisfied. Without restrictions on marginal cost, an interior tax level that dominates the quantity policy can always exist with a sufficiently large ν.