DISCUSSION PAPER SERIES

Quasi-Experimental Methods in Environmental Economics: Opportunities and Challenges

Olivier Deschenes
Kyle C. Meng

Olivier Deschenes
UC Santa Barbara, IZA and NBER

Kyle C. Meng
UC Santa Barbara and NBER

ABSTRACT

Quasi-Experimental Methods in Environmental Economics: Opportunities and Challenges*

This paper examines the application of quasi-experimental methods in environmental economics. We begin with two observations: i) standard quasi-experimental methods, first applied in other microeconomic fields, typically assume unit-level treatments that do not spill over across units; (ii) because public goods, such as environmental attributes, exhibit externalities, treatment of one unit often affects other units. To explore the implications of applying standard quasi-experimental methods to public good problems, we extend the potential outcomes framework to explicitly distinguish between *unit-level source* and the resulting *group-level exposure* of a public good. This new framework serves as a foundation for reviewing and interpreting key papers from the recent empirical literature. We formally demonstrate that two common quasi-experimental estimators of the marginal social benefit of a public good can be biased due to externality spillovers, even when the source of the public good itself is quasi-randomly assigned. We propose an unbiased estimator for the valuation of local public goods and discuss how it can be implemented in future studies. Finally, we consider how to preserve the advantages of the quasi-experimental approach when valuing global public goods, such as climate change mitigation, for which no control units are available.

Corresponding author:
Olivier Deschenes
Department of Economics
2127 North Hall
University of California
Santa Barbara, CA 93106
USA

E-mail: olivier@econ.ucsb.edu

1 Introduction

The optimal provision of environmental goods is a long-standing question in environmental economics. Environmental policy relies on estimates of the marginal social benefit and marginal private cost of providing environmental goods. In recent years, concerns over the potential endogeneity of environmental exposure have prompted the environmental economics literature to increasingly focus on causality. Knowledge of causal relationships is crucial for policy design: without it one cannot credibly claim that any environmental policy leads to specific outcomes. To that end, the literature has primarily turned to quasi-experimental methods. The central feature of quasi-experimental research is the care taken to understanding the nature of treatment assignment, and in defining treatment and control groups. These methods have fundamentally transformed the field of environmental economics. According to our review, quasi-experimental papers, which did not appear in the literature until 2000, now make up more than 30% of published environmental economics papers in prominent economics journals.

This paper examines the quasi-experimental literature in environmental economics over the last two decades. In particular, it is organized around the following question: what are the implications of applying standard quasi-experimental methods to questions involving environmental goods? Our overall perspective is based on two observations. First, the standard set of quasi-experimental methods, which were largely developed for applications in labor economics, typically assume unit-level treatment assignment that is "as good" as random and do not spill over across units.[1] Second, environmental economics, with its strong connection to public economics, is fundamentally interested in the study of environmental public goods. Public goods exhibit externalities.[2] This implies that a public good treatment received by one unit often affects other units.

We formally explore what happens when standard quasi-experimental methods are applied to public or environmental goods settings. In particular, the presence of externalities complicates estimation of the marginal social benefit of a public good. To understand what happens to standard quasi-experimental estimates of this parameter, we first extend the potential outcomes framework to explicitly model the distinction between the *unit-level source* of a public good and the resulting *group-level exposure* to that public good. For example, consider an air quality regulation imposed on a randomly chosen power plant that reduces emissions. This treatment will lead to cleaner air over all downwind locations even if those locations were not directly regulated. Using this framework, we examine the two most commonly used quasi-experimental estimators for the marginal social benefit of a public good, both of which assume unit-level quasi-random assignment of public good sources with no treatment spillovers. We show that when there are externality spillovers, these assumptions alone are no longer sufficient for recovering unbiased estimates of the average treatment effect.

The first estimator, which we call the average source effect estimator (ASEE), compares average outcomes across units that are and are not the source of the public good. A typical application of this estimator appears in studies of the U.S. Clean Air Act that compares average outcomes across more regulated (i.e., nonattainment) and less regulated (i.e., attainment) counties (see, for example, Chay and Greenstone (2005)). We show that for a local public good, the presence of externality spillovers violates the Stable Unit Treatment

---

[1]In environmental economics, units are typically households, firms, or locations.

[2] According to Baumol and Oates (1988, p.17), "an externality is present whenever some individual's...utility or production relationships include real ... variables, whose values are chosen by others...without particular attention to the effects on [that individual's] welfare." From this definition, a public good is a good which exhibits an externality that is non-rival, whereby consumption of the public good by one agent does not deplete public good consumption by another agent. There is a further distinction in the literature. An externality is typically the unintended consequence of a private action, whereas a public good can result from either government or private action.

Value Assumption (SUTVA) making the ASEE biased downwards. For a global public good, where every control unit experiences public good spillovers, the ASEE produces an estimate of zero.

The second estimator, which we call the average exposure effect estimator (AEEE), compares average outcomes across units that are and are not exposed to a public good following quasi-random assignment of a public good source. For example, this estimator is used in studies where outcomes from a given location are regressed on pollution exposure that is sourced from elsewhere (see, for example, Schlenker and Walker (2016)). Unlike the ASEE, the AEEE for a local public good need not be biased. A form of selection bias arises, however, if potential outcomes are correlated with the likelihood of spillover. To illustrate this bias, consider a setting along a river. Even if water pollution abatement were randomly assigned across locations next to the river, downstream locations are more likely to experience cleaner water. If downstream locations also have different baseline characteristics, perhaps due to residential sorting, the AEEE will produce a biased estimate. In general, this bias is of unknown sign. Furthermore, for a global public good, we show that the AEEE is undefined.

Out of this formal analysis emerges several paths forward for future research. For local public goods, we suggest that researchers obtain unit-level likelihoods of externality spillovers. We propose an unbiased version of the ASEE for local public goods which uses spillover likelihoods as regression weights. The ASEE, however, may not be applicable for all empirical settings. For example, suppose the population of interest is households residing in locations without polluting firms. Here, only the AEEE can be estimated. To indirectly test for selection bias in the AEEE, we suggest that researchers examine whether spillover likelihoods are correlated with pre-determined characteristics. Of course, such tests would not be informative if the bias in the AEEE arises only from selection on unobservables. Unfortunately, when studying global public goods, such as climate change mitigation, adjustments or tests for either estimator do not address the fundamental problem of not having control units available. We offer some thoughts on how to work around this challenge while maintaining the advantage of causal inference provided by the quasi-experimental approach.

Our observation that treatment spillovers may undermine experimental or quasi-experimental methods is not new. This critique has been made convincingly in other contexts (e.g., Miguel and Kremer (2004) and Manski (2013)). For example, labor market interventions may lead workers to move from treated to control locations. Information treatments may inadvertently be available to individuals in the control group. Vaccine treatments may result in herd immunity for nearby control individuals. In all these contexts, identification requires the researcher to observe spillover probabilities. This is often very difficult. For example, researchers may not know where one labor market ends and another begins. Similarly, the spread of information is typically hard to observe. These factor market and informational spillovers are also present in environmental economics. However, we argue that environmental economists have an advantage when it comes to spillover of environmental goods. The availability of physical spatial models of pollution dispersal together with pollution measures at high-spatial resolutions allow researchers to observe where pollution is generated and where it goes.

Finally, we note what this review does not cover. First, while our focus is on the use of quasi-experimental methods in environmental economics, this paper does not provide a thorough review of the methodological issues associated with quasi-experimental techniques. Nor do we contrast quasi-experimental and structural approaches.[3] We also do not review papers in environmental economics that use randomized experiments,

---

[3] Interested readers can turn to excellent reviews in environmental economics by Timmins and Schlenker (2009) and Greenstone and Gayer (2009) and to more general reviews by Rosenzweig and Wolpin (2000), DiNardo (2008), Imbens and Wooldridge (2009), Heckman (2010), Angrist and Pischke (2010), Keane (2010), and Nevo and Whinston (2010), among others.

which have also become increasingly prominent.[4] Our objective is more focused: we study the implications of applying standard quasi-experimental methods to environmental good problems. Second, while we will discuss at length how to obtain causal estimates of the marginal social benefit of an environmental good, we do not address how to select outcome variables that captures the marginal willingness to pay for that good. The correct use and welfare interpretation of various outcome variables is the subject of a long and important literature on revealed preferences, which is beyond our scope.[5] Here, we simply assume that a proxy for marginal willingness to pay is available.

The rest of the paper is structured as follows: Section 2 reviews a central objective for empirical environmental economics: estimating components of the Lindahl-Samuelson condition for optimal public good provision. We review the Lindahl-Samuelson condition and discuss estimation challenges. Section 3 reviews the standard quasi-experimental approach. We define the standard quasi-experimental approach both broadly and formally using the canonical potential outcomes framework. Section 4 extends the standard potential outcomes framework to explicitly distinguish between public good source and exposure. We then evaluate two commonly used estimators in environmental economics both formally and through numerical simulations. Section 5 summarizes recent publication trends for quasi-experimental papers in environmental economics across methods, topics, and journals. We then review select papers in this literature through the lens of our extended potential outcomes framework. Section 6 offers suggestions for future quasi-experimental research in environmental economics for both local and global public goods. Section 7 concludes.

2 The Lindahl-Samuelson condition

A central concern in environmental economics is determining the socially optimal provision of an environmental good such as pollution abatement. This section reviews the optimality condition for public good provision and discusses its estimation. We begin by presenting a standard model that produces the well-known Lindahl-Samuelson condition, which states that optimality is achieved when the marginal social benefits of a public good equals the marginal private cost of its provision. We then discuss challenges to estimating these parameters both in terms of measuring welfare-relevant outcomes and identification of causal effects.

2.1 A model of optimal public good provision

There are $i = 1, \ldots, N$ agents in an economy, denoted by set $\mathcal{N}$. The subset of households, $\mathcal{H} \subset \mathcal{N}$, is indexed by $h = 1, \ldots, H$. The subset of producers, $\mathcal{L} \subset \mathcal{N}$, is indexed by $\ell = 1, \ldots, L$. The public good consumed by agent i is the sum of that generated by all producers, $Q_i = \sum_{\ell \in \mathcal{L}} q_\ell$. Households consume a private good x_h and possibly a public good Q_h, with utility function $U_h(x_h, Q_h)$, where $\partial U_h / \partial x_h \geq 0$ and $\partial U_h / \partial Q_h \geq 0$. Producers take r_ℓ as an input to produce private good y_ℓ and q_ℓ. Production can be affected by the presence of public good Q_ℓ, which we model as an input. This implies the following transformation function for producers: $F_\ell(r_\ell, Q_\ell, y_\ell, q_\ell)$, with $\partial F_\ell / \partial r_\ell \geq 0$, $\partial F_\ell / \partial Q_\ell \geq 0$, $\partial F_\ell / \partial y_\ell \leq 0$, and $\partial F_\ell / \partial q_\ell \leq 0$.

To simplify the setting, we consider Q_i to be a local, quasi-fixed, pure public good. It is "local" in the sense that the set of households and producers exposed to the public good is not the entire population. The subset of exposed households and producers, $\mathcal{M} \subset \mathcal{N}$, is indexed by $m = 1, \ldots, M$. Q_i is "quasi-fixed" in that an affected agent has no direct control over its production.[6] Q_i is "pure" or non-exclusive implying

---

[4] See for example Duflo et al. (2013), Allcott and Rogers (2014), and Fowlie, Greenstone and Wolfram (2018).

[5] We encourage interested readers to turn to Bockstael and McConnell (2007) and Freeman, Herriges and Kling (2014) for extensive reviews on revealed preference methods.

[6] In the current setup, this is reasonable if the number of producers are sufficiently large such that $q_i << Q_i$.

that, absent costly defensive behavior, the public good enters positively into either the utility function or production transformation function of an affected agent. Finally, Q_i is "public" or non-rival with exposure being the same for all affected agents. The last two assumptions imply that $U_h(x_h, Q_h = Q > 0) \ \forall h \in \mathcal{M}$ and $F_\ell(r_\ell, Q_\ell = Q > 0, y_\ell, q_\ell) \ \forall \ell \in \mathcal{M}$.

The planner's objective is to determine the socially optimal level of Q with resource constraint R. Without loss of generality, suppose agent $i = 1$ is a household affected by the public good. The Pareto problem is

$$\max_{x_1,\ldots,x_H,r_1,\ldots,r_L,y_1,\ldots y_L,q_1,\ldots,q_L} \quad U_1(x_1, Q)$$

$$s.t. \quad U_h(x_h, Q) \leq \bar{u}_h \quad \forall h = \{2, \ldots, H\}$$

$$F_\ell(r_\ell, Q, y_\ell, q_\ell) \leq 0 \quad \forall \ell = \{1, \ldots, L\}$$

$$\sum_{h \in \mathcal{H}} x_h + \sum_{\ell \in \mathcal{L}} r_\ell - \sum_{\ell \in \mathcal{L}} y_\ell \leq R$$

The socially optimal Q^* must satisfy

$$\underbrace{\underbrace{\sum_{h \in \mathcal{M}} \frac{\partial U_h / \partial Q^*}{\partial U_h / \partial x_h^*}}_{\text{amenity benefits}} - \underbrace{\sum_{\ell \in \mathcal{M}} \frac{\partial F_\ell / \partial Q^*}{\partial F_\ell / \partial y_\ell^*}}_{\text{productivity benefits}}}_{\text{marginal social benefit}} = \underbrace{\frac{\partial F_\ell / \partial q_\ell^*}{\partial F_\ell / \partial y_\ell^*}}_{\text{marginal private cost}} \quad \forall \ell = \{1, \ldots, L\} \tag{1}$$

Equation (1) is the Lindahl-Samuelson condition (Lindahl, 1919; Samuelson, 1954). It states that at the social optimum, the marginal social benefit of the public good must equal the marginal private cost of providing the public good across all producers. Marginal social benefit, the left hand side of equation (1) has two components. The first term captures the sum of the marginal rates of substitution between the public good and the private good for all affected households. This is typically referred to as amenity benefits. The second term captures the sum of the marginal rates of transformation between the public good and the private good for all producers. Because we treat the public good Q as an input to production, this captures the marginal productivity of Q in terms of output and is often referred to as productivity benefits. The marginal private cost of public good provision, shown on the right hand side of equation (1) is simply the marginal cost of increasing production of q_ℓ in terms of output, y_ℓ.

To determine the optimal level of public good provision, all three components of equation (1) must be empirically estimated. Estimation of these components using observational data entails several challenges. Section 2.2 discusses the challenges associated with measuring variables that approximate the terms in equation (1). Section 2.3 then turns to the challenge of causally identifying each effect.

2.2 Estimating the Lindahl-Samuelson condition: measurement challenges

There are significant challenges associated with observing the components of the Lindahl-Samuelson condition shown in equation (1). The first difficulty has to do with extrapolation from the observed market equilibrium to the desired social optimum. Applying the Lindahl-Samuelson condition in practice requires knowledge of the marginal social benefit and marginal private cost at the socially optimum public good level, Q^*. In general, observed data will only reveal marginal rates of substitution and transformation around the market equilibrium and not at the socially optimum. Unless marginal rates of substitution and transformation are linear functions, which we have no particular reason to assume, estimates near the market equilibrium will

not approximate estimates at the social optimum, Q^*. Thus, observational data may help inform whether public good policies are Pareto improving relative to the market equilibrium but may not help select Pareto efficient policies. Unfortunately, there is little that can be done to address this issue without imposing functional form assumptions.

The second challenge involves knowing which agents are affected by the public good. Equation (1) requires the researcher to know the marginal rate of substitution for every affected household $h \in \mathcal{M}$ and marginal rate of transformation for every affected producer $\ell \in \mathcal{M}$. This is empirically very demanding if affected households and producers have heterogeneous preferences and technologies, respectively. Instead, researchers often define population averages as estimands of interest. That, together with information on the number of affected households and producers, is then used to assemble the marginal social benefit. We return to this point again in Section 4.

The third challenge has to do with missing or socially mis-priced markets which occurs when externalities are present. For example, there is no naturally-occurring market for pollution from power plants. Furthermore, wholesale electricity prices may reflect a power plant's private value for abating pollution but is unlikely to internalize the full social benefit of pollution abatement. This is particularly problematic when trying to estimate amenity effects. As an alternative, environmental economics has developed non-market approaches based on linking preferences for a public good with preferences for private goods. Under certain assumptions, demand shifts in the linked private good market arising from changes in public good exposure can be used to infer the marginal willingness to pay for the public good. Commonly used linked private goods include housing, health-producing goods, and the time cost of travel. A long literature details the challenges of using linked private good markets to infer the marginal willingness to pay for a public good. They include, among other issues, whether the chosen outcomes measure all the relevant margins of behavioral adjustment, whether there exists non-use values associated with the public good, and whether welfare can be recovered using observed Marshallian demand of the linked private good (see Bockstael and McConnell (2007) and Freeman, Herriges and Kling (2014)).

2.3 Estimating the Lindahl-Samuelson condition: identification challenges

The measurement challenges discussed in Section 2.2 have long been the focus of empirical research in environmental economics. Conditional on observing some welfare-relevant outcome, estimation itself has traditionally employed ordinary least squares (OLS), which provides regression-adjusted differences between units exposed and non-exposed to the public good. The resulting estimates, however, may not causally identify parameters of interest. Specifically, OLS estimates may not have a causal interpretation when there are unobserved confounders - variables that affect the outcome but are also correlated with the treatment - that cannot be controlled for by regression adjustment or matching. Obtaining credibly causal estimates is particularly critical for policy design. Policy-makers need to know whether observed outcome differences can be attributed to the policy of interest and not to the influence of confounding variables.

Why might public good exposure be endogenous to the outcomes of interest? There are several classic examples in environmental economics. The first example comes from studies of the effect of air pollution on health outcomes. Locations with higher levels of pollution tend to also be more urbanized and face different economic and social conditions (e.g., average income and crime rates). Such differences may lead to sorting of households with different underlying health levels and preferences for clean air into locations with different pollution levels. If so, the difference in observed health outcomes by air pollution status may reflect both the effect of air pollution on health and the effect of unobserved household differences.

Identification problems also arise in cross-sectional hedonic estimates of the effect of environmental attributes on land values. For local temperature studies, cross-sectional temperature and land values may be correlated with unobserved differences in soil quality or in the option value of converting farmland to alternative use. If so, resulting estimates of the temperature-land value gradient may suffer from omitted variable bias by including both the causal effect of temperature and that of the unobserved determinant. Reverse causality may also be an issue. Consider a hedonic study of the effects of local air pollution on land values. If increasing land values induce local polluters to move, leading in cleaner air, a regression of land values on local pollution would not reflect the correct direction of causality.

A final example comes from the literature on the impact of environmental regulations on workers' employment status and earnings. Here, the identification challenge is that workers employed in plants subjected to the regulation may have different unobservable characteristics than workers employed in plants not subjected to the regulation. For example, workers employed in polluting plants may have different skill levels and labor market experience. As a result, the OLS estimate of the earnings differential between workers across regulated and non-regulated plants may reflect differences in other characteristics in addition to the effect of the regulation.

In each of these examples, identification of causal effects using observational data requires a research design that is experimental or quasi-experimental, where public good exposure is "as good" as randomly assigned. We now turn to Section 3 which describes how standard quasi-experimental approaches address these endogeneity concerns in general settings that employ unit-level quasi-random assignment. This is then followed by Section 4 which shows that resulting estimates may still be biased in public good settings due to externality spillovers.

3   The standard quasi-experimental approach

The identification problem appearing in non-experimental research designs arises because treatment assignment is not random and thus may be correlated with unobserved determinants of an outcome under study. In recent years, a growing literature in environmental economics has attempted to address such identification problems by making use of quasi-experiments. This section first defines the standard quasi-experimental approach broadly and then formally within the potential outcomes framework.

3.1   Background

We begin by broadly defining the scope of empirical methods, which we label as "quasi-experimental." The term quasi-experiment originates in Campbell and Stanley (1963) who wrote:

> "There are many natural social settings in which the research person can introduce something like experimental design into his scheduling of data collection procedures (e.g., the when and to whom of measurement), even though he lacks the full control over the scheduling of experimental stimuli (the when and to whom of exposure and the ability to randomize exposures) which makes a true experiment possible. Collectively, such situations can be regarded as quasi-experimental designs." (Campbell and Stanley (1963), pg. 34)

In economics, the terms quasi-experiment and natural experiment are often used interchangeably.[7] The presumed equivalence of natural and quasi experiments in economic research is stated in DiNardo (2008)'s

---

[7] The distinction between quasi-experiment (where the treatment can be manipulated, e.g., a policy reform) and natural

New Palgrave Dictionary of Economics entry: "Natural experiments or quasi-natural experiments ... are also serendipitous situations where assignment to treatment 'approximates' randomized design or a well-controlled experiment." This is echoed in Angrist and Pischke (2008) who write: "We hope to find natural or quasi-experiments that mimic a randomized trial by changing the variable of interest while other factors are kept balanced."

Another useful definition of quasi-experiments (inclusive of natural experiments) appears in Stock and Watson (2017):

> "In a quasi-experiment, also called a natural experiment, randomness is introduced by variations in individual circumstances that make it appear as if the treatment is randomly assigned. These variations in individual circumstances might arise because of vagaries in legal institutions, location, timing of policy implementation, natural randomness such as birth dates, rainfall, or other factors unrelated to the causal effect under study." (Stock and Watson (2017), pg. 493)

3.2 Potential outcomes framework

We now formally show how quasi-experimental observational data settings provide identification of causal effects within a standard potential outcomes framework. To present the problem in its most general form for now, we do not consider any particularities that arise from a public good setting. We expand the standard framework to accommodate public goods in Section 4.

Consider N units which can be households, firms, or locations. For each unit $i = 1, \ldots, N$, there are two potential outcomes, each associated with a binary status of being exposed and not exposed to the public good, $Q_i \in \{0,1\}$.[8] Define the set of potential outcomes for unit i under the two exposure statuses as $\mathcal{Y}_i \in \{Y_{0i}, Y_{1i}\}$. We place no homogeneity restrictions on $\mathcal{Y}_i$, which is important for public good settings discussed in Section 4. The N-component vectors of treatment assignment and potential outcomes are $\mathbf{Q}$, $\mathbf{Y_0}$, and $\mathbf{Y_1}$.

In many applied microeconomic settings, the causal estimand of interest is the population average treatment effect (ATE):

$$\tau = \frac{1}{N} \sum_{i=1}^{N} (Y_{1i} - Y_{0i}) \tag{2}$$

where $Y_{1i} - Y_{0i}$ is the unit-level treatment effect. The empirical challenge is that it is impossible to directly estimate τ as the researcher never observes both potential outcomes for the same unit. This is often called the fundamental problem of causal inference (Holland, 1986).

An assumption is needed in order to define the mapping between potential outcomes and observed outcome. The stable-unit treatment value assumption (SUTVA) requires that there is no interference between units and that there are no hidden variations in the treatments (Imbens and Rubin, 2015). Under SUTVA, the potential outcomes for unit i depends only on own treatment and not on the treatment of other units. This implies the following definition for the observed outcome Y_i

---

experiment (where the treatment cannot be manipulated, e.g., an earthquake) is made more explicit by certain authors (Cook, Campbell and Shadish, 2002).

[8] We consider a binary treatment variable for ease of exposition. There are generally analogues for applications with multivalued treatments (e.g., the average causal response vs. the average treatment effect) but these extensions are beyond the scope of this paper.

$$Y_i = \begin{cases} Y_{0i} & \text{if } Q_i = 0 \\ Y_{1i} & \text{if } Q_i = 1 \end{cases}$$
$$= Q_i Y_{1i} + (1 - Q_i) Y_{0i} \tag{3}$$

Suppose there are a fixed number of M treated units. Furthermore, for simplicity, suppose the potential outcomes are also fixed such that assignment of $\mathbf{Q}$ is the only source of uncertainty in the observed outcomes (see Section 3.4 of Imbens and Rubin (2015)). A common estimator for τ compares average observed outcomes between treated and control units. Formally, it is

$$\hat{\tau} = \frac{1}{M} \sum_{i:Q_i=1} Y_i - \frac{1}{(N-M)} \sum_{i:Q_i=0} Y_i$$
$$= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Q_i Y_{1i}}{M/N} - \frac{(1-Q_i)Y_{0i}}{(N-M)/N} \right) \tag{4}$$

where the second line applies equation (3). Denote $Pr(\mathbf{Q}|\mathbf{Y_0}, \mathbf{Y_1})$ as the assignment mechanism. Imbens and Rubin (2015) define a completely randomized experiment as having an assignment mechanism that (i) has a known functional form; (ii) is probabilistic, with unit-level probability of treatment between 0 and 1[9]; (iii) is individualistic, with unit-level probability of treatment that is independent across units, and (iv) is unconfounded, meaning that the probability of treatment does not depend on potential outcomes.

In most empirical applications in environmental economics, the research design is non-experimental and instead uses observational data. In observational studies, the researcher does not know the functional form of $Pr(\mathbf{Q}|\mathbf{Y_0}, \mathbf{Y_1})$. However, provided the assignment mechanism is still probabilistic, individualistic, and unconfounded, $\hat{\tau}$ can still be an unbiased estimator for the population average treatment effect, τ. To see that, one can take the expected value of $\hat{\tau}$ over all realizations of treatment assignment vectors $\mathbf{Q}$

$$\mathbb{E}_{\mathbf{Q}}[\hat{\tau}|\mathcal{Y}_i] = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Y_{1i}\mathbb{E}_{\mathbf{D}}[Q_i|\mathcal{Y}_i]}{M/N} - \frac{Y_{0i}\mathbb{E}_{\mathbf{D}}[(1-Q_i)|\mathcal{Y}_i]}{(N-M)/N} \right)$$
$$= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Y_{1i}Pr(Q_i=1)}{M/N} - \frac{Y_{0i}(1-Pr(Q_i=1))}{(N-M)/N} \right)$$
$$= \tau$$

where the second line applies probabilistic, unconfounded, and individualistic assignment such that $\mathbb{E}_{\mathbf{Q}}[Q_i|\mathcal{Y}_i] = Pr(Q_i = 1|\mathcal{Y}_i) = Pr(Q_i = 1) = \frac{M}{N}$. In this paper, we formally define studies using observational data with probabilistic, unconfounded, and individualistic treatment assignment as quasi-experimental. The most common concern with observational non-experimental data is whether the unconfoundedness assumption is satisfied, which requires that treated and control units are otherwise similar on average if not for the assignment of treatment. To address concerns about unconfoundedness, the quasi-experimental literature commonly employs three methods to ensure valid comparisons, which we now briefly summarize.

---

[9]More formally, a probabilistic assignment requires that all members of the population can potentially be exposed to all possible values of a treatment with non-zero probability.

3.3 Three quasi-experimental methods

Difference-in-difference (DID) and panel fixed effects If treatment and control units differ because of time-invariant unobserved characteristics, researchers can assume common time trends in these characteristics. The difference-in-difference (DID) method can be applied to settings in which some units experience a change in treatment status over time while other units do not.[10] That is, rather than comparing average outcomes between treated and control units, the DID method compares average change in outcomes between treated and control units over time. Often the change in treatment status arises because of a change in policy, making the method well-suited for measuring the impact of regulations or policies that change over time for some units, but not for others. The key identifying assumption is that the time evolution of outcomes in the control group provides a valid counterfactual for the time evolution of outcomes in the treatment group absent the treatment. This is often referred to as the "common trends" assumption. With more than one period of pre-treatment observations, this assumption can be indirectly tested by examining whether trends in outcomes are similar before the change in treatment status.

A generalized version of the DID method is the panel data fixed effects model. Here, units are observed over multiple time periods where treatment status changes over time for some units and not for others in a more arbitrary manner, not necessarily as the result of a "discrete" change in policy. When the treatment effect is constant across groups and over time, panel data FE regressions that include fixed effects for units and time periods identify the population average treatment effect under the common trends assumption.

Instrumental variables (IV) The instrumental variables method attempts to isolate the exogenous component of treatment assignment in order to identify a causal effect. Specifically, it makes use of a measurable relationship between the treatment of interest and an exogenous variable, or instrument. Two requirements must be met in order for the IV method to provide causal estimates. First, the instrument must be relevant in the sense that the relationship between the instrument and treatment variable, known as the first stage relationship, must be sufficiently strong and monotonic. Second, the instrument must be exogenous in the sense that it affects the outcome variable only through the endogenous treatment variable. Under these conditions, the IV method estimates the local average treatment effect for the subpopulation who experiences a change in treatment status as a result of the instrument.

Regression discontinuity (RD) In some settings, treatment assignment may follow a discontinuous rule: units with values of a "forcing" variable above some threshold receive the treatment while those with values below the threshold do not. Such a setting, which often emerges from specific policies or institutional rules, allows for the regression discontinuity (RD) design. A typical RD study compares the outcomes of units with values of the forcing variable that are "near" either sides of the threshold. The key requirement for the RD method is that only the probability of receiving the treatment jumps discontinuously as the forcing variable crosses the threshold.[11] All other factors that determine the outcome must be continuous around the threshold. Under these conditions, the RD method estimates the local average treatment effect for the subpopulation close to the threshold.

---

[10] The introduction of DID method in the seminal papers of Ashenfelter (1978) and Ashenfelter and Card (1985) began the "natural experiment" revolution in labor economics.

[11] Note that because of externality spillovers, it may not be appropriate to apply the "spatial" RD method for public goods where the forcing variable is some form of spatial distance.

4　The quasi-experimental approach for public goods

Section 3 presented the standard quasi-experimental approach for addressing endogeneity concerns in observational data settings. In this section, we discuss what happens when the standard quasi-experimental approach is applied to public goods treatments, which, by definition, exhibit externalities. To start, we return to the problem of optimal public good provision summarized by the the Lindahl-Samuelson condition of Section 2.1. In particular, estimates of the marginal social benefit of a public good, captured by the left hand side of equation (1) requires knowing: (i) the number of agents exposed to the public good and, (ii) the population average causal effect of exposure to the public good. The latter is the estimand of interest. As noted in Section 2.3, OLS regression of any welfare-relevant outcome on public good exposure may produce biased estimates due to non-random assignment of treatment. The application of standard quasi-experimental methods discussed in Section 3.3 aims to address this identification problem by exploiting quasi-random treatment assignment. However, because public goods exhibit externalities, a new issue arises that potentially complicates identification. Specifically, one must consider the distinction between *unit-level source* and the resulting *group-level exposure* of a public good.

This section formalizes how this distinction affects identification by first extending the potential outcomes framework from Section 3.2 to explicitly account for externality spillovers. As with the standard quasi-experimental approach, we assume that assignment of the public good source is probabilistic, individualistic, and unconfounded, or "as good" as randomly assigned. However, these assumptions are not sufficient for identifying the population average treatment effect in the presence of externality spillovers. Instead, we show that two commonly used quasi-experimental estimators may be biased, though for different reasons. From this identification framework emerges a weighted estimator that is unbiased, with unit-level weights based on the likelihood of receiving externality spillovers. The section concludes by illustrating the potential bias in each estimator via simulations.

## 4.1　Distinguishing public good source and exposure

The standard potential outcomes framework presented in Section 3.2 assumes that treatment assignment is individualistic, or that the likelihood of receiving treatment is independent across units. There is a further assumption of the Stable Unit Treatment Value Assumption (SUTVA) such that treatment applied to one unit does not affect outcomes of other units. In many empirical settings, these assumptions are typically innocuous as there is often no need to distinguish between source and exposure of a treatment.

This distinction is essential when the treatment of interest is a public good. The "point-source" nature of public goods is consistent with unit-level assignment. For example, air pollution abatement policy in a given location reduces pollution in that location. However, because public goods exhibit externalities, public good sourced from one location leads to changes in exposure over a group of locations. When a policy reduces pollution from a location, all downwind locations are jointly exposed to cleaner air, irrespective of whether they directly receive the policy.

Figure 1 illustrates this issue for a simple 2-dimensional setting with 3×3 locations. Public good produced in one location spills over to all downwind locations according to the map of prevailing wind directions in the left panel. The middle and right panels of Figure 1 illustrates the distinction between public good source (in orange, or dark shading) and exposure (in hatched lines) when the public good is sourced from two particular locations. When a location is assigned to be a public good source, exposure to the public good is experienced by both the treated location and all downwind locations. Observe that for this particular wind pattern, the

likelihood of receiving an externality spillover is heterogeneous across locations, with the bottom right corner unit experiencing the highest likelihood of spillovers. We return to this important feature later.

Figure 1: Illustrating public good source and exposure in the presence of spillovers
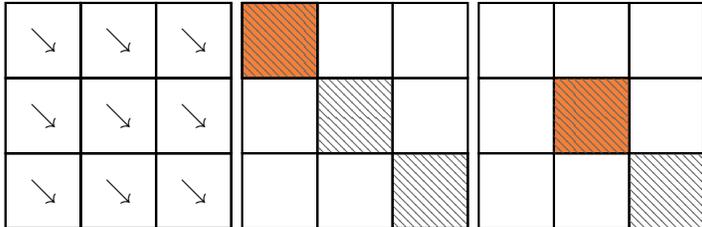
NOTES: The left panel shows the direction of public good spillovers. The middle and right panels show a different spatial configuration of public good source and exposure when different locations serve as the public good source. Locations in orange, or dark shading, are assigned to be the public good source. Locations with hatched lines are exposed to the public good both directly and via spillovers.

We have thus far discussed how unit-level public good exposure does not occur in observational data settings. It should also be noted that unit-level public good exposure in lab or field-based experiments may not be practical or necessarily policy-relevant. Experiments to protect individuals from pollution exposure in the field or lab may be prohibitively costly or unethical.[12] Furthermore, even if unit-level treatments were feasible, the resulting estimates of the marginal social benefit of a public good may not be policy relevant. Returning to our air pollution example, if health-related expenditures were used to recover the marginal willingness to pay for pollution abatement, the relevant estimate should include any price effects from a downward shift in aggregate demand for health-producing goods as downwind individuals jointly experience cleaner air. Experimental estimates based on unit-level pollution exposure would not capture such price effects.

To formalize the identification challenges for public goods, we extend the potential outcomes framework presented in Section 3.2 to explicitly account for externality spillovers.

4.2 A potential outcomes framework for public goods

Rather than assuming quasi-experimental assignment of public good exposure $\mathbf{Q}$ as in Section 3.2, we now consider quasi-experimental assignment of public good source. Define $\mathbf{D}$ as the vector of public good sources. A typical element of $\mathbf{D}$ is a Bernoulli random variable, $D_i \in \{0, 1\}$. Location i is the source of a public good when $D_i = 1$. We assume the standard conditions for a quasi-experimental setting, with assignment mechanism for $\mathbf{D}$ being probabilistic, individualistic, and unconfounded. The number of units assigned to be sources, $L < N$, is fixed. Potential outcomes are also fixed such that assignment of $\mathbf{D}$ is the only driver of uncertainty in observed outcomes. Altogether, this implies that the expected value of D_i over all draws of $\mathbf{D}$ is $\mathbb{E}_{\mathbf{D}}[D_i | \mathcal{Y}_i] = Pr(D_i = 1 | \mathcal{Y}_i) = Pr(D_i = 1) = \frac{L}{N}$, where the equalities apply the probabilistic, unconfounded, and individualistic assignment mechanism assumptions, respectively.

The vector of public good sources, $\mathbf{D}$, maps onto the vector of exposure, $\mathbf{Q}$. For simplicity, we assume that unit i's exposure is also a Bernoulli random variable, $Q_i \in \{0, 1\}$. Location i is exposed to the public

---

[12]For example, it would be costly to exclusively restrict subjects to environmentally-controlled indoor settings.

11

good when $Q_i = 1$. $\mathbf{Q}$ is related to the source vector $\mathbf{D}$ in the following way

$$\mathbf{Q} = \mathbf{1}\{\mathbf{W'D} > 0\} \tag{5}$$

where $\mathbf{W}$ is a deterministic N-by-N adjacency matrix. The bilateral weight w_{ki} characterizes the transport of a public good that is sourced from unit k and exposes "downwind" unit i. $\mathbf{W}$ is sometimes referred to as the source-receptor matrix in environmental economics.

We place two restrictions on $\mathbf{W}$. First, we assume unit diagonal elements, $w_{ki} = 1 \ \forall k = i$. This implies $Pr(Q_i = 1 | D_i = 1, \mathcal{Y}_i) = 1$, which is consistent with the non-exclusive nature of public goods: conditional on being a public good source, that unit is also exposed to the public good. Second, we allow for weakly positive off-diagonal elements, $w_{ki} \geq 0 \ \forall k \neq i$. We place no additional structure on these off-diagonal elements such that conditional on not being a public good source, the probability of public good exposure varies across units. Denote this likelihood of externality spillover as $Pr(Q_i = 1 | D_i = 0, \mathcal{Y}_i) = S_i$. This probability of externality spillover captures the degree of SUTVA violation. It implies that the probability of being exposed to the public good Q_i is no longer individualistic, even if assignment of D_i is. The probability of spillover S_i plays a crucial role in identification of the population average treatment effect, which we discuss in detail below. These two restrictions on $\mathbf{W}$ create four categories of units, with associated conditional probabilities of exposure displayed in Table 1.

Table 1: Public good source (D_i) and exposure (Q_i)

|		D_i	
		0	1
Q_i	0	$Pr(Q_i = 0 | D_i = 0, \mathcal{Y}_i) = 1 - S_i$	$Pr(Q_i = 0 | D_i = 1, \mathcal{Y}_i) = 0$
	1	$Pr(Q_i = 1 | D_i = 0, \mathcal{Y}_i) = S_i$	$Pr(Q_i = 1 | D_i = 1, \mathcal{Y}_i) = 1$

Having established this notation, we now discuss whether two commonly used quasi-experimental estimators in environmental economics identify the population average treatment effect, τ, from equation (2).

4.3 Two quasi-experimental estimators in the literature

4.3.1 Average source effect estimator

The average source effect estimator (ASEE) compares average outcomes across units that are and are not assigned to be a public good source. A classic example of the ASEE in environmental economics are comparisons of average outcomes across counties designated as nonattainment and attainment under the US Clean Air Act. Figure 2 illustrates the comparison across units that underlies the ASEE for the spatial configuration of public good assignments shown in the middle and right panels of Figure 1. Red (dark-shaded) cells are locations that are the source of the public good with $D_i = 1$. Blue (light-shaded) cells are locations that are not the source of the public good with $D_i = 0$. The ASEE is essentially a comparison of average outcomes across red (dark-shaded) and blue (light-shaded) cells in Figure 2.

Figure 2: Illustrating the average source effect estimator
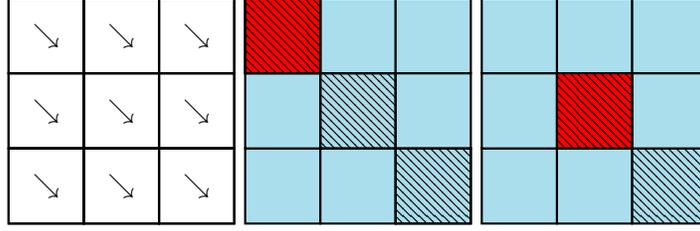
NOTES: The left panel shows the direction of public good spillovers. The middle and right panels show comparisons for the ASEE corresponding to the public good source assignments shown in the middle and right panels of Figure 1. Red (dark-shaded) cells represent locations that are public good sources with $D_i = 1$. Blue (light-shaded) cells represent locations that are not public good sources with $D_i = 0$. Cells with hatched lines show locations that are exposed to the public good both directly and via spillovers.

Formally, the ASEE is

$$
\begin{aligned}
\hat{\tau}^S &= \frac{1}{L} \sum_{i:D_i=1} Y_i - \frac{1}{(N-L)} \sum_{i:D_i=0} Y_i \\
&= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{D_i(Q_i Y_{1i} + (1-Q_i)Y_{0i})}{L/N} - \frac{(1-D_i)(Q_i Y_{1i} + (1-Q_i)Y_{0i})}{(N-L)/N} \right)
\end{aligned} \tag{6}
$$

where the second line applies equation (3). The expected value of $\hat{\tau}^S$ over all $\mathbf{D}$ realizations is

$$
\begin{aligned}
\mathbb{E}_{\mathbf{D}}[\hat{\tau}^S|\mathcal{Y}_i] &= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Y_{1i}\mathbb{E}_{\mathbf{D}}[D_i Q_i|\mathcal{Y}_i]}{L/N} + \frac{Y_{0i}\mathbb{E}_{\mathbf{D}}[D_i(1-Q_i)|\mathcal{Y}_i]}{L/N} \right. \\
&\left. \quad - \frac{Y_{1i}\mathbb{E}_{\mathbf{D}}[(1-D_i)Q_i|\mathcal{Y}_i]}{(N-L)/N} - \frac{Y_{0i}\mathbb{E}_{\mathbf{D}}[(1-D_i)(1-Q_i)|\mathcal{Y}_i]}{(N-L)/N} \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} (1-S_i)(Y_{1i} - Y_{0i})
\end{aligned} \tag{7}
$$

where the first line expands the expression for the estimator. The second line applies $\mathbb{E}_{\mathbf{D}}[D_i Q_i|\mathcal{Y}_i] = \mathbb{E}_{\mathbf{D}}[D_i|\mathcal{Y}_i] = \frac{L}{N}$ and $\mathbb{E}_{\mathbf{D}}[Q_i|\mathcal{Y}_i] = \frac{L+S_i(N-L)}{N}$.

The externality associated with public good source D_i leads to spillovers in exposure, Q_i, as characterized by equation (5). A comparison of average outcomes between units that are and are not the source of the public good yields a biased estimate of the population average treatment effect τ because some units in the control group also experience changes in public good exposure due to externality spillovers. These units are shown in gray hatched lines in Figure 2. This non-individualistic assignment of public good exposure constitutes a violation of SUTVA. The degree of SUTVA violation is captured by S_i, the probability that unit i is exposed to the public good conditional on not being a public good source. Since S_i is bounded between 0 and 1, the sign of the bias in $\hat{\tau}^S$ is known.

It is helpful to consider three cases. In the first case, Q_i is a private good and all off-diagonal terms in the matrix $\mathbf{W}$ are zero so that there are no externality spillovers. In that case, we return to the standard unit-level randomization setting with $S_i = 0 \ \forall i$ and the ASEE is unbiased, $\mathbb{E}_{\mathbf{D}}[\hat{\tau}^S|\mathcal{Y}_i] = \tau$. The second case is the other extreme where Q_i is a global public good and all off-diagonal terms in $\mathbf{W}$ are positive. Because the entire population is now exposed to the public good with $S_i = 1 \ \forall i$, we have $\mathbb{E}_{\mathbf{D}}[\hat{\tau}^S|\mathcal{Y}_i] = 0$. Finally,

there is the intermediate case with a local public good where some off-diagonal elements in $\mathbf{W}$ are zero and others are positive such that $S_i \in [0, 1] \ \forall i$. In that case, $0 < \mathbb{E}_{\mathbf{D}}[\hat{\tau}^S | \mathcal{Y}_i] < \tau$. In summary, externalities from local public goods generate spillovers that biases estimates from the ASEE towards zero.

4.3.2 Average exposure effect estimator

Another common quasi-experimental estimator compares average outcomes across units that are and are not exposed to the public good arising from quasi-random assignment of public good source $\mathbf{D}$. We call this the average exposure effect estimator (AEEE). A typical AEEE study regresses outcomes on pollution that originates from elsewhere. Figure 3 illustrates how the AEEE compares units for the spatial configuration of public good assignments shown in the middle and right panels of Figure 1. Brown (dark-shaded) cells are locations that are exposed to the public good with $Q_i = 1$. Gray (light-shaded) cells are locations that are not exposed to the public good with $Q_i = 0$. The AEEE takes the difference in average outcomes across brown (dark-shaded) and gray (light-shaded) cells.

Figure 3: Illustrating the average exposure effect estimator
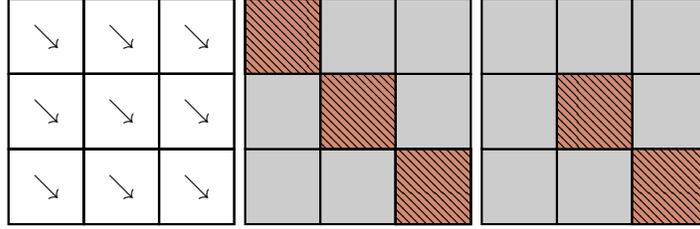
NOTES: The left panel shows the direction of public good spillovers. The middle and right panels show comparisons for the AEEE corresponding to the public good source assignments shown in the middle and right panels of Figure 1. Brown (dark-shaded) cells represent locations that are exposed to the public good with $Q_i = 1$. Gray (light-shaded) cells represent locations that are not exposed to the public good with $Q_i = 0$. Cells with hatched lines show locations exposed to the public good both directly and via spillovers.

For simplicity, suppose the number of units receiving the public good, M, is fixed, and $L \leq M \leq N$. Formally, the AEEE is

$$\hat{\tau}^E = \frac{1}{M} \sum_{i:Q_i=1} Y_i - \frac{1}{(N-M)} \sum_{i:Q_i=0} Y_i$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Q_i Y_{1i}}{M/N} - \frac{(1-Q_i)Y_{0i}}{(N-M)/N} \right) \quad (8)$$

where the second line applies equation (3). The expected value of $\hat{\tau}^E$ over all $\mathbf{D}$ realizations is

$$\mathbb{E}_{\mathbf{D}}[\hat{\tau}^E | \mathcal{Y}_i] = \frac{1}{N} \sum_{i=1}^{N} \left(Y_{1i} \frac{\mathbb{E}_{\mathbf{D}}[Q_i | \mathcal{Y}_i]}{M/N} - \frac{Y_{0i} \mathbb{E}_{\mathbf{D}}[(1-Q_i) | \mathcal{Y}_i]}{(N-M)/N} \right)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(Y_{1i} \left(\frac{L + S_i(N-L)}{M} \right) - Y_{0i} \left(\frac{(N-L)(1-S_i)}{N-M} \right) \right) \quad (9)$$

where the second line applies the probabilities in Table 1 and $\mathbb{E}_{\mathbf{D}}[Q_i | \mathcal{Y}_i] = \frac{L+S_i(N-L)}{N}$. Equation (9) shows that the AEEE heterogeneously weights potential outcomes for each unit. The weight on Y_{1i} is the combined

14

likelihood of experiencing public good exposure from being the source, $\frac{L}{M}$, and from receiving spillovers, $\frac{S_i(N-L)}{M}$. The weight on Y_{0i} is the likelihood of jointly not being the public good source and not receiving spillovers, $\frac{(N-L)(1-S_i)}{N-M}$.

Again, it is useful to consider three cases. First, in the private good setting with no spillovers, the number of units receiving exposure equals the number of source units, $M = L$, and $S_i = 0 \ \forall i$. In that case, the AEEE is unbiased, with $\mathbb{E}_{\mathbf{D}}[\hat{\tau}^E | \mathcal{Y}_i] = \tau$. Second, in the case of a global public good, all units become exposed such that $M = N$ and $S_i = 1 \ \forall i$. As a consequence, $\mathbb{E}_{\mathbf{D}}[\hat{\tau}^E | \mathcal{Y}_i]$ becomes undefined.

For the intermediate of a local public good with $L < M < N$ and $S_i \in [0, 1]$, equation (9) shows that the AEEE places greater relative weight on Y_{1i} than on Y_{0i} as S_i, the likelihood of exposure from spillovers, increases. As a consequence, the AEEE is biased if potential outcomes and the likelihood of experiencing a public good spillover are correlated, or formally whenever $\frac{1}{N} \sum_{i=1}^N Y_{1i} S_i \neq (\frac{1}{N} \sum_{i=1}^N Y_{1i})(\frac{1}{N} \sum_{i=1}^N S_i)$ and $\frac{1}{N} \sum_{i=1}^N Y_{0i} S_i \neq (\frac{1}{N} \sum_{i=1}^N Y_{0i})(\frac{1}{N} \sum_{i=1}^N S_i)$. This is a form of selection bias. In our air pollution example, poorer households may tend to live downwind from polluting plants (Heblich, Trew and Zylberberg, 2016). Thus, even if pollution abatement policy were quasi-randomly assigned, sorting implies a greater likelihood that poorer households are exposed to a change in air pollution. Unlike the ASEE, unless the exact nature of sorting is known, the sign of the bias in the AEEE is generally ambiguous.

When is the AEEE unbiased? It is unbiased when potential outcomes are uncorrelated with the likelihood of experiencing a spillover. To see this, observe that with fixed number of public good source and exposed units, we have $\sum_{i=1}^N \mathbb{E}_{\mathbf{D}}[(1 - D_i)Q_i] = M - L$. Applying $\mathbb{E}_{\mathbf{D}}[D_i Q_i | \mathcal{Y}_i] = \mathbb{E}_{\mathbf{D}}[D_i | \mathcal{Y}_i] = \frac{L}{N}$ and $\mathbb{E}_{\mathbf{D}}[Q_i | \mathcal{Y}_i] = \frac{L + S_i(N-L)}{N}$, we have

$$\frac{1}{N} \sum_{i=1}^N S_i = \frac{M - L}{N - L} \tag{10}$$

This implies that equation (9) can be further reduced to

$$
\begin{aligned}
\mathbb{E}_{\mathbf{D}}[\hat{\tau}^E | \mathcal{Y}_i] &= \frac{1}{N} \sum_{i=1}^N \left(Y_{1i} \left(\frac{L + S_i(N-L)}{M} \right) - Y_{0i} \left(\frac{(N-L)(1-S_i)}{N-M} \right) \right) \\
&= \frac{L}{M} \left(\frac{1}{N} \sum_{i=1}^N Y_{1i} \right) + \frac{N-L}{M} \left(\frac{1}{N} \sum_{i=1}^N Y_{1i} \right) \left(\frac{1}{N} \sum_{i=1}^N S_i \right) \\
&\quad - \frac{N-L}{N-M} \left(\frac{1}{N} \sum_{i=1}^N Y_{0i} \right) + \frac{N-L}{N-M} \left(\frac{1}{N} \sum_{i=1}^N Y_{0i} \right) \left(\frac{1}{N} \sum_{i=1}^N S_i \right) \\
&= \frac{1}{N} \sum_{i=1}^N (Y_{1i} - Y_{0i}) = \tau
\end{aligned}
$$

where the second equality applies the assumption that potential outcomes are uncorrelated with spillover likelihood, or $\frac{1}{N} \sum_{i=1}^N Y_{1i} S_i = (\frac{1}{N} \sum_{i=1}^N Y_{1i})(\frac{1}{N} \sum_{i=1}^N S_i)$ and $\frac{1}{N} \sum_{i=1}^N Y_{0i} S_i = (\frac{1}{N} \sum_{i=1}^N Y_{0i})(\frac{1}{N} \sum_{i=1}^N S_i)$. The third equality applies equation (10). In summary, even if public good sources were quasi-randomly assigned, the AEEE is biased for local public goods when potential outcomes are correlated with the likelihood of externality spillovers. This bias may be due to selection on observables or on unobservables. In the case of selection on observables, we discuss how one can test for bias in the AEEE in Section 6.1. Of course, there are no direct remedies for selection on unobservables.

4.4 An unbiased estimator for local public goods

Equation (7) suggests that a simple re-weighting of observations will allow the ASEE to produce an unbiased estimate of τ. Specifically, one can divide observations by $1 - S_i$, the probability of not receiving the externality spillover conditional on not being the public good source. This weighting scheme will down weight control units that are likely to experience spillovers while up weight those that are less likely to experience spillovers. Specifically, the weighted average source effect estimator (WASEE) is

$$
\begin{aligned}
\hat{\tau}^W &= \frac{1}{L} \sum_{i:D_i=1} \frac{Y_i}{(1-S_i)} - \frac{1}{(N-L)} \sum_{i:D_i=0} \frac{Y_i}{(1-S_i)} \\
&= \frac{1}{N} \sum_{i=1}^N \left(\frac{D_i(Q_i Y_{1i} + D_i(1-Q_i)Y_{0i})}{(1-S_i)L/N} - \frac{(1-D_i)(Q_i Y_{1i} + (1-Q_i)Y_{0i})}{(1-S_i)(N-L)/N} \right)
\end{aligned} \tag{11}
$$

The same derivation leading to equation (7) shows that $\mathbb{E}_{\mathbf{D}}[\hat{\tau}^W | \mathcal{Y}_i] = \tau$ for local public goods. Unfortunately, as with the AEEE, the WASEE is undefined in the presence of a global public good.

Equation (11) seems to suggest that if unit-level spillover likelihoods were available, a researcher should always implement the WASEE, especially since the possibility of selection on unobservables will generate bias in the AEEE. Unfortunately, the ASEE and by extension, the WASEE may not be possible to implement in all empirical settings. Consider again the case of air pollution regulation. Suppose the subpopulation of interest are households residing in locations exposed to air pollution. However, these locations do not contain polluting firms and thus would never be pollution sources. In such a setting, only the AEEE can be implemented for the subpopulation of interest and the only remedy available to a researcher is to test for selection on observables, as will be discussed in Section 6.1.

4.5 Illustrative simulations

We now turn to simulations to illustrate the potential biases in our three estimators: ASEE ($\hat{\tau}^S$), AEEE ($\hat{\tau}^E$), and WASEE ($\hat{\tau}^W$). To avoid unnecessary complications, we consider the simple geography displayed in Figure 1, with $N = 9$ units located on a 3×3 grid. For each policy realization, one randomly chosen unit receives the public good, $D_i = 1$.

Under the spatial pattern of externality spillover shown in the left panel of Figure 1, the bottom right corner unit experiences the highest likelihood of spillovers, S_i. To examine what happens when spillover likelihood and potential outcomes are correlated, we consider two configurations of baseline potential outcomes $Y_{0i} \in \{1, \ldots, 9\}$. In the first configuration, the bottom right unit has $Y_{0i} = 1$ while elements in the set $\{2, \ldots, 9\}$ are randomly assigned to Y_{0i} of other units. This generates a negative correlation between Y_{0i} and S_i. The second configuration exhibits positive correlation between Y_{0i} and and S_i by assigning the bottom right unit $Y_{0i} = 9$ with elements in the set $\{1, \ldots, 8\}$ randomly assigned to Y_{0i} of other units.

The simulation draws 10,000 realizations of $\mathbf{D}$ with the population average treatment effect set at $\tau = 5$ and assumed to be constant across units. Table 2 shows the expected value for the ASEE, AEEE, and WASEE when Y_{0i} is negatively correlated $(-)$ and positively correlated $(+)$ with spillover likelihood, S_i. When treatment involves a private good, that is when off-diagonal terms in the weighting matrix $\mathbf{W}$ are zero and $S_i = 0 \; \forall i$, all three estimators are unbiased and the spatial arrangement of baseline characteristics does not matter. When the treatment involves a local public good, the ASEE and AEEE are both biased. The ASEE is biased towards 0 due to a SUTVA violation. This bias does not depend on whether baseline characteristics, Y_{0i} are correlated to spillover probabilities. The AEEE is also biased and the direction of

Table 2: Sample means of ASEE and AEEE estimates based on simulated data

	$\mathbb{E}_{\mathbf{D}}[\hat{\tau}_-^S|\mathcal{Y}_i]$	$\mathbb{E}_{\mathbf{D}}[\hat{\tau}_+^S|\mathcal{Y}_i]$	$\mathbb{E}_{\mathbf{D}}[\hat{\tau}_-^E|\mathcal{Y}_i]$	$\mathbb{E}_{\mathbf{D}}[\hat{\tau}_+^E|\mathcal{Y}_i]$	$\mathbb{E}_{\mathbf{D}}[\hat{\tau}_-^W|\mathcal{Y}_i]$	$\mathbb{E}_{\mathbf{D}}[\hat{\tau}_+^W|\mathcal{Y}_i]$
Private good	4.99	4.99	4.99	4.99	4.99	4.99
Local public good	4.64	4.65	3.9	5.4	4.99	4.99

NOTES: Table 2 shows estimates for the ASEE ($\hat{\tau}^S$), AEEE ($\hat{\tau}^E$), and WASEE ($\hat{\tau}^W$) averaged over 10,000 simulations when baseline potential outcomes, Y_{0i}, are negatively ($-$) or positively ($+$) correlated with likelihood of spillovers, S_i. The private good simulations assume the off-diagonal terms in the weighting matrix $\mathbf{W}$ are zero. The local public good simulations use $\mathbf{W}$ based on the spatial pattern of externality spillovers shown in the left panel of Figure 1.

this bias is negative when the likelihood of spillover effects is negatively correlated with baseline potential outcomes and positive when spillover likelihood is positively correlated with baseline potential outcomes. Table 2 also shows that estimates from the WASEE are always unbiased. The empirical distribution of the ASEE and AEEE estimates under the two correlation structures are shown in Figure 4.

Figure 4: Distribution of estimates of ASEE and AEEE for different configuration of baseline characteristics
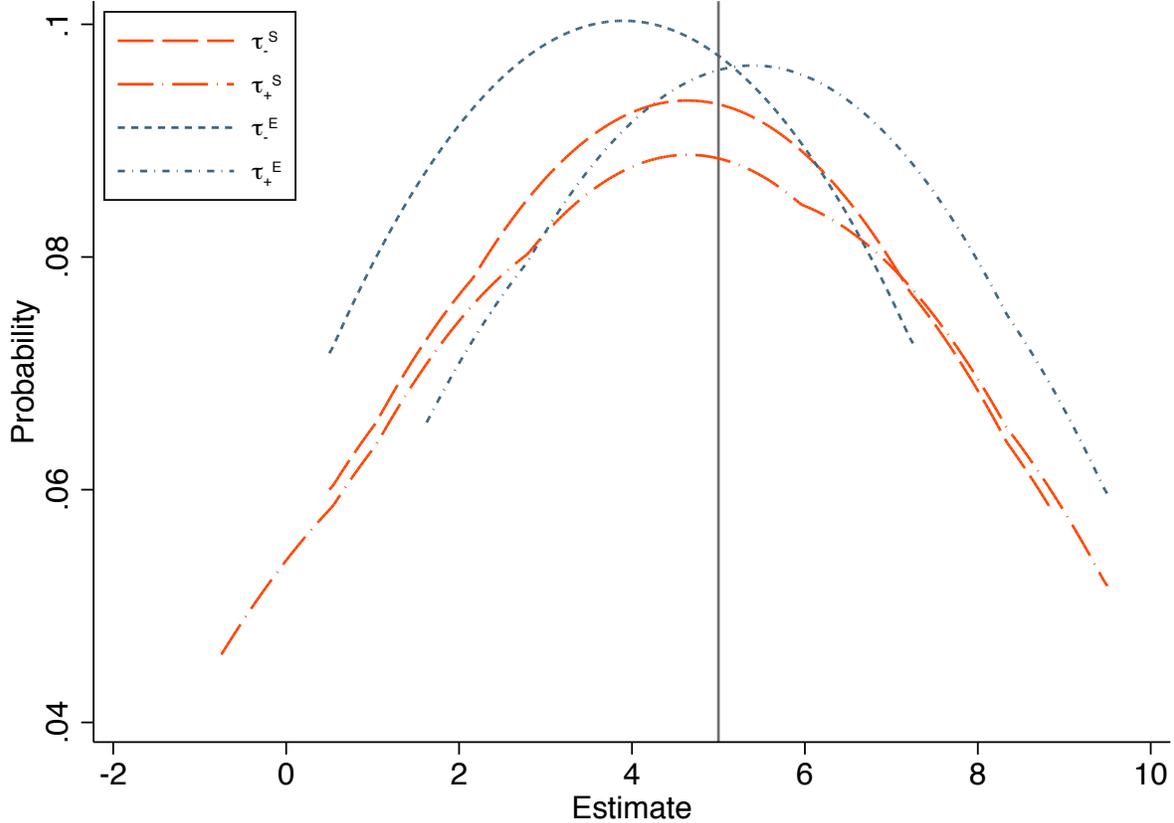
NOTES: Figure 4 shows the empirical distribution of the estimates for ASEE ($\hat{\tau}^S$) and AEEE ($\hat{\tau}^E$) for 10,000 simulations when baseline potential outcomes, Y_{0i}, are negatively ($-$) or positively ($+$) correlated with likelihood of spillovers, S_i.

5   Literature review

This section reviews published environmental economics papers that apply quasi-experimental methods. We begin by summarizing recent publication trends. We then review select papers that employ the average source effect estimator and average exposure effect estimator to obtain the marginal social benefits of environmental goods. Finally, we also review quasi-experimental papers that estimate the marginal cost of environmental good provision.

5.1   Publication trends

We searched for environmental economics papers using quasi-experimental methods published between 2000-2017 in the Journal of Environmental Economics and Management, the Journal of Public Economics, and 10 general interest economics journals.[13] To perform this query, we searched for articles in the Econlit database using an extensive list of keywords.[14] This initial search produced a total of 557 papers. We then manually examined each article to determine if they employed quasi-experimental methods, and then classified them according to the method used: difference-in-difference, panel fixed effect, instrumental variables, and regression discontinuity.[15] We acknowledge from the onset that our search criteria and classification method may have omitted environmental economics papers in these journals that used quasi-experimental methods. Furthermore, articles published in other economics journals or journals outside of economics were excluded from this review.

We begin by looking at the relative role of quasi-experimental methods in environmental economics. Figure 5 plots the annual share of papers using quasi-experimental methods among the set of environmental economics papers identified through our search criteria. Amongst the journals we consider, quasi-experimental methods have become increasingly prominent over the last two decades. The first quasi-experimental environmental economics paper identified in our literature review is "Effects of Air Quality Regulations on Polluting Industries" by Vernon Henderson and Randy Becker, published in the Journal of Political Economy in 2000. In 2017, 32% of published papers used quasi-experimental methods.

Figure 6 shows just the numerator from Figure 5, i.e., the annual number of published papers in environmental economics using quasi-experimental methods. Following a relatively slow start in the 2000s, publications pick up around 2010 (one year after the introduction of the American Economic Journals), reaching a peak of 18 papers in 2017. Over this period, a cumulative total of 90 quasi-experimental papers have been published. Figure 6 also breaks down annual published papers by quasi-experimental method.

Table 3 shows the number of environmental economics papers and the subset of papers using quasi-experimental methods by journal and method during the 2000-2017 period. The Journal of Environmental Economics and Management has the most with 387 papers returned by our search criteria, of which 38 employed quasi-experimental methods. The next highest are the American Economic Journals, which have published 24 environmental economics papers, of which 13 were quasi-experimental. No quasi-experimental

---

[13]The general interest journals are: American Economic Review (excluding papers published in the May Papers & Proceedings), the American Economic Journals, Econometrica, the Journal of Political Economy, the Quarterly Journal of Economics, the Review of Economics and Statistics, and The Review of Economic Studies. We omit the Journal of the Association of Environmental and Resource Economists, an important field journal that only began publication in 2014.

[14]The full list of title words is: environment, environmental, resource, natural, air, water, climate, temperature, pollution, agriculture, land, forest, deforestation, and energy. The search was conducted in February 2018.

[15] We classify papers as using the difference-in-difference method when the treatment and control groups in the analysis are defined on the basis of a discrete environmental policy or regulation. We classify papers as panel fixed effects in cases where the causal effect of interest is identified from within unit comparisons of changes in treatment status over time, where the changes in the treatment status are more arbitrary (i.e., not tied to a policy or regulation).

Figure 5: Trend in the share of published environmental economics papers using quasi-experimental methods
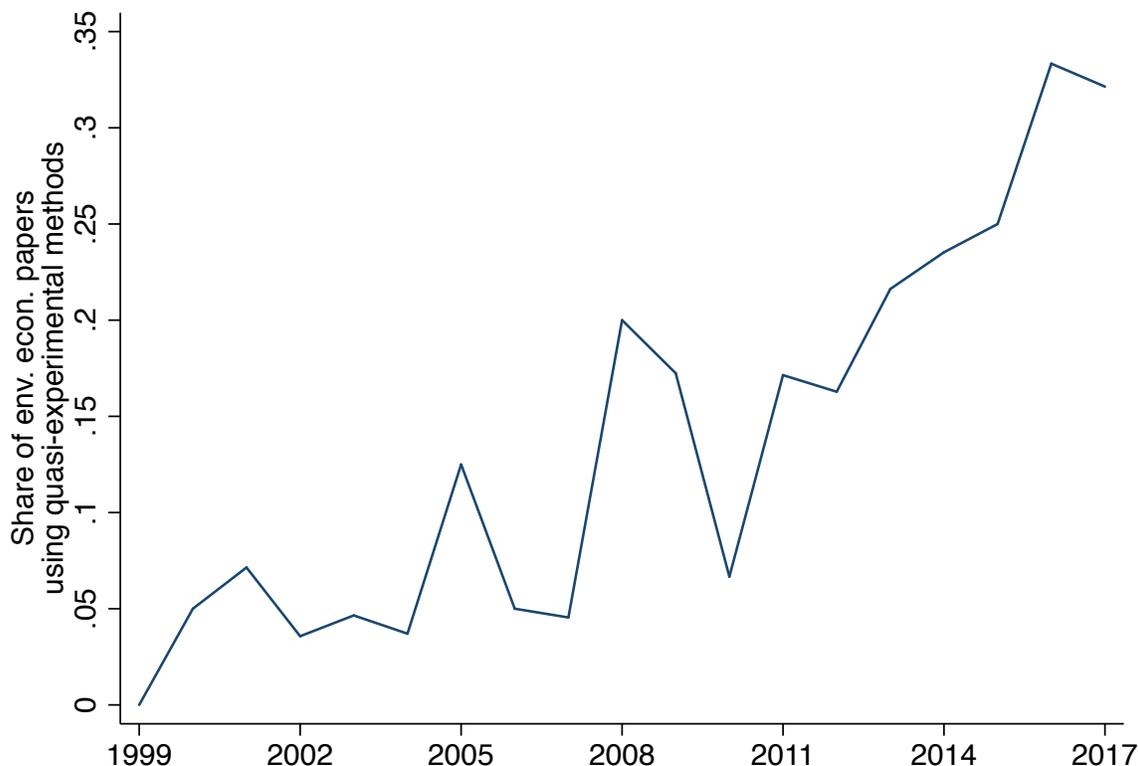
NOTES: Share of published environmental economics papers using quasi-experimental methods year of publication. Journals include American Economic Journals, American Economic Review, Econometrica, Journal of Environmental Economics and Management, Econometrica, the Journal of Political Economy, the Journal of Public Economics, the Quarterly Journal of Economics, the Review of Economics and Statistics, and the Review of Economic Studies.
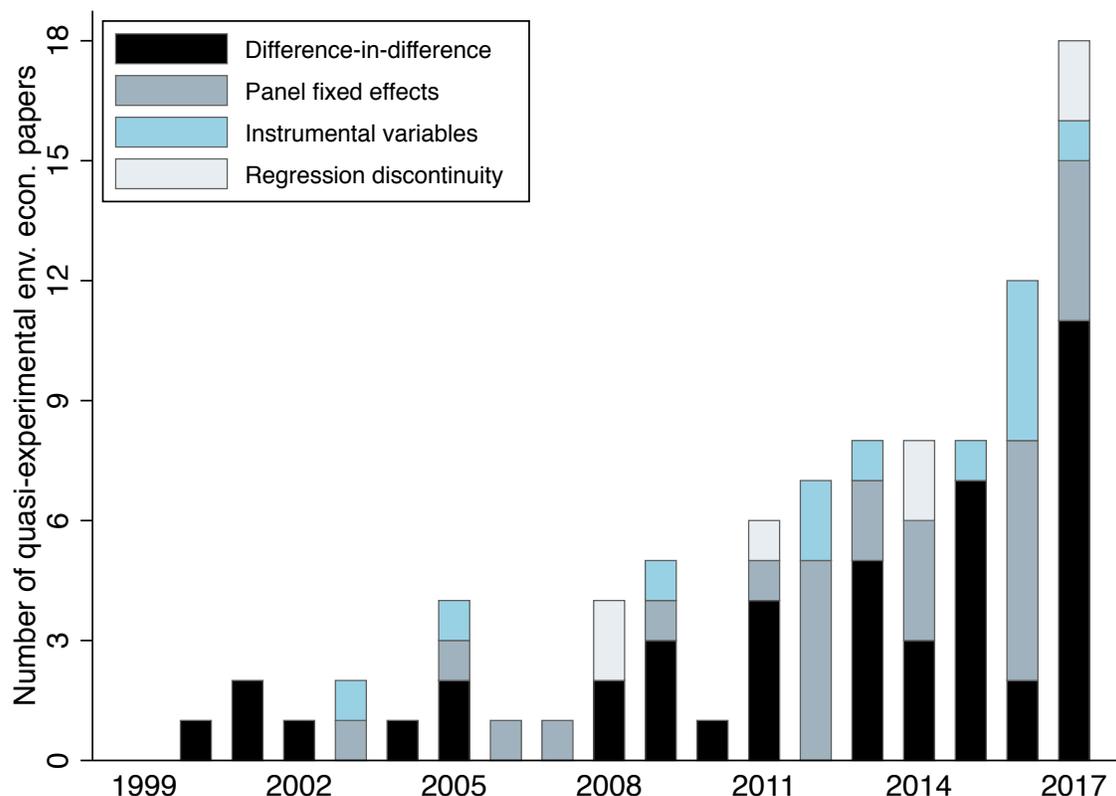
papers in environmental economics were published in Econometrica and only 2 were published in The Review of Economics Studies. Across the four quasi-experimental methods, difference-in-difference (including triple-difference and synthetic control methods) is the most popular, accounting for 50% of all quasi-experimental papers. Panel fixed effects, instrumental variables, and regression discontinuity account for 29%, 13%, and 8% of all environmental economics papers in these journals, respectively.

Figure 7 organizes our subset of published quasi-experimental papers in environmental economics (total=90) into 4 groups based on the estimator, parameter of interest, environmental good, and outcome under study. These 4 categories account for 45, 14, 12, and 19 papers, respectively. The largest category, Box A, contains papers that use the ASEE to estimate the marginal social benefit of environmental policies. Outcomes examined in this part of the literature are (i) air pollution (11 papers); (ii) labor market outcomes and productivity (7 papers), health outcomes (4 papers), housing values (3 papers), and other.[16]

The papers in Box B estimate the marginal social benefit of air pollution, using the AEEE. These papers rely on sources of quasi-experimental variation in the source of air pollution originating from other locations.

---

[16]These other outcomes include environmental expenditures, patents, electricity consumption, and foreign direct investment (20 papers).

Figure 6: Trends in the number of published quasi-experimental environmental economics papers by method

NOTES: The number of published papers using quasi-experimental methods in environmental economics by year of publication and method. Journals include American Economic Journals, American Economic Review, Econometrica, Journal of Environmental Economics and Management, Econometrica, the Journal of Political Economy, the Journal of Public Economics, the Quarterly Journal of Economics, the Review of Economics and Statistics, and the Review of Economic Studies.

For example, some papers use the opening and/or closing of polluting plants nearby to generate exogenous local variation in pollution exposure. Other papers use traffic patterns, network congestion, and public transport strikes elsewhere as instrumental variables for local air pollution exposure. The main outcome variables in this group are health outcomes (6 papers), housing values (3 papers), labor market outcomes and productivity (3 papers), and school outcomes (2 papers).

Box C summarizes an emerging literature on the economic effects of temperature fluctuations. These papers use the AEEE with the goal of informing the marginal social benefit of greenhouse gas abatement by estimating the damage associated with higher temperatures. The 12 published studies generally focus on identifying the effect of temporal variation in temperature for a given location across periods (e.g., days, months, years). The primary outcomes studied are agricultural outcomes (4 papers), health outcomes (4 papers), and other outcomes such as GDP growth rate (4 papers).

The last category, represented by Box D, is composed of papers that concern other important parameters in environmental economics as well as other environmental goods. In addition to estimating the marginal social benefit of environmental goods, these quasi-experimental papers also examine the marginal cost of

Table 3: Number of published papers in environmental economics by journal and method

	All EE papers	Quasi-experimental EE papers				
		All QE methods	DID	Panel FE	IV	RD
American Economic Journals	24	13	6	6	0	1
American Economic Review	31	10	5	4	0	1
Econometrica	1	0	0	0	0	0
Journal of Env. Econ. and Mgmt	387	38	23	9	4	2
Journal of Political Economy	12	6	2	1	2	1
Journal of Public Economics	53	6	3	1	1	1
Quarterly Journal of Economics	7	5	1	2	1	1
Review of Economics and Statistics	37	10	5	2	3	0
Review of Economic Studies	5	2	0	1	1	0
TOTAL	**557**	**90**	**45**	**26**	**12**	**7**

NOTES: Published papers in environmental economics by journal and method during 2000-2017.

environmental policy as well as various related political economy questions. These papers also study other environmental goods such as toxic chemicals and water pollution. One striking observation from Figure 7 is just how few quasi-experimental studies there are of environmental attributes besides air pollution and temperature. Of the 90 identified quasi-experimental papers published between 2000-2017, only a handful of papers examine other environmental goods. The general lack of quasi-experimental evidence for other environmental goods suggests potential opportunities for future research.

5.2 A selected review of average source effect estimates

The average source effect estimator (ASEE) compares units that are and are not the source of an environmental good that is quasi-randomly assigned. The canonical application of this estimator exploits the introduction of a spatially-differentiated environmental policy, such as national environmental policies with sufficiently intricate local implementation to estimate the policy's marginal social benefit. In such settings, the assignment of local regulation is argued to be exogenous to local economic circumstances. Here, we review several prominent papers in this literature.

The Clean Air Act (CAA) is the primary national policy in the United States that regulates sulfur dioxide, particulates, nitrogen dioxide, carbon monoxide, ozone, and lead air pollution. Beginning with the seminal papers by Henderson (1996) and Becker and Henderson (2000), researchers have since examined over five decades of the CAA, and its amendments in 1970, 1977 and 1990. Most empirical analyses of the CAA employ the ASEE by comparing average outcomes between counties that are more regulated under "non-attainment" status with those that are less regulated under "attainment" status. The predominant

Figure 7: Breakdown of the quasi-experimental literature in environmental economics by parameter, estimator, treatment, and outcome
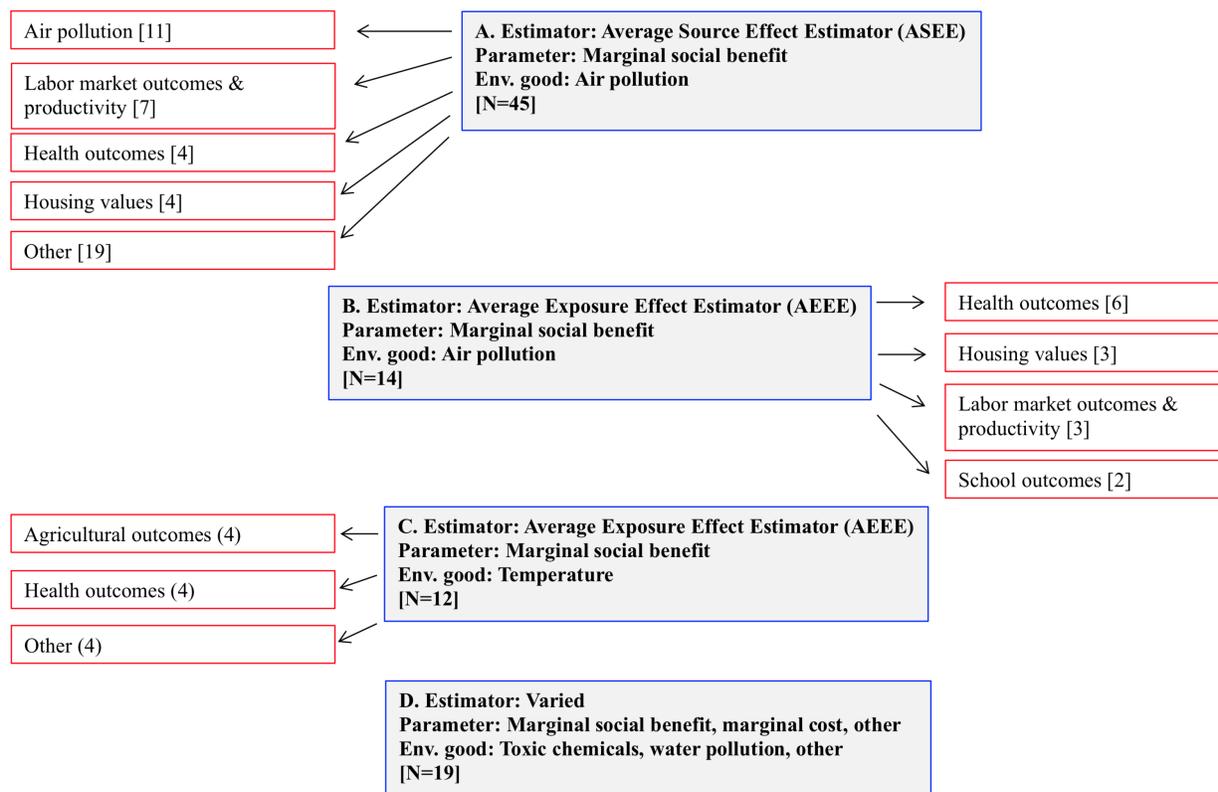
identification strategy, pioneered by Chay and Greenstone (2005), uses a difference-in-difference research design, sometimes combined with an instrumental variables approach.

These papers typically begin with a first-stage analysis estimating the change in ambient pollution before and after the introduction of a particular CAA amendment and between counties that are designated non-attainment and attainment. For example, Chay and Greenstone (2005) find the implementation of the first amendments to the Clean Air Act in the early 1970s led to a large relative change in TSP concentrations in non-attainment counties. In fact they attribute virtually all the national-level decline in TSP concentrations to the CAA regulation. Auffhammer, Bento and Lowe (2009) study the 1990 CAA and find substantial heterogeneity in the effect of this regulation on PM_{10} concentrations.[17] In particular, they find no effect of the regulation on concentrations measured at the typical non-attainment county. However, there is an 11-14% reduction in PM_{10} concentrations for monitors located in high-pollution non-attainment counties.[18]

The second-stage analysis then proceeds by estimating how changes in local air pollution induced by regulations such as the CAA affects an economic outcome that help to recover the marginal social benefit of improving air quality. To do this, an instrumental variables approach is usually implemented on a first-differenced regression equation relating changes in the economic outcome in a location over time to corresponding changes in measured air pollution that is instrumented by early period non-attainment status

---

[17]Under the CAA, particulate matter was denoted as total suspended particulates (TSP) until 1987 when EPA began explicitly regulating PM2.5 and PM10.

[18]Another part of this literature on the effect of environmental regulations on air quality examines transportation policies (e.g., Auffhammer and Kellogg (2011), Davis (2008)).

in that location. Thus the identifying assumption is that early period non-attainment status only affects an outcome through its effect on the regulated air pollutant. A violation of this assumption would occur if early period non-attainment status led to changes in the outcome of interest through a channel other than air pollution, or if early period non-attainment status affects more than one pollutant. Since non-attainment status is a function of lagged air pollution levels, the assumption requires that unobserved shocks in the process under study are unrelated to past air pollution levels that determine non-attainment status. For example, non-attainment status may affect air pollution and local labor demand, thereby making the identification strategy invalid for estimating the effect of pollution on housing values, since the housing market may respond to local labor market conditions.

Motivated by hedonic theory suggesting that willingness to pay for non-market amenities can be inferred from housing values, the primary outcomes studied in this part of the quasi-experimental literature are housing values or rents.[19] Chay and Greenstone (2005) present the first such study using quasi-experimental methods. Focusing on changes in average housing values for U.S. counties between 1970 and 1980 and using TSP non-attainment status in 1975 to instrument for pollution changes, they report estimates indicating that one $\mu g/m^3$ reduction in TSP caused a 0.2 to 0.4% increase in housing values. This corresponds to elasticities ranging from -0.20 to -0.35. Chay and Greenstone (2005) also note that their IV estimates are remarkably larger than previous non-experimental estimates and attribute this difference to omitted variables bias.

Grainger (2012) and Bento, Freedman and Lang (2015) extend the Chay and Greenstone (2005) approach to study the 1990 CAA amendments and exploit spatial heterogeneity in the impact of the amendments on air pollution. An innovation in Grainger (2012) is to estimate separate models for rents and owner-occupied housing values. He finds that a one $\mu g/m^3$ reduction in PM_{10} leads to a 1% to 2% decrease in owner-occupied housing values. As expected, the corresponding estimates for rents are significantly smaller. Bento, Freedman and Lang (2015) further extend this idea by using track-level data and by estimating separate models for housing units located "closer" to and "farther" from PM_{10} monitoring stations. They find that most of the benefits of PM_{10} reductions concentrations accrue to housing units located within 5 miles of monitoring stations.

This approach for measuring the benefits of air quality improvements has also been extended to studies of the contemporaneous effect of air pollution on health (Chay, Dobkin and Greenstone, 2003; Chay and Greenstone, 2003; Sanders and Stoecker, 2015), and to studies of the long-run effect of air pollution on labor market outcomes such as earnings and employment (Isen, Rossin-Slater and Walker, 2017). Specifically, Isen, Rossin-Slater and Walker (2017) relate earnings (measured around age 30) to average TSP levels in the county and year of birth for a large sample of U.S. workers, and instrument for birth-year TSP levels with the early 1970s non-attainment status of each county. They those exposed to lower concentrations of TSP in early life as a result of the 1970 CAA amendments earn about 1% more on average 30 years later.

The pollutant of interest in much of the quasi-experimental CAA literature is particulate matter (or TSP). In the context of the identification framework in Section 4, environmental externality spillovers likely result in estimates of CAA affects that are biased towards zero. However, a focus on particulate matter over other regulated ambient pollutants has a distinct advantage: particulates tends to disperse locally. For example, a power plant in a non-attainment county may emit particulate matter that travels to an adjacent county, especially if that plant is located near the county border. However, it is unlikely for particulates

---

[19] Quasi-experimental estimates of housing values allow welfare interpretations only under certain assumptions, as summarized by Kuminoff, Smith and Timmins (2013). The literature offers various tests of these assumptions within quasi-experimental settings (Klaiber and Smith, 2013; Gamper-Rabindran and Timmins, 2013; Kuminoff and Pope, 2014).

to continue onto counties that are further away. This implies that the downward bias in the ASEE for particulate matter from a SUTVA violation will be smaller than for pollutants traveling greater distances.

Deschenes, Greenstone and Shapiro (2017) estimates a ASEE on the direct health impacts and defensive expenditures related to nitrogen oxides (NO_x) and ozone pollution. Specifically, they estimate the health impact of EPA's NO_x Budget Program (NBP), a cap-and-trade market affecting over 2,500 electricity generating units and industrial boilers located in the Eastern and Midwestern United States operating between 2003 and 2008.[20] An important feature of this market is that it was only operational during the "ozone summer season," defined as May 1st to September 30th of each year.

Deschenes, Greenstone and Shapiro (2017) first reports triple-difference estimates of the impact of the cap-and-trade market on NO_x emissions, ambient ozone concentrations, and health outcomes. Using county-by-season-by-year level data for 2,500 U.S. counties, they find that NBP led to a 40% decline in NO_x emissions during the ozone summer season as well as large reductions in high ozone days and sizable health benefits as captured by reductions in medication expenditures and premature mortality.

While such reduced-form analysis can be interpreted under reasonably weak assumptions (e.g., the absence of county-by-season-by-year level unobserved shocks), Deschenes, Greenstone and Shapiro (2017) also implement an instrumental variable strategy that follows in the spirit of prior CAA studies. Using the implementation of the NBP market as an instrument for NO_x and ozone requires that the NBP market does not affect other determinants of health outcomes. Unlike previous instrumental variables studies of air pollution, Deschenes, Greenstone and Shapiro (2017) report first stage estimates associated with the implementation of NBP for all air pollutants regulated under the CAA. They find that the reduction in NO_x emissions caused by the regulation drives virtually all of the changes in ambient ozone concentrations. They also find less robust evidence that the NBP market led to changes $PM_{2.5}$ and therefore caution readers about interpreting the resulting instrumental variable estimates in light of this potential violation of the exogeneity assumption. The fact that environmental policies may simultaneously affect more than one pollutant, thereby invalidating the exclusion restriction, should be taken on more directly in future research.

In contrast to particulate pollution, NO_x can travel large distances and create a SUTVA violation if downwind locations are assigned as control units.[21] Deschenes, Greenstone and Shapiro (2017) indirectly address this concern by estimating a coarse version of the weighted ASEE, introduced in equation (11). Specifically, they drop states from the sample that are immediately adjacent to states regulated under NBP to minimize concerns about pollution spillover.

The ASEE has also been applied in various developing country settings. Greenstone and Hanna (2014) study the effect of major air and water quality policies in India using city-level data over 1986-2007. Using a difference-in-difference estimator to compare outcomes within Indian cities pre- and post-regulation adoption, they find that air quality regulations (in particular the mandate for new vehicles to adopt catalytic converters) led to large improvements in air quality (PM_{100} and SO_2). However, water pollution regulations had no detectable effect on water quality (e.g., biochemical oxygen demand and dissolved oxygen). Overall the reductions in air pollution caused by regulations in India did not lead to significant reductions in infant mortality rates.

Air quality regulations are sometimes introduced as temporary measures. An interesting example is the series of air pollution regulations implemented by the Chinese government in Beijing and surrounding cities

---

[20]The NBP market later became part of the Clean Air Interstate Rule, which was replaced by the Cross-State Air Pollution Rule in 2011.

[21]For example, according to Husar and Renard (1997), on the 96 percent of days in the NBP region where wind speeds are below 6 meters per second, ozone and its precursors can travel up to 300 miles.

starting around the time of the 2008 Beijing Olympics. Such regulations required the temporary shutdown of polluting industries, the installation of pollution control equipment at coal-fired power plants, and the implementation of traffic control policies. Because of air pollution spillovers, neighboring cities and provinces were also required to implement pollution control programs. He, Fan and Zhou (2016) document the effects of these regulations on air quality and mortality rates. Using city-month data for 34 large cities in China and the 2008 regulation as an instrument for PM_{10}, they find that PM_{10} concentrations fell by 25 $\mu g/m^3$, or roughly 25%, as a result of these policies. They also find a positive and statistically significant effect of PM_{10} concentration on mortality from cardiovascular disease.

5.3 A selected review of average exposure effect estimates

The average exposure effect estimator (AEEE) compares units that are and are not exposed to the public good. Quasi-experimental AEEE implicitly assume that the origin source of this environmental good is sufficiently "far-away" such that whatever caused it has otherwise no direct effect on local outcomes. Here, we briefly discuss two broad sets of AEEE applications looking at the effects of ambient air pollution and local temperature on various outcomes.

Graff Zivin and Neidell (2012) study the effect of daily ambient ozone pollution on the labor productivity of agricultural workers in California's Central Valley. Their setting has three main empirical advantages. First, outdoor agricultural field workers, in this case blueberry pickers, are directly exposed to ambient pollution. Second, their productivity is quantifiable in terms of fruit harvested per hour and observable through payroll systems. Third, the authors argue that local ozone fluctuations over California's Central Valley are plausibly exogenous to unobserved local determinants. This is because ozone is not directly emitted, but rather produced through nonlinear interactions between local environmental factors such as temperature and sunlight and the presence of nitrogen oxides (NO_x) and volatile organic chemicals ($VOCs$). In particular, it is assumed that conditional on controls, the factors causing NO_x and $VOCs$ emissions from distant coastal locations and transported to the Central Valley are unrelated to local outcomes in the Central Valley.

Graff Zivin and Neidell (2012) combine daily data from the 2009 and 2010 growing seasons on 1,600 workers with daily ozone concentrations measured from nearby agricultural sites. Using a panel fixed effects regression with worker fixed effects, they find that a 10 ppb decrease in daily ozone concentration leads to a 5.5% increase in worker productivity, although some point estimates are not statistically significant. This approach for documenting productivity benefits of environmental regulation has since been extended to other countries, other air pollutants, indoor office settings, and manufacturing settings (e.g., Chang et al. (2016) and He, Fan and Zhou (2016)). Neidell (2017) presents an overview of this literature.

Another prominent application of the AEEE is Schlenker and Walker (2016) who study the effects of carbon monoxide (CO), nitrogen dioxide (NO_2), and ozone air pollution generated by major airports in California on hospital visit rates for households living near airports. Schlenker and Walker (2016) devise an identification strategy that exploits local air pollution shocks generated by conditions elsewhere. Specifically, they observe that plane taxi times in large California airports is driven, in part, by congestion occurring at large Eastern U.S. airports. These delays serve as instruments for cumulative taxi times in California airports and are unlikely to directly affect local health outcomes near California airports.[22]

---

[22] The use of congestion from other parts of a transportation network to derive instruments for local air pollution has appeared elsewhere in the literature (e.g., Moretti and Neidell (2011), using boat traffic in ports, and Knittel, Miller and Sanders (2016), using road traffic).

Schlenker and Walker (2016) first document that taxi times in remote Eastern U.S. airports drive local air pollution in large California airports. They find that a 1,000 minute increase in daily taxi time increases ambient CO concentrations close to airports by 45 ppb, an 8% increase relative to the mean daily concentration. They then report IV estimates of the effect of CO on hospital visits related to respiratory and heart diseases and find that a one standard deviation increase in CO concentrations leads to a 21% increase in daily hospital admissions for asthma, and to an 18% increase in daily hospital admissions for heart disease.

A third air pollution example AEEE is the Currie et al. (2015) study which examines the effects of toxic hazardous airborne pollutants (HAPs) such as benzene, cumene, and nickel. Currie et al. (2015) estimate the economic costs associated with HAPs by comparing housing values and infant health across households exposed to different HAPs levels.[23] Specifically they combine data on toxic releases by individual industrial plants from the Toxic Release Inventory database, housing transactions, and infant health at a high degree of spatial granularity from Texas, New Jersey, Pennsylvania, Michigan, and Florida. Their difference-in-difference strategy exploits the timing of "toxic" plant openings and closing and compares outcomes for locations that are within 0.5 miles (treatment) vs 1-2 miles (control) away from the opening and/or of a closing toxic plant. This spatial distinction between treatment and control groups emerges from first empirically examining the spatial decay of hazardous pollution as a function of radial distance from a toxic plant. In this context, the identifying assumption is that unobserved circumstances leading to a toxic plant opening or closing has a uniform effect on housing market and infant health trends across treatment and control groups. Currie et al. (2015) finds empirical support for this assumption by failing to detect differential pre-trends in outcomes across treatment and control groups.

Using this approach, Currie et al. (2015) find that toxic emissions have sizable negative effects on housing values. Specifically, they estimate that the opening of a toxic plant lowers housing prices by 11% while plant closings have little effect on housing values. They interpret this as suggesting that the presence of a toxic plant continues to create disamenities even after operation ceases, perhaps because of concerns about long-term environmental contamination. Finally, Currie et al. (2015) also find that the opening and closing of toxic plants reduces the probability of low birth weight for infants whose mother lives within 1 mile of a toxic plant.

In all these studies, there is an acknowledged understanding that estimates pertain to particular subpopulations and may not generalize to a broader population. Graff Zivin and Neidell (2012) note that "While the impacts of ozone on agricultural productivity are large, the generalizability of these findings to other pollutants and industries is unclear." Similarly Schlenker and Walker (2016) observe that zip codes closer to a major airport in California are more urban, more populated, wealthier, and have higher housing prices than zip codes from the rest of California, writing "Therefore, we would caution against interpreting the estimated dose–response relationship as representative for the entire population at large." Finally, the subpopulation of interest in the Currie et al. (2015) study is U.S. residents living near industrial plants that emit toxic pollutants. Indeed, they report that locations that are further away from an industrial plant have higher housing values and different maternal characteristics.

Our concerns about potential selection bias in the AEEE still hold for estimands defined over subpopulations. For example, in the case of air pollution from industrial plants, rather than define the set of locations that emit air pollution as the population of interest, one can instead consider a subpopulation of locations that contain industrial plants. The estimand of interest will then be the average treatment effect for this

---

[23]Currie et al. (2015) is part of a rich literature in environmental economics that studies the economic, social, and health outcomes associated with the pollutants generated by industrial activity in the United States and elsewhere. See for example, Bui and Mayer (2003), Davis (2011), and Greenstone and Gallagher (2008).

subpopulation. Identification of this subpopulation estimand using the AEEE requires that the likelihood of spillovers, S_i, be uncorrelated with potential outcomes, Y_{1i} and Y_{0i}.

We conclude this section with papers that apply the AEEE to examine the effects of atmospheric variables, such as temperature and precipitation, on various economic and social outcomes. These papers are primarily motivated by an interest in informing the marginal social benefit of climate change policy. Before describing these papers, it is first important to distinguish between climate and weather. In atmospheric physics, climate typically describes the distribution of an atmospheric variable while weather denotes a particular realization of that distribution. The distribution of interest may be defined either across space or time. For example, for a given time period, the climate can be summarized by the average temperature across locations and weather is the temperature in a particular location. Similarly, for a given location, the climate can be described by the average temperature over time while weather is temperature at a single period.

Anthropogenic climate change (ACC) alters the distribution of atmospheric conditions in both spatial and temporal dimensions. ACC poses a particular challenge for quasi-experimental methods along the spatial dimension. It is the result of greenhouse gas emissions, a global pollutant by virtue of its atmospheric mixing and long decay properties. When a unit of greenhouse gas is emitted into the atmosphere anywhere on the planet, it changes local atmospheric conditions everywhere. There is no control group as every location on the planet is exposed. As we showed in Section 4.3.2, for a global public good the ASEE produces an estimate of and both the AEEE and WASEE are undefined. Thus, the key question along the spatial dimension is how to interpret estimates based on local weather as proxies for global ACC impacts. That is, how relevant are comparisons across locations that experience different weather conditions when the ultimate policy question of interest involves a global pollutant that simultaneously changes the global climate. We defer this important issue for now and return to it in Section 6.2.

Much of the existing AEEE literature estimating the effects of atmospheric variables focuses on the distinction between climate and weather along the temporal dimension. In particular, these papers have explored whether quasi-experimental estimates include adaptation behavior in anticipation of expected changes in average temperature over any location. These studies have been labeled by Dell, Jones and Olken (2014) as the "New Climate-Economy Literature."

The quasi-experimental literature using local weather variability begins with the Deschenes and Greenstone (2007) study of how weather shocks affect U.S. agricultural profits. The paper was motivated by the seminal cross-sectional Ricardian approach of Mendelsohn, Nordhaus and Shaw (1994). Deschenes and Greenstone (2007) argue that such cross-sectional estimates of the effect of temperature on agricultural land values were not identified by quasi-experimental variation and so possibly biased. This bias would arise in the land value application when cross-sectional average temperatures are correlated with unobserved time-invariant determinants of land values such as unmeasured soil quality or the potential for conversion to non-agricultural land use.

Deschenes and Greenstone (2007) propose using year-to-year variation in temperature and rainfall to estimate effects on agricultural profits. Under the assumption that temporal changes in weather in a given location are as good as randomly assigned, identification of the temperature effect on profits (or any other economic outcome) is straightforward via an application of the panel fixed effects method.

Deschenes and Greenstone (2007) implement this approach using a county-year panel for the United States over 1987-2002 and report a statistically insignificant relationship between weather shocks and U.S. agricultural profits. Based on that, they argue that climate change would have a relatively small economic impact on the U.S. agricultural sector, especially since damages measured from short-run variation are likely

to overstate long-term damages. This paper has spawned a large subsequent literature on various outcomes beside agricultural outcomes including 12 of the published papers tabulated in Section 5.1 and numerous other papers not reviewed here.[24]

The findings in Deschenes and Greenstone (2007) attracted interest and also criticism. In particular, Fisher et al. (2012) identifies errors in the data used by Deschenes and Greenstone (2007) and argues as a result that the negative impact of climate change on agriculture was larger than the estimate reported in the original study. In their response to Fisher et al. (2012), Deschenes and Greenstone (2012) correct their estimates to account for such errors and use a more recent set of projections from the CCSM model (unavailable when the initial paper was published in 2007). Based on that, Deschenes and Greenstone (2012) conclude that the present discounted value of lost agricultural profits due to climate change over the next 90 years was $164 billion (in 2002 dollars), or about 5 times annual U.S. agricultural profit.

Burke and Emerick (2016) proposed an extension of the original analysis by using long-differences (i.e., defined over 20 years) instead of the 5-year differences used in Deschenes and Greenstone (2007). The motivation for using longer differences is that it leverages quasi-experimental variation in temperature trends over longer time periods. In principle, this longer time horizon allows for a wider range of possible adaptation behavior to offset temperature effects provided that realized trends were indeed anticipated by agents ex-ante. In their application for the U.S. farming sector, Burke and Emerick (2016) find similar effects of high temperatures on corn and soybeans yields using the long and short difference approaches. They interpret this finding as evidence of limited longer term adaptation by farmers to mitigate the negative effects of high temperature on crop yields in the shorter term.[25]

Deschenes and Greenstone (2011) provides the first comprehensive study of the health cost of climate change accounting for both the direct cost in terms of reduced health, and the economic cost of adaptation, as approximated by electricity consumption. The paper uses a county-by-year panel for the U.S. over 1968-2002 to estimate the relationship between mortality rates (including age- and cause-specific mortality rates) and the daily distribution of temperature realizations, captured by the number of days per year that fall in various daily temperature categories. The benchmark regression model include county and state-by-year fixed effects, so the identification comes from temperature deviations relative to county and state-by-year averages. This variation is assumed to be uncorrelated with unobserved determinants of health.[26] Using this variation, Deschenes and Greenstone (2011) find evidence of heat- and cold-related mortality: each day with temperature above 90°F and below 40°F is associated with statistically significant increases in the annual mortality rate. Using the same type of panel fixed effect regression as the one described above, they detect a similar "flat U" relationship between energy expenditures and temperature extremes. Interestingly, the adaptation response to temperature variability (in proportional terms) is four times larger than the mortality response to the same temperature variability, suggesting that household defensive investments are important. The Deschenes and Greenstone (2011) analysis was expanded in subsequent studies to further consider the role of humidity in shaping the temperature-mortality relationship (Barreca, 2012), to examining cross-sectional differences in the temperature-mortality relationship across climates (Barreca et al., 2015; Heutel, Miller and Molitor, 2017), and across long time periods (Barreca et al., 2016).

An important finding in the Barreca et al. (2016) study is the remarkable reduction in the effect of high

---

[24]For example, see the highly cited study by Schlenker and Roberts (2009) on the nonlinear relationship between temperature and yields of corn and soybeans in the U.S.

[25] An interesting question is whether the higher frequency variation is more suitable for estimating causal effects than lower frequency variation. Hsiang (2016) discusses this "frequency-identification trade-off" and argues that lower frequency variation has lower internal validity for measuring causal effects than higher frequency variation.

[26]See Figure 1 in Deryugina and Hsiang (2014) for a visual description of this variation.

temperatures on mortality over time. Using a panel of state-by-year-by-month data, they find that high temperature effect on U.S. mortality rates declined by 75% after the 1960s. A second key finding is the protective effect of residential air conditioning on health. Barreca et al. (2016) document large negative interaction terms between air conditioning and high temperature (80-89°F and especially >90°F), while at the same finding no interaction effect for colder temperatures (<70°F). Taken together, this evidence suggests that the diffusion of residential air conditioning significantly reduced vulnerability to extreme heat, explaining virtually the entire decline in heat-related mortality in the United States over the twentieth century. More broadly, the Barreca et al. (2016) provides an empirical methodology for estimating the effect of adaptation (using existing technologies) to long-term environmental change.

Important progress has been made in understanding the relationship between atmospheric conditions and economic outcomes. However, one key limitation is that studies based on short-run variation are likely to overstate potential long-run economic damages under anthropogenic climate change because the set of possible adaptations may be larger when agents are responding to expected long-run changes. Recent results by Deryugina and Hsiang (2017) describe the assumptions under which effects identified by short-run weather variation may correspond to effects of long-run climate change.

5.4   A selected review of marginal cost estimates

We have thus far focused on papers that estimate the marginal social benefits of public goods. This is because the externality associated with public goods creates distinct identification challenges for quasi-experimental methods that require treatment and control groups. We now discuss several quasi-experimental papers in environmental economics that estimates the other key empirical parameter in the Lindahl-Samuelson condition: the marginal private cost of environmental good provision. Here, the externality present in public goods creates a different empirical challenge. Because an externality leads to either missing or mis-priced markets for the public good, the general-equilibrium marginal cost of environmental policies is rarely directly observed.

What is the marginal cost of an environmental policy? Pizer and Kopp (2005) distinguishes between partial and general equilibrium costs. Partial equilibrium costs are borne by regulated agents while holding prices and activity in other markets fixed. Following the public finance tradition (Harberger, 1971; Diamond and Mirrlees, 1971), general equilibrium costs of a policy are those borne by all agents that compose final good demand after prices across all markets in the economy equilibrate. This distinction is also maintained in quasi-experimental estimates of marginal costs.

As with estimates of marginal social benefits, the U.S. CAA has been the subject of pioneering studies of environmental policy costs using quasi-experimental methods. Greenstone (2002) uses manufacturing plant-level data and a triple-difference research design to estimate the effects of the 1970 and 1977 CAAA on plant level employment, capital stock, and value-added output. Specifically, he compares average outcomes for manufacturing plants (i) before and after the introduction of the amendments, (ii) across nonattainment and attainment counties, and (iii) across plants in industries that are classified as emitting or not emitting at least one of the four regulated pollutants. In practice, this approach amounts to a regression of plant-level outcomes on a set of two-way fixed effects indexed at the county-by-year, non-attainment status-by-year, and industry-by-year levels. Thus, the identifying assumption in this paper is that there are no trending factors that differentially affect input and output decisions in regulated and non-regulated plants. Importantly, the granularity of the data used also allows for the inclusion of plant fixed effects in further remove potentially confounding time-invariant plant characteristics. Based on this approach, Greenstone (2002) finds plants

regulated by the 1970 and 1977 CAAA lost close to 600,000 jobs, \$37 billion in capital stock, and \$75 billion (in 1987 dollars) in value-added revenue.

In general equilibrium, an environmental policy's marginal cost would alter the marginal product of factor inputs, such as labor. Walker (2013) observes that if those made unemployed by an environmental policy are able to find new jobs without significant earnings loss, then the policy would have little effect on final good demand by workers that were laid off by the regulation. Transitional costs, however, may be high if workers lose job- or industry-specific skills and/or experience long unemployment spells. To address this general equilibrium channel, Walker (2013) refines the methodology in Greenstone (2002) in several important ways to examine the effects of the 1990 CAAA. In particular, Walker (2013) uses longitudinal data tracking the employment status and earnings of individual workers in the manufacturing over time and across jobs (rather than tracking the outcomes of individual manufacturing plants over time). Another notable difference is the use of a plant-level data set from the U.S. EPA that identifies which plants are regulated under the Clean Air Act (instead of relying on industry-level proxies of regulation).

Using data on nearly every worker in four U.S. states, and implementing a triple-difference estimator, Walker (2013) finds that the 1990 CAAA resulted in persistent earnings decline for affected workers, with the average worker in a regulated industry experiencing a 20% percent earnings decline. In aggregate terms these losses amount to \$5.4 billion (in 1990 dollars) in forgone earnings in the years following the 1990 CAAA. Walker (2013) also notes that the forgone earnings are two orders of magnitude smaller than the EPA estimated benefits of the 1990 CAAA.

A full general equilibrium estimate of a policy's marginal cost should include effects that propagate across all inputs and outputs of an economy. Such a characterization would require researchers to correctly specify and credibly estimate parameters from a complete model of technology, preferences, and government behavior. In settings where such structural primitives may not be fully known, researchers interested in causal inference may be able to approximate marginal costs under general equilibrium via a sufficient statistics approach. This approach examines "higher-level" variables which theory implies are "sufficient" for welfare analysis regardless of underlying primitive economic parameters (Chetty, 2009). In standard public finance applications, where a private good is taxed, the general equilibrium welfare cost of the tax can be approximated by the product of the tax and the effect of the tax on the taxed good (Harberger, 1964).

Most environmental policies, however, are quantity constraints. In such cases, the marginal cost of a policy is the shadow price on the constraint. For a private good, the First Fundamental Welfare Theorem states that the shadow price on a quantity constraint is the market price for the constrained good. Unfortunately, that condition does not apply for public goods where the externality results in either missing or mis-priced markets.

One class of quantity-restricting environmental policies where the shadow price could be directly observed is cap-and-trade. Under certain assumptions, the equilibrium permit price of the cap-and-trade market is equal to the shadow price of the pollution cap (or constraint).[27] If such assumptions are met, then one needs only to observe the permit price associated with a cap-and-trade market directly to obtain the general equilibrium marginal cost. Unfortunately, there are few cap-and-trade markets in existence. Furthermore, jurisdictions where estimates of the marginal costs of cap-and-trade policies are most needed in policy deliberations are exactly the places that lack such policies.

---

[27]Specifically, the permit price of a cap-and-trade program equals the shadow price of the constraint if the allocation of abatement across polluters were allocated in a least-cost manner (Montgomery, 1972). Violations of this equivalence occur whenever distortions lower the residual demand for permits. Examples include the presence of binding auxiliary environmental policies and when polluters can easily relocate operations in unregulated jurisdictions.

To get around this "chicken-and-egg problem," Meng (2017) develops an empirical method to forecast the general-equilibrium marginal cost of climate change policy based on market expectations. In contrast to the standard approach which requires specifying a computable general equilibrium model, Meng (2017)'s approach allows for causal identification of a reduced-form parameter that serves as a sufficient statistic for the marginal cost. The policy studied is the Waxman-Markey bill, a cap-and-trade climate policy that passed the U.S. House of Representatives in 2009, failed passage in the US Senate in 2010, and is to date the US climate legislation that came closest to becoming implemented.

Meng (2017) exploits a grandfathering permit rule in the proposed policy which grants firms in sectors above a certain historical energy intensity threshold free permits. This threshold-based variation in permit holdings would naturally lead to a RD estimator based on the assumption that firms on either sides of the threshold are similar in terms of unobservable characteristics. However, because the Waxman-Markey policy was never actually implemented, a standard RD comparing firm market values around this threshold is not possible. Instead, Meng (2017) extends the RD framework to account for policy uncertainty, with market-held beliefs captured by prices from a prediction market on the likelihood of eventual Waxman-Markey implementation during the 2009-2010 period. Using this approach, Meng (2017) estimates that markets expected the marginal cost of the Waxman-Markey policy to be in the range of $5 to $19 per ton of carbon dioxide (CO_2e) in 2015 (in 2009 dollars).

Three main findings emerge from this brief review of the quasi-experimental environmental economics literature. First, environmental regulations such as the Clean Air Act have led to notable improvements in air quality over the last decades, with the biggest gains occurring in the early stages of the regulation. Second, improvements in air quality generally lead to important benefits, including increased housing values and better health outcomes, as well as significant impacts on a range of other outcomes. These benefits must be carefully evaluated in comparison to the costs associated with environmental regulations. For costs, quasi-experimental methods are increasingly used towards obtaining estimates that have a general equilibrium interpretation. Third, extreme temperatures are now linked to negative impacts on health, manufacturing productivity, and crop yields. As we explain below, more research is needed in order to interpret these findings in the context of global climate change.

6   Moving forward

Section 4 details how commonly used quasi-experimental estimators in environmental economics may produce biased estimates of the marginal social benefit of environmental goods. This section suggests ways for future quasi-experimental studies in environmental economics to avoid such potential biases. We proceed by separately discussing the problems associated with local and public goods.

6.1   What to do with local public goods

For local public goods, externality spillovers exist but do not affect the entire population. Since there are control units that are not exposed to the public good, one can employ quasi-experimental methods. However, studies should take care to incorporate information on the spatial extent of externality spillovers in the estimation procedure.

In the extended potential outcomes framework of Section 4, externality spillovers are characterized by the likelihood that unit i is exposed to a public good conditional on not being a source, denoted by $S_i = Pr(Q_i = 1 | D_i = 0, \mathcal{Y}_i)$. This spillover term is not unique to environmental externalities; it broadly covers

spillovers from, among other things, regional factor markets and informational treatments. However, while most spillovers are generally hard to observe, environmental economists have the advantage of being able to characterize S_i in many settings. Recent empirical work has begun to incorporate concerns about pollution spillovers indirectly via sample selection of control units (Deschenes, Greenstone and Shapiro, 2017) and directly by modeling pollution transport (Grainger and Ruangmas, 2017; Sullivan, 2017).

We advise future empirical studies to incorporate the following steps. First, regardless of whether the ASEE or the AEEE is implemented, it is useful to know S_i. The availability of atmospheric transport models, combined with pollution exposure data from monitoring stations and satellites at high-spatial resolutions, often allows for the full characterization of the public good source-receptor matrix, $\mathbf{W}$.[28] Indeed, such matrices are already commonly used in prominent integrated assessment models of air pollution for the U.S. at the county level (e.g., Perry et al. (2005); Muller and Mendelsohn (2007, 2009)). With $\mathbf{W}$ in hand, and returning to the notation of source k and receptor i, one can empirically define spillover likelihood for unit $i = 1, \ldots, N$ as

$$S_i = \frac{\sum_{k=1}^{N} w_{ki} - 1}{N - 1} \tag{12}$$

The researcher can then proceed with either an estimator that compares units that are and are not the source of the public good (i.e., the ASEE), or an estimator that compares units that are and are not exposed to the public good (i.e., the AEEE). If the population of interest resides in a location which also produces the environmental good, then the ASEE is available. In that case, we advise that researchers uses the weighted average source effect estimator (WASEE) in equation (11) with unit-level weights of $1 - S_i$. For example, in the U.S. Clean Air Act setting, this estimator would compare average outcomes across counties with nonattainment and attainment status, but would down weight attainment counties that are likely to receive cleaner air spillovers from nonattainment counties. If, however, the population of interest does not reside in locations that also produce the environmental good, then only the AEEE can be implemented. In that case, the researcher should be cautious about any possible correlation between the unit-level likelihood of pollution spillover and potential outcomes. For example, if locations that are more likely to be downwind of pollution sources are also poorer on average, then selection bias arises in the AEEE even if environmental policy were assigned in a quasi-random manner. If this selection bias is due to systematic differences in observed characteristics, then it can be detected by examining whether the likelihood of spillover, S_i, is correlated with observed pre-determined characteristics. For cross-sectional methods such as instrumental variables and regression discontinuity design, that means demonstrating that pre-determined covariates are uncorrelated with S_i. For difference-in-difference settings, that implies showing pre-trends in covariates are uncorrelated with S_i. Of course, if the bias is due to selection on unobservables, there are no direct remedies for bias in the AEEE.

6.2 What to do with global public goods

As the spatial extent of the externality grows, the number of available control units unexposed to the public good decreases. In the limit with a global public good, there are no control units and quasi-experimental estimators are no longer directly suitable. In fact, we showed in Section 4 that for global public goods, the ASEE produces a zero estimate and both the AEEE and WASEE are undefined.

---

[28] In some settings, the source-receptor matrix may not be fixed and needs to be estimated. We leave such considerations for future research.

This difficulty is relevant for anthropogenic climate change (ACC). Of particular importance is the marginal social benefit of greenhouse gas abatement, also known in the literature as the social cost of carbon. Anthropogenic climate change presents a fundamental challenge to quasi-experimental techniques because of its global nature. When a unit of greenhouse gas is emitted somewhere in the world, every location on the planet eventually becomes exposed. As a consequence, there are no unexposed control units. Instead, researchers often rely on two empirical approaches to learn about the possible impacts of greenhouse gas emissions on economic outcomes. The first approach, which uses the AEEE and reviewed in Section 5.3, exploits historical local temperature variation arising from natural climate variability. Implemented typically via a panel fixed effects model, this approach asks what happens when temperature for a given location changes while temperatures elsewhere on the planet are fixed. Anthropogenic climate change, however, is expected to alter atmospheric conditions across the globe. As a consequence, the estimates provided by this approach is unlikely to capture general equilibrium effects that would occur when climate change jointly warms *all* locations. A second approach is to run a time-series analysis of global spatially-averaged temperature on some global economic outcome. This approach may not identify causal effects if other drivers of the global economy are changing over time and are correlated with temperature changes. In short, because of the global nature of ACC, cross-sectional comparisons are not available while time-series comparisons may not allow a causal interpretation.

In light of this fundamental challenge, how does one inform climate policy with causal estimates? We propose a two-pronged approach which combines causal effects of atmospheric variables on economic primitives with causal validation of general equilibrium models. The first step is to "go small", by estimating the direct effects of local weather on local economic primitives that are invariant to prices. This allows one to isolate effects that are unaffected by general equilibrium mechanisms. The second step is to "go big", by causally validating general equilibrium models of economic exchange. When combined, these two approaches should allow researchers to credibly assemble forecasts of the social cost of carbon that include both direct partial equilibrium effects and indirect general equilibrium effects.[29]

Go small... Outcomes are considered "economic primitives" if they do not respond to prices. For such outcomes, local weather has the same effect regardless of whether it is driven by anthropogenic climate change or by local environmental conditions. For example, extreme temperature is likely to lower cereal yields and increase human mortality whether it arises from greenhouse gas emissions or a naturally-occurring local heat wave. These direct local effects can be causally identified using the AEEE. To estimate such effects, researchers should take care to select appropriate outcome variables and to isolate local weather variation.

First, researcher should select, as best as possible, outcome variables that are unaffected by prices. Examples of such economic primitives include physical measures of productivities and amenities such as cereal yields (Schlenker and Roberts, 2009; Burke and Emerick, 2016), manufacturing productivity (Zhang et al., 2018), and mortality (Deschenes and Greenstone, 2011; Barreca et al., 2016). In practice, however, it is hard to guarantee that even these measures are unaffected by prices. For example, because cereal yields may depend on fertilizer as input, shocks to fertilizer prices would affect yields. Similarly, mortality rates may be a function of the cost of medical care. To address this potential issue, researchers should additionally exploit weather variation that is spatially uncorrelated. Such variation can be isolated by employing, as much as possible, fixed effects at spatial levels broader than the unit of observation. By ensuring that the

---

[29] Our approach has flavors of the seminal Lucas critique in macroeconomics (Lucas, 1976). Here, we argue that historical local temperature effects on economic outcomes alone may not be pertinent for forecasting anthropogenic climate change impacts. Such local relationships may not hold under climate change because of the general equilibrium consequences of its global nature.

identifying weather variation is local to the unit of observation and uncorrelated with weather elsewhere, the estimated effect is less likely to capture price responses.

...and go big   Anthropogenic climate change is a global event. A full accounting of its global economic impacts involve not just knowing what happens when a single location warms but what happens when all locations warm at the same time. As a consequence, patterns of economic exchange will shift and price effects are likely to be important components of the overall economic impact of climate change. To provide credible economic predictions, one needs to combine partial equilibrium estimates of local temperature effects on economic primitives with models of economic exchange, both across sectors in a given location and across locations. Such models should also be empirically validated.

Empirical validation of economic models is a particularly hard task because two objectives must be satisfied. First, one must not only causally validate a model's prediction, but the prediction must also be unique to that model. If unique predictions do not exist for a model, one cannot claim that the evidence is consistent with one model and not another. Second, predictions typically describe market-level features of the model. If the market is global, we return to the problem of not having control units necessary for quasi-experimental estimation.

A weaker version of the first objective is to validate a predicted effect from a model of economic exchange against a null hypothesis of no effect. This is the approach taken by Dingel, Meng and Hsiang (2018) who show using a standard model of international trade that a rise in the spatial correlation of productivities amplifies cross-country welfare inequality by increasing the correlation between a country's productivity and its gains from trade. This prediction is then causally validated against a null of no change in the global variance of the gains from trade. Specifically, the authors exploit a naturally-occurring, quasi-random global climatic phenomenon - the El Niño Southern Oscillation (ENSO) - which drives the annual global spatial correlation of cereal productivity. Using a half-century of global agricultural trade data, Dingel, Meng and Hsiang (2018) find that year-to-year ENSO variation alters the distribution of the gains from trade so as to increase welfare inequality in a way that is consistent with their model's theoretical prediction. Having validated their prediction, Dingel, Meng and Hsiang (2018) consider the consequences of accounting for this general equilibrium mechanism in projections of climate change impacts. They show that projections that include a rise in the spatial correlation of cereal productivity due to climate change predict a 20% greater increase in global welfare inequality by the end of the 21st century compared to projections that omit this general equilibrium effect.

7   Conclusion

Quasi-experimental empirical methods have fundamentally transformed environmental economics. This paper reviews the recent application of such techniques from the perspective of a central problem in environmental economics: the optimal provision of public goods. We show formally that the presence of externalities in public goods creates an extra complication when applying standard quasi-experimental methods. Specifically, externality spillovers imply that estimates of the marginal social benefits of a public good using two common quasi-experimental estimators may be biased even under the standard assumption that the public good source is randomly assigned.

We review the existing literature focusing on this particular issue and offer suggestions for future research. For valuation of local public goods, we advise the explicit modeling of externality spillovers and introduce

an unbiased estimator that incorporates this information. For global public goods where, by definition, no control units exist, we suggest future research directions that preserve the credibility of causal inference.

We leave two important topics for future research. One issue is how to conduct statistical inference and construct standard errors for the proposed estimators in the presence of treatment spillovers. Another area is the development of data sources and econometric methods for estimating the unit-level likelihood of spillover. All the derivations in this paper, including the unbiased weighted average source effect estimator, assume that spillover likelihoods are known. In some cases, spillover likelihoods may not be known and require estimation. Further work is needed to develop tractable estimation approaches in such settings.

Finally, our review highlights two other shortcomings in the existing quasi-experimental literature in environmental economics. First, air pollutants are typically analyzed in isolation even when most tend to be co-produced and thus correlated with one another. Future research should devise identification strategies that can quasi-randomly alter one pollutant while not affecting others. Second, we observe that there are few quasi-experimental papers that focus on other environmental goods besides air pollution and temperature. Increasing data availability on other environmental attributes from remote sensing and other sources may help to fill this literature gap in the years to come.

References

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review*, 104(10): 3003–37.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Angrist, Joshua D., and Jrn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, 24(2): 3–30.

Ashenfelter, Orley. 1978. "Estimating the effect of training programs on earnings." *The Review of Economics and Statistics*, 47–57.

Ashenfelter, Orley, and David Card. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." *The Review of Economics and Statistics*, 67(4): 648–660.

Auffhammer, Maximilian, and Ryan Kellogg. 2011. "Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality." *The American Economic Review*, 101(6): pp. 2687–2722.

Auffhammer, Maximilian, Antonio M Bento, and Scott E Lowe. 2009. "Measuring the effects of the Clean Air Act Amendments on ambient PM10 concentrations: The critical importance of a spatially disaggregated analysis." *Journal of Environmental Economics and Management*, 58(1): 15–26.

Barreca, Alan I. 2012. "Climate Change, Humidity, and Mortality in the United States." *Journal of Environmental Economics and Management*, 63(1): 19–34.

Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S. Shapiro. 2016. "Adapting to Climate Change: The Remarkable Decline in the U.S. Temperature-Mortality Relationship over the Twentieth Century." *Journal of Political Economy*, 124(1): 105–159.

Barreca, Alan, Karen Clay, Olivier Deschnes, Michael Greenstone, and Joseph S. Shapiro. 2015. "Convergence in Adaptation to Climate Change: Evidence from High Temperatures and Mortality, 1900-2004." *American Economic Review*, 105(5): 247–51.

Baumol, William J., and Wallace E. Oates. 1988. *The Theory of Environmental Policy, Second Edition.* Cambridge University Press.

Becker, Randy, and Vernon Henderson. 2000. "Effects of Air Quality Regulations on Polluting Industries." *Journal of Political Economy*, 108(2): pp. 379–421.

Bento, Antonio, Matthew Freedman, and Corey Lang. 2015. "Who Benefits from Environmental Regulation? Evidence from the Clean Air Act Amendments." *The Review of Economics and Statistics*, 97(3): 610–622.

Bockstael, Nancy E, and Kenneth E McConnell. 2007. *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models.* Vol. 7, Springer Science & Business Media.

Bui, Linda TM, and Christopher J Mayer. 2003. "Regulation and Capitalization of Environmental Amenities: Evidence from the Toxic Release Inventory in Massachusetts." *Review of Economics and statistics*, 85(3): 693–708.

Burke, Marshall, and Kyle Emerick. 2016. "Adaptation to Climate Change: Evidence from US Agriculture." *American Economic Journal: Economic Policy*, 8(3): 106–40.

Campbell, Donald T, and Julian C Stanley. 1963. "Experimental and Quasi-experimental Designs for Research." In *Handbook of Research on Teaching.* Rand McNally.

Chang, Tom, Joshua Graff Zivin, Tal Gross, and Matthew Neidell. 2016. "Particulate Pollution and the Productivity of Pear Packers." *American Economic Journal: Economic Policy*, 8(3): 141–69.

Chay, Kenneth, Carlos Dobkin, and Michael Greenstone. 2003. "The Clean Air Act of 1970 and Adult Mortality." *Journal of Risk and Uncertainty*, 27(3): 279–300.

Chay, Kenneth Y., and Michael Greenstone. 2003. "Air Quality, Infant Mortality, and the Clean Air Act of 1970." National Bureau of Economic Research Working Paper 10053.

Chay, Kenneth Y., and Michael Greenstone. 2005. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy*, 113(2): 376–424.

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods." *Annual Review of Economics*, 1(1): 451–488.

Cook, Thomas D, Donald Thomas Campbell, and William Shadish. 2002. *Experimental and Quasi-experimental Designs for Generalized Causal Inference.* Houghton Mifflin Boston.

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker. 2015. "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings." *American Economic Review*, 105(2): 678–709.

Davis, Lucas W. 2008. "The Effect of Driving Restrictions on Air Quality in Mexico City." *Journal of Political Economy*, 116(1): pp. 38–81.

Davis, Lucas W. 2011. "The Effect of Power Plants on Local Housing Values and Rents." *Review of Economics and Statistics*, 93(4): 1391–1402.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken. 2014. "What Do We Learn from the Weather? The New Climate-Economy Literature." *Journal of Economic Literature*, 52(3): 740–98.

Deryugina, Tatyana, and Solomon Hsiang. 2017. "The Marginal Product of Climate." National Bureau of Economic Research Working Paper 24072.

Deryugina, Tatyana, and Solomon M. Hsiang. 2014. "Does the Environment Still Matter? Daily Temperature and Income in the United States." National Bureau of Economic Research Working Paper 20750.

Deschenes, Olivier, and Michael Greenstone. 2007. "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather." *The American Economic Review*, 97(1): pp. 354–385.

Deschenes, Olivier, and Michael Greenstone. 2011. "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US." *American Economic Journal: Applied Economics*, 3(4): 152–85.

Deschenes, Olivier, and Michael Greenstone. 2012. "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Reply." *American Economic Review*, 102(7): 3761–73.

Deschenes, Olivier, Michael Greenstone, and Joseph S. Shapiro. 2017. "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program." *American Economic Review*, 107(10): 2958–89.

Diamond, Peter A, and James A Mirrlees. 1971. "Optimal Taxation and Public Production II: Tax Rules." *The American Economic Review*, 61(3): 261–278.

DiNardo, J. 2008. "Natural Experiments and Quasi-natural Experiments. The New Palgrave Dictionary of Economics."

Dingel, Jonathan I., Kyle C. Meng, and Solomon M. Hsiang. 2018. "The Spatial Structure of Productivity, Trade, and Inequality: Evidence from the Global Climate."

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *The Quarterly Journal of Economics*, 128(4): 1499–1545.

Fisher, Anthony C., W. Michael Hanemann, Michael J. Roberts, and Wolfram Schlenker. 2012. "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment." *American Economic Review*, 102(7): 3749–60.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2018. "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program*." *The Quarterly Journal of Economics*, 133(3): 1597–1644.

Freeman, A Myrick, Joseph A Herriges, and Catherine L. Kling. 2014. *The Measurement of Environmental and Resource Values: Theory and Methods.* Routledge.

Gamper-Rabindran, Shanti, and Christopher Timmins. 2013. "Does Cleanup of Hazardous Waste Sites Raise Housing Values? Evidence of Spatially Localized Benefits." *Journal of Environmental Economics and Management*, 65(3): 345 – 360.

Graff Zivin, Joshua, and Matthew Neidell. 2012. "The Impact of Pollution on Worker Productivity." *American Economic Review*, 102(7): 3652–73.

Grainger, Corbett A. 2012. "The Distributional Effects of Pollution Regulations: Do Renters Fully Pay for Cleaner Air?" *Journal of Public Economics*, 96(9): 840 – 852.

Grainger, Corbett, and Thanicha Ruangmas. 2017. "Who Wins from Emissions Trading? Evidence from California." *Environmental and Resource Economics*, forthcoming.

Greenstone, Michael. 2002. "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures." *Journal of Political Economy*, 110(6): pp. 1175–1219.

Greenstone, Michael, and Justin Gallagher. 2008. "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program." *The Quarterly Journal of Economics*, 123(3): 951–1003.

Greenstone, Michael, and Rema Hanna. 2014. "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India." *The American Economic Review*, 104(10): 3038–3072.

Greenstone, Michael, and Ted Gayer. 2009. "Quasi-experimental and Experimental Approaches to Environmental Economics." *Journal of Environmental Economics and Management*, 57(1): 21 – 44. Frontiers of Environmental and Resource Economics.

Harberger, Arnold C. 1964. "The Measurement of Waste." *The American Economic Review*, 54(3): 58–76.

Harberger, Arnold C. 1971. "Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay." *Journal of Economic literature*, 9(3): 785–797.

He, Guojun, Maoyong Fan, and Maigeng Zhou. 2016. "The Effect of Air Pollution on Mortality in China: Evidence from the 2008 Beijing Olympic Games." *Journal of Environmental Economics and Management*, 79: 18 – 39.

Heblich, Stephan, Alex Trew, and Yanos Zylberberg. 2016. "East Side Story: Historical Pollution and Persistent Neighborhood Sorting." Spatial Economics Research Centre, LSE SERC Discussion Papers 0208.

Heckman, James J. 2010. "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy." *Journal of Economic literature*, 48(2): 356–98.

Henderson, J. Vernon. 1996. "Effects of Air Quality Regulation." *The American Economic Review*, 86(4): pp. 789–813.

Heutel, Garth, Nolan H. Miller, and David Molitor. 2017. "Adaptation and the Mortality Effects of Temperature Across U.S. Climate Regions." National Bureau of Economic Research Working Paper 23271.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association*, 81(396): 945–960.

Hsiang, Solomon. 2016. "Climate Econometrics." *Annual Review of Resource Economics*, 8(1): 43–75.

Husar, R.B., and W.P. Renard. 1997. "Ozone as a Function of Local Wind Speed and Direction: Evidence of Local and Regional Transport." Air and Waste Management Associations 90th Annual Meeting and Exhibition, Toronto, Ontario.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences.* Cambridge University Press.

Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, 47(1): 5–86.

Isen, Adam, Maya Rossin-Slater, and W. Reed Walker. 2017. "Every Breath You TakeEvery Dollar Youll Make: The Long-Term Consequences of the Clean Air Act of 1970." *Journal of Political Economy*, 125(3): 848–902.

Keane, Michael P. 2010. "A Structural Perspective on the Experimentalist School." *Journal of Economic Perspectives*, 24(2): 47–58.

Klaiber, H Allen, and V Kerry Smith. 2013. "Quasi Experiments, Hedonic Models, and Estimating Trade-offs for Local Amenities." *Land Economics*, 89(3): 413–431.

Knittel, Christopher R, Douglas L Miller, and Nicholas J Sanders. 2016. "Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health." *Review of Economics and Statistics*, 98(2): 350–366.

Kuminoff, Nicolai V, and Jaren C Pope. 2014. "Do "Capitalization Effects" for Public Goods Reveal the Public's Willingness to Pay?" *International Economic Review*, 55(4): 1227–1250.

Kuminoff, Nicolai V., V. Kerry Smith, and Christopher Timmins. 2013. "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets." *Journal of Economic Literature*, 51(4): 1007–1062.

Lindahl, Erik. 1919. "Just Taxation - a Positive Solution." In *Reprinted: Classics in the theory of public finance.* 168–176. Springer.

Lucas, Robert E. 1976. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy*, 1: 19 – 46.

Manski, Charles F. 2013. "Identification of Treatment Response with Social Interactions." *The Econometrics Journal*, 16(1): S1–S23.

Mendelsohn, Robert, William D. Nordhaus, and Daigee Shaw. 1994. "The Impact of Global Warming on Agriculture: A Ricardian Analysis." *The American Economic Review*, 84(4): pp. 753–771.

Meng, Kyle C. 2017. "Using a Free Permit Rule to Forecast the Marginal Abatement Cost of Proposed Climate Policy." *American Economic Review*, 107(3): 748–84.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*, 72(1): 159–217.

Montgomery, W. David. 1972. "Markets in Licenses and Efficient Pollution Control Programs." *Journal of Economic Theory*, 5(3): 395 – 418.

Moretti, Enrico, and Matthew Neidell. 2011. "Pollution, Health, and Avoidance Behavior: Evidence from the Ports of Los Angeles." *Journal of Human Resources*, 46(1): 154–175.

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.

Muller, Nicholas Z., and Robert Mendelsohn. 2009. "Efficient Pollution Regulation: Getting the Prices Right." *The American Economic Review*, 99(5): 1714–1739.

Neidell, Matthew. 2017. "Air Pollution and Worker Productivity." *IZA World of Labor*, , (363).

Nevo, Aviv, and Michael D. Whinston. 2010. "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference." *Journal of Economic Perspectives*, 24(2): 69–82.

Perry, Steven G, Alan J Cimorelli, Robert J Paine, Roger W Brode, Jeffrey C Weil, Akula Venkatram, Robert B Wilson, Russell F Lee, and Warren D Peters. 2005. "AERMOD: A Dispersion Model for Industrial Source Applications. Part II: Model Performance Against 17 Field Study Databases." *Journal of Applied Meteorology*, 44(5): 694–708.

Pizer, William A, and Raymond Kopp. 2005. "Calculating the Costs of Environmental Regulation." *Handbook of Environmental Economics*, 3: 1307–1351.

Rosenzweig, Mark R, and Kenneth I Wolpin. 2000. "Natural "Natural Experiments" in Economics." *Journal of Economic Literature*, 38(4): 827–874.

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *The Review of Economics and Statistics*, 387–389.

Sanders, Nicholas J., and Charles Stoecker. 2015. "Where have all the Young Men Gone? Using Sex Ratios to Measure Fetal Death Rates." *Journal of Health Economics*, 41: 30 – 45.

Schlenker, Wolfram, and Michael J Roberts. 2009. "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change." *Proceedings from the National Academy of Science*, 106(37): 15594–8.

Schlenker, Wolfram, and W. Reed Walker. 2016. "Airports, Air Pollution, and Contemporaneous Health." *The Review of Economic Studies*, 83(2): 768–809.

Stock, James, and Mark W Watson. 2017. *Introduction to Econometrics.* Prentice Hall.

Sullivan, Daniel M. 2017. "The True Cost of Air Pollution: Evidence from the Housing Market."

Timmins, Christopher, and Wolfram Schlenker. 2009. "Reduced-Form Versus Structural Modeling in Environmental and Resource Economics." *Annual Review of Resource Economics*, 1(1): 351–380.

Walker, W. Reed. 2013. "The Transitional Costs of Sectoral Reallocation: Evidence From the Clean Air Act and the Workforce." *The Quarterly Journal of Economics*, 128(4): 1787–1835.

Zhang, Peng, Olivier Deschenes, Kyle Meng, and Junjie Zhang. 2018. "Temperature Effects on Productivity and Factor Reallocation: Evidence from a Half Million Chinese Manufacturing Plants." *Journal of Environmental Economics and Management*, 88: 1–17.