What is Additionality?

Part 2: A framework for more precise definitions and standardized approaches

Michael Gillenwater

Discussion Paper No. 002
January 2012 | Version 03

ghg management institute
www.ghginstitute.org

What is Additionality?
Part 2: A framework for more precise definitions and standardized approaches

MICHAEL GILLENWATER[†,‡]

[†]*Greenhouse Gas Management Institute, Silver Spring, MD*

[‡]*Science, Technology and Environmental Policy Program, Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ*

(email: gillenwater@alum.mit.edu)

Discussion Paper (version 03)
January 2012

ABSTRACT

This article proposes a conceptual framework for offset policy makers and program administrators to shift the additionality and baseline assessment process from subjective "tests" towards more standardized approaches based on explicitly recognized policy interventions, theories of behavior, and objective models. This framework is intended to be generalizable to any type of offset program. Explicitly elaborating and using a model with variables potentially brings greater transparency, objectivity, and replicability to the assessment process. Such an approach also permits the development of a falsifiable hypothesis that can then be challenged and improved. The challenge for offset policy makers is evaluating whether offset program administrators can create models for additionality and baseline assessments for a given class of activities that are good enough to provide net benefits greater than the alternative policies available to achieve a given objective. A key message of this article is that offset mechanisms, to be credible, need to better define the scientific basis of additionality and baselines. And even though perfect certainty is unattainable, offset quality can be improved with clearer assumptions, better models, and greater attention to the rules of scientific inference, which are critically important when considering the use of standardized approaches.

KEYWORDS
additionality, offsets, standardized approaches, baseline scenario, environmental markets

1 Introduction

Within climate change policy, and environmental markets more broadly, the concept of additionality has been an especially charged topic, eliciting frustration, disagreement, and even despair. At the heart of these reactions is a fundamental obstacle preventing constructive debate. Lacking are both well-framed definitions of additionality and baseline and an accepted conceptual framework within which discussion and policy formulation can take place.

This article is the second in a three-part series, the overall aim of which is to more precisely define the terms "additionality" and "baseline" to better enable their application within offset policies. The intended audience includes policy makers, environmental market practitioners, and social scientists. Significant attention is given to greenhouse gas (GHG) emission offsets as a case study; however, the framework presented is intended to be generalizable to any type of offset credit policy or program.

Part 1 focused on definitional problems within existing offset programs, standards, and climate change policy literature. It was shown that within the context of GHG emission offset programs, the language used to discuss and define additionality and baseline is, with few exceptions, imprecise, varied, and internally inconsistent. It concluded by proposing the following definitions.

Additionality is the property of an activity being *additional*. A proposed activity is *additional* if the recognized policy interventions are deemed to be causing the activity to take place. The occurrence of additionality is determined by assessing whether a proposed activity is distinct from its baseline (see below).

A **baseline** is a prediction of the quantified amount of an input to or output from an activity resulting from the expected future behavior of the actors proposing, and affected by, the proposed activity in the absence of one or more policy interventions, holding all other factors constant (*ceteris paribus*). The conditions of a baseline are described in a baseline scenario.

This second article in the series builds from these definitions by outlining and then systematically analyzing a series of fundamental questions with respect to additionality and baselines that have been neglected, but are critical to offset policy design. Specifically, I explain why the definitions of additionality and baseline are contingent on the specification of a policy intervention and provide a framework for the development of standardized approaches for the assessment of additionality and setting of baselines that is founded on scientifically credible investigations. These investigations are to serve as the basis for objective decision rules and algorithms for the assessment of additionality and selection of baselines. I refer to these rules and algorithms, which make up a standardized approach, as a model.

I build on the valuable work of Chomitz (Chomitz 1998) and Trexler, Broekhoff et al. (2006) that advanced our understanding of additionality and baselines by applying concepts from program evaluation[1] and statistical hypothesis testing. However, as was shown in Part 1, most of the climate change policy literature on additionality and baselines has focused on techniques (i.e., tests) for assessing additionality, especially under the Kyoto Protocol's Clean Development Mechanism (CDM), and gives little attention to the theoretical justifications for these techniques.

[1] The concept of additionality is similar to the concept of the *ex ante* assessment of policy intervention effects within the field of program and policy evaluation. Evaluating the effect of a program or policy can also be done *ex post*, which is discussed below in the context of the testing and falsification of additionality and baseline models.

To address this gap, I draw on causal inference theory (King, Keohane et al. 1994; Epstein and King 2002; Brady and Collier 2004; George and Bennett 2004) and approaches from program and policy evaluation (Rossi, Lipsey et al. 2004; Ferraro and Pattanayak 2006; EP 2008; BIS 2009; Khandker, Koolwal et al. 2010). *Inference* means using facts we know to learn about or draw conclusions regarding behaviors that we do not know (Epstein and King 2002); and *causal inference* is the process of learning about causation from observed data (King, Keohane et al. 1994; Brady and Collier 2004). For complex social and economic behaviors, it is challenging to empirically observe cases both with and without a policy intervention under perfectly controlled conditions (i.e., ideal experimentation). However, it is possible to use causal inference—the basis for much of social science—to develop acceptable levels of confidence in an estimation of threshold values for specific factors that result in behavioral changes.

For the assessment of additionality and baselines, causal inference investigations can lead to predictive models of behavior where a recognized policy intervention is assumed to be absent (i.e., the baseline scenario). By explicitly specifying variables and a model, the assessment process becomes subject to scientific testing and improvement. The goal is to develop models in the form of decision rules (i.e., a decision tree) using easily measured variables, although in some cases a model may take the form of a more complex algorithm. Regardless, the investigation of causal inference serves as the basis and evidence supporting the development of model algorithms and decision threshold values. With a model in hand, a proposed activity (e.g., project) can then be compared to its predicted baseline (or a set of equally likely baselines).

2 Outlining the framework

Additionality is the process of assessing whether a proposed activity is different than its baseline. For example, in the context of climate change policy the question of additionality is whether GHG emissions from a proposed activity will be different than baseline emissions.[2] To further elaborate and apply the above definitions of additionality and baseline, several underlying questions must be asked and answered:

- What is the recognized policy intervention (i.e., treatment or causal factor) created by the offset policy, which is intended to change behavior?
- What specific treatment variables represent that policy intervention?
- What theory of behavior is assumed for modeling the effect of the policy intervention on actors?
- In an additionality and baseline scenario model, what type of dependent variable (i.e., outcome variable) will be used and what are the roles of other variables in the model?

These questions outline a theoretical framework for additionality that can provide a generalized foundation for realistic offset policy implementation. Collectively, these questions have neither been systematically asked nor thoroughly analyzed in previous work within the literature on offsets policy.[3] This section explores these questions and discusses options for how they can be answered. The following section then presents how this framework can be applied for use in offset programs.

The framework presented below is intended to enable the development of more objective standardized approaches for the assessment of additionality and baselines, which are both more credible and closer to meeting scientific principles of reliability and falsifiability. Although largely conceptual in character, the answers to these questions are the foundation of a transparent additionality and baseline assessment

[2] An activity's emissions could be less than or greater than the baseline, but if they are greater than the baseline then no offset credits will be awarded. So, the question of additionality can be more simply formulated as a difference question rather than a less than question.

[3] Individual questions have been partially discussed by Chomitz (1998), Greiner and Michaelowa (2003), Trexler, Broekhoff et al. (2006), and Bennett (2010).

process. Offset policy makers and program administrators that ignore these questions and move directly to "just coming up with tests for additionality" will create assessment processes without transparent assumptions and will thereby open the design and implementation of their programs to the vagaries of subjective and political biases.

2.1 What is the policy intervention?

Additionality is typically referred to as the determination of whether or not an activity or project is different than what would have happened otherwise. Yet, this question cannot be answered unless the question "otherwise except for what" is answered first. As demonstrated in Part 1, saying "otherwise except for the project" begets circular definitions of additionality and baseline. Instead the "what" in this question is a policy "intervention" (or "treatment").

Specifically, additionality is about whether and in which cases a policy intervention is causing behavior change. The assessment of additionality, then, must take into account the possibility that behavior may remain unchanged even when the policy intervention is present. By policy intervention, we mean any type of incentive (i.e., "carrots"), assistance, mandate, or threat (i.e., "sticks"). Policy interventions can vary in magnitude, timing, and form. The issuance of credits is not necessarily the only form of policy intervention created by an offset program. Programs can take actions that directly or indirectly have the potential to lead to behavior change.

A few authors have made reference to, intentionally or not, the type of policy intervention they recognize as being relevant for the assessment of additionality and baselines for emission offset projects. References to a policy intervention, where present in the literature, fall into one of the four types presented in Table 1.[4] The type of policy intervention that an offset program recognizes has major implications for how additionality and baseline are assessed. Specifically, each type leads to a different set of factors that are acknowledged as influencing the behavior of target actors.

Table 1. Policy intervention descriptions for additionality and baseline scenario assessments in the GHG emission offset literature

<table>
<tr><th>Type</th><th>Description of policy intervention type</th><th>Typical language in literature</th><th>Example References</th></tr>
<tr><td>A</td><td>Expected economic value of offset credits</td><td>Income from offset credits or carbon funding</td><td>(Chomitz 1998; Sutter and Parreño 2007; Au Yong 2009; Bennett 2010)</td></tr>
<tr><td>B</td><td>A + Measures directly taken by offset program or policy</td><td>Offset program</td><td rowspan="2">(Asuka and Takeuchi 2004; Fischer 2005; Kartha, Lazarus et al. 2005; Au Yong 2009; CBO 2009; Müller 2009; Bennett 2010)</td></tr>
<tr><td>C</td><td>A + B + Indirect market spillover effects</td><td>Offset credit market</td></tr>
<tr><td>D</td><td>Emergence or presence of climate change as an issue for policy or decision making consideration</td><td>Issue of climate change</td><td>(Sugiyama and Michaelowa 2001; Asuka and Takeuchi 2004; WRI/WBCSD 2005; Trexler, Broekhoff et al. 2006)</td></tr>
</table>

Type *A* from Table 1 recognizes a policy intervention that is narrowly interpreted. With this type, factors other than the expected economic value of offset credits are considered to have an insignificant influence

[4] Schneider (2009a) identifies the act of an offset program registering a project as the recognized policy intervention., Registration is, however, a proximate, not ultimate policy intervention. The policy intervention would be what registration of a project makes possible (e.g., the opportunity to earn revenue from credits).

on behavioral change; offset policy is effectively viewed as a specific type of economic subsidy. The resulting treatment variable (i.e., causal variable) under Type *A* would be related to the expected value of the subsidy. For example, the treatment variable could be defined as the Net Present Value (NPV) (i.e., discounting for time as well as uncertainty/risk) of the offset credits expected to be earned by the proposed activity. A policy intervention of this type would probably be used within a model that focuses on: i) the effects of an offset credit price signal on the behavior of investors or other relevant actors and ii) factors, in the form of variable thresholds, that identify the actors and contexts where changes in behavior due to this price signal are likely to occur (e.g., producers on the margin).[5]

A narrow definition of policy intervention represented by Type *A* does not necessarily lead to a model for assessing additionality and baselines that solely uses financial analysis and financial variables to predict behavior. Obviously, not every decision is solely a based on a financial calculation; non-financial factors are important for some decisions. The purpose of Table 1, however, is not it identify all of the potential factors that affect decision makers. It is instead to identify the specific policy intervention recognized by an offset program. It is critically important to the understanding of additionality and baselines that these two concepts be distinguished. Even if behavior can be influenced by a number of different factors or is prevented from occurring by one or more factors (i.e., barriers), this does not mean that any of those factors are affected by a policy intervention. Only factors that are influenced by an offset program's policy intervention are relevant. For example, one must show (i.e., model) how the recognized policy intervention overcomes barriers to the implementation of a proposed activity.

A policy intervention of Type *B* would include a broader set of treatment variables than a Type *A*. Type *B* recognizes that an offset program may include a mixture of policy interventions intended to influence behavior and that policy interventions are more than simply a price signal. They potentially include other measures directly taken by the offset program. For example, a program could provide new market infrastructure that lowers the transaction costs of implementing activities (e.g., through education or capacity building) or offer other technical, legal, and/or financial support. Table 2 lists some generic categories of policy interventions, any combination of which could be directly made by an offset program.

Table 2. Categories of policy interventions

1. Economic subsidies for an activity[6]
2. Measures to reduce transaction costs, which can decrease the cost to an actor of a known activity and/or add to the list of known options for a given decision or behavioral choice a. Information or educational campaigns b. Changes to existing rules (e.g., remove regulations that discourage or increase the costs of certain activities) c. Providing new infrastructure (e.g., communication networks) d. Research and development funding e. Development of voluntary technical standards
3. Enforced government mandates
4. Reputation-based programs a. Recognition programs for superior performing actors b. Shaming programs for poorly performing actors

[5] The magnitude of this price signal is typically thought of as a function of market forces (assuming some source of demand for credits), but it could also be a price bound (e.g., through price floors and/or ceilings) or directly set by a program administrating agency.
[6] An offset program could provide other subsidies beyond those represented by offset credits.

A policy intervention of Type *C* includes yet a broader set of treatment variables than Type *B*. Type *C* recognizes the actual policy intervention to be the creation of a broader offset credit market, and therefore assumes the policy intervention includes broader market effects[7] as well as the offset credit price signal (Type *A*) and other direct measures taken by the offset program (Type *B*).[8] These market effects could include a host of network, positive, and negative externalities. For example, the offset credit price signal could lead to the invention of a new emission reduction technology that would not have developed had the offset market not existed.[9] If such a new technology turns out to be the lowest cost option for a given activity then it may not be assessed as additional under Type *A* or *B*, where its use would be seen as the likely baseline scenario. But if the new technology was created in response to the offset market then it would not been available in the baseline scenario, assuming Type *C*.

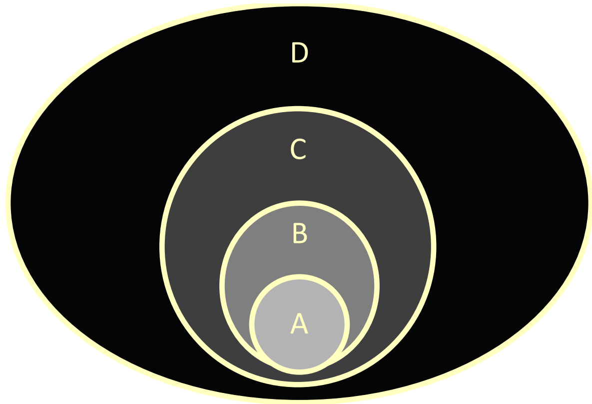
If one recognizes a policy intervention of Type *C*, then a baseline scenario would be modeled as an alternative world that does not (or never did) include an offset credit market and offset program. Clearly, such an assumption increases the number of relevant treatment variables and the complexity of any model used to predict baselines.

Type *D* dramatically broadens the view of a policy intervention. It recognizes the policy intervention to be the emergence or presence of humanity's knowledge of anthropogenic climate change. Under Type *D*, any change in human behavior resulting from this knowledge would be additional. A baseline scenario would then be assessed using a model of the world in which: i) the problem of climate change did not exist, ii) society was unaware of the existence of anthropogenic climate change, or iii) society showed no interest in the issue and so this knowledge had no effect on human behavior. Clearly, modeling an alternative past and future under the assumption of this type implies an unrealistically complex model as well as a host of subjective assumptions. Yet, authors who define additionality in terms of what would have happened in the absence of climate change as a policy issue are implying just this sort of approach (see Table 1).

Figure 1 illustrates the relationship between the four types of policy intervention presented in Table 1. Each succeeding type increases the number of treatment variables that would be necessary to model a baseline scenario and assess additionality. Once treatment variables representing a policy intervention are specified and quantified, then one can attempt to model the effect of a change in these treatment variables on behavior. Offset program administrators must decide whether each added treatment variable excessively grows the complexity of the assessment process.

[7] Also referred to as secondary, indirect, spillover, or market transformation effects (Vine and Sathaye 2000; Bernow, Kartha et al. 2001; Nadel, Thorne et al. 2003).
[8] Project leakage (or affected GHGs under ISO 14064-2) is also fundamentally an indirect effect. See Millard-Ball and Ortolano (2010) for a useful discussion of indirect effects and leakage issues related to transport projects.
[9] The presence of an offset market, by offering an alternative compliance option to regulated entities under an emissions cap-and-trade program, could also be argued to reduce the incentive for those capped entities (versus those supplying offset credits) to innovate relative to a scenario in which this alternative compliance option did not exist.

Figure 1. Relationship between policy intervention types presented in Table 1 in terms of number of treatment variables likely needed to represent it



2.2 What is the theory of behavioral causation?

Separate from defining the type of policy intervention and specifying treatment variable(s) that represent it, it is also necessary to assume an overarching theory of behavior. This theory will guide the selection of other factors (i.e., variables) and optimization function (i.e., causal mechanism) to include in the models used by offset programs to assess additionality and baselines. "Without a theoretical model, [we] cannot decide which potential explanatory variables should be included in [the] analysis" (King, Keohane et al. 1994). A baseline is predicated on the assumption that we can reasonably model behavior under conditions where a policy intervention is absent and that the preferences of actors—and how they act on those preferences—are adequately understood and captured by the model. The climate policy literature on additionality and baselines has typically been silent on its assumed theory of behavior.

The first consideration in elaborating a theory of behavior is identifying the actors whose behavior is being modeled. Economists use the term "representative agent" to refer to a typical decision-maker of a certain type (e.g., typical consumer or typical firm) who is choosing behaviors based on some decision making model (Hartley 1996).[10] In the context of emission offset programs, the relevant actors for a particular type of activity would be individuals and/or organizations (e.g., firms, coalitions of investors, or government agencies) with decision making authority over investments or operations that affect emissions related to a class of activities.

There are several theories of behavior that could be assumed for the development of additionality and baseline assessment models. Table 3 summarizes the major alternatives.[11] The fundamental difference in these theories is an agent's optimization function?

[10] An important consideration is that standardized approaches to additionality can misrepresent (e.g., lead to a false positive) the behavior of early adopters of a technology or practice, to the extent that they differ from the recognized representative agent.

[11] See van den Bergh, Ferrer-i-Carbonell et al. (2000) and Venkatachalam (2008) for broader discussions of behavior theory in the context of environmental policy.

Table 3. Alternative theories of behavior for additionality and baseline models

Theory	**Description**	**Reference(s)**
Pure rationality	Assumes actors optimize their decision making, given the available information, by maximizing benefits minus costs to themselves.	(van den Bergh, Ferrer-i-Carbonell et al. 2000; Asuka and Takeuchi 2004; Shogren and Taylor 2008; Venkatachalam 2008)
Bounded rationality	Assumes that when actors make decisions they lack information, cognitive resources, and/or the time to analyze choices and achieve an optimal rational choice. As a result, actors use simplifying heuristics to reach a satisfactory, rather than an optimal, solution. A number of decision making biases have been elaborated in literature and empirically demonstrated, including: • *Certainty effect*: actors are biased towards outcomes they perceive as less uncertain; • *Reflection effect*: actors seek risk when all choices involve losses, while they avoid risk when all choices yield gains; • *Loss-aversion or endowment effect*: actors' decisions are influenced more by a potential loss than an equal-sized potential gain. • *Isolation effect*: actors' decisions focus on perceived differences in choices and ignore the elements that their choices have in common, thereby assuming they are identical; • *Habitual, behavioral inertia or status quo effect*: actors will continue to use existing decision making heuristics even after they no longer offer optimal results.	(Simon 1957; Lindblom 1959; Kahneman and Tversky 1979; DeCanio 1993; Fox and Tversky 1995; van den Bergh, Ferrer-i-Carbonell et al. 2000; Kahneman 2003; Williamson 2005; Shogren and Taylor 2008; Venkatachalam 2008; Shogren, Parkhurst et al. 2010; Kragt and Bennett 2012)
Altruism	Assumes actors make decisions to optimize the welfare (i.e., benefits minus costs) of others rather than based on their own self interest.	(van den Bergh, Ferrer-i-Carbonell et al. 2000; Shogren and Taylor 2008)

Rational actors behave "in a manner that will maximize their own well-being" (Grafton, Pendleton et al. 2001). Pure rationality, the basis of neoclassical economic theory, views behavior as predictable and consistent across situations and time by assuming constant preferences. In the context of additionality and baselines, the primary model of behavior for a rational economic actor would be a cost/benefit analysis. The variables relevant to the model would be economic, and could include both financial variables (e.g., revenue and prices) as well as non-financial variables (i.e., transaction costs and economic utility).

Typically, only private (versus public) benefits and costs would be considered under this theory of behavior, and actors would be assumed to maximize their expected private utility (van den Bergh, Ferrer-i-Carbonell et al. 2000). But this assumption does not necessarily reduce the process to a simple financial analysis. Grounding assessments in an economically rational actor theory of behavior can account for behavior where non-financial benefits and costs are significant. For example, some activities may entail public relations benefits to a company that can be valued (albeit with difficulty). Similarly, transaction costs associated with overcoming information and other barriers can be included in an economic model.[12]

There are some important classes of activities for which pure rationality is likely to be a poor theory of behavior. Bounded rationality assumes that real-world actors are not able to make perfectly rational decisions in many circumstances. "The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world—or even for a reasonable approximation to such objective rationality" (Simon 1957). Beyond the specific biases listed in Table 3, examples of other effects that may lead actors to deviate from pure rationality are a human evolutionary bias for fairness and justice as well as reference-dependent preferences (i.e., keeping up with the Joneses) (van den Bergh, Ferrer-i-Carbonell et al. 2000; Venkatachalam 2008).[13]

Both laboratory and field experiments indicate that individuals (versus firms or entire markets) exhibit behaviors that are often better represented by a theory of bounded rationality (Kahneman 2003; Venkatachalam 2008). However, firms and other institutions—as collections of individuals with institutional rules and cultures—as well as markets as a whole are not necessarily shaped by the same findings (Shogren and Taylor 2008). For example, the social and economic context of firms are likely to moderate some individual biases, while simultaneously introducing some new biases through social phenomena (Cyert and March 1963). A firm will typically have more cognitive resources to identify and analyze choices than a single individual, especially when a decision involves large costs and benefits. However, when an activity involves small decisions by many actors and transaction costs are large (e.g., residential energy efficiency projects), bounded rationality theory is likely to be highly relevant (van den Bergh, Ferrer-i-Carbonell et al. 2000). In these cases, a policy intervention may have less of an impact than would be expected under a pure economic rationality theory of behavior.

Lastly, altruism assumes that an actor behaves in a way that benefits other actors or the broader public even when it costs them. In the context of environmental protection, the behavior of some individuals can only be explained by a theory of altruism (e.g., voluntarily purchasing GHG emission offsets).[14] The challenge for modeling is that altruistic preferences are less likely to be consistent across actors. At the firm-level, behavior is less likely to be guided by altruism. Plus, actors in competitive markets are more likely to express self-interested rational preferences over pro-social norms (Reeson and Tisdell 2010).

2.3 What is an additionality and baseline model?

The specification of a policy intervention and a theory of behavior assumption are two of the key inputs into the process of developing a model to assess additionality and a baseline. Generically, a model is a mathematical and/or logical system of relationships between independent (i.e., treatment and control) variables and a dependent variable. In the context of additionality and baseline assessments, a model's purpose is to predict behavior of actors under an unobserved scenario. These models do not need to be sophisticated computer programs with hundreds of equations and variables. Preferably, they will take the form of a simple equation or set of objective rules embedded in a decision tree. The goal is a simple

[12] Under the CDM, one of the three options for the basis of a baseline scenario in the Marrakesh Accords (see paragraph 48b) specifies the following: "emissions from a technology that represents an economically attractive course of action, taking into account barriers to investment." This option implicitly assumes a rational economic actor theory of behavior (Asuka and Takeuchi 2004).

[13] A potential psychological implication of establishing a program that pays for public goods is that actors may no longer voluntarily supply them and insist on being paid instead (Gneezy and Rustichini 2000).

[14] Seemingly altruistic behavior to privately provide a public good may be better explained by treating the public good as impure. In other words, contributors benefit both from the public good as well as a "warm glow" satisfaction or social approval and prestige from the act of giving (Andreoni 1990; Menges, Schroeder et al. 2005; Kotchen 2009).

model that utilizes as few variables as possible to distinguish additional from non-additional proposals and assign a baseline to them.

The rules (e.g., equations or algorithm) in such a decision tree are a function of measurable variable values, rather than subjective considerations. The kinds of additionality tests used by offset programs, such as the CDM additionality tool, are also models. But they are models that overly rely on subjective considerations rather than objective and standardized criteria developed using input from evidence-based investigations. In some cases a model can be reduced to a single additionality threshold metric (e.g., a performance benchmark) where the model's algorithm is simply whether or not the performance of the proposed activity will be above or below a performance standard and then assigns the only offered baseline scenario candidate to the proposal.

A single model can both assess additionality and assign a baseline scenario from a list of candidates. Because performance (i.e., behavior) in the presence of the policy intervention (e.g., project scenario) is observed, it can be measured after the activity is implemented and does not require a predictive model. However, for the same specific case it is not possible to simultaneously observe behavior under baseline conditions (Gustavsson, Karjalainen et al. 2000). This fact is referred to as the fundamental problem of causal inference (Holland 1986; Collier, Seawright et al. 2004), which states that "*for a given case at a given point in time*--the researcher can observe either the presence of the cause (and of its presumed effect), or the absence of the cause (and hence potentially the absence of its presumed effect), but not both. Therefore, the researcher…must instead turn to imperfect real-world comparisons among cases" (Brady and Collier 2004).

A reason for using a decision model with clearly specified variables is that it offers greater transparency, objectivity, and replicability, as well as a falsifiable hypothesis that can then be challenged and improved upon. Scientifically credible predictions must be falsifiable, meaning that they should be specified so that a knowledgeable person could identify a case, if it exists, which demonstrates that the predictive model is incorrect. In the context of additionality and baselines, if a case was identified where proposed activity was implemented without the presence of the policy intervention and yet the applicable model said it would require the intervention to occur, then this would theoretically falsify the model. This result would then lead to the model's modification or replacement.

In the context of offset-related activities, although it is not possible to perfectly predict behavior or perfectly measure all relevant variables, neither is behavior fully random. For example, project investment decisions have explanatory variables and causal relationships. Additionality and baseline assessments will involve uncertainty, but the presence of some uncertainty should not lead to the conclusion that offsets are an unusable or impractical policy mechanism.[15] The challenge of offsets is evaluating whether the models used for assessing additionality and baselines entail so much uncertainty (i.e., error rates in additionality determinations and/or poor baseline predictions) that an offset program is inferior to alternative policy mechanisms in meeting a specific policy objective (see Part 1).

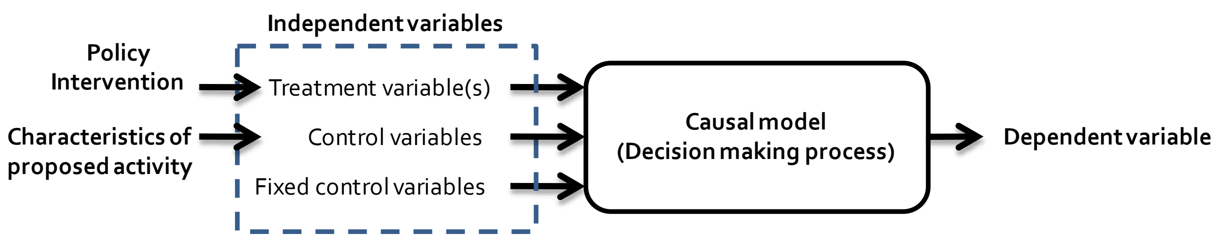
A key message of this article is that "even though [perfect] certainty is unattainable, we can improve the reliability,[16] validity,[17] certainty, and honesty of our conclusions by paying attention to the rules of scientific inference" (King, Keohane et al. 1994); and that offset mechanisms, to be credible, need to better define the scientific basis of additionality and baselines.

[15] See Begg and Van der Horst (2004) for a discussion of uncertainties with additionality and baseline scenario assessments.
[16] Applying the same procedure in the same way should produce the same results (King, Keohane et al. 1994).
[17] Are we measuring what we think we are measuring (King, Keohane et al. 1994)?

Generically, a model predicts changes in the value of a dependent variable resulting from changes in the values of independent variables (i.e., explanatory or causal variables). Treatment variables are one type of independent variable. Offset programs need to identify one or more specific treatment variables that represent their recognized policy intervention(s). The treatment variables may be identical for all classes of activities (e.g., project types or regions) or they may vary by class.[18]

Independent variables in a model, other that treatment variables, are control variables. The applicability of a model (i.e., what is included in the class of activities for a model) is defined by an acceptable range of values for select control variables, the range for some of which may be a fixed value. A model would not then be applicable to cases (e.g., proposed projects) with control variable values outside these ranges.[19] Certain control variables would also used to predict variations in behavior between cases with a class of activities (i.e., which proposals were additional and which were not). These relationships are illustrated in Figure 2.

Figure 2. Illustration of a causal model and variable relationships



Model variables used can use nominal, ordinal, or interval values. The objectivity and replicability of variable coding or scoring—such as by expert elicitation rather than direct measurement—can be increased by basing values on the answer to a specific question. (e.g., what is the capacity in mega-watts of the facility being proposed?) and a well-specified protocol.

The models described here are standardized approaches to additionality and baselines.[20] The use of a model, though, may not entirely eliminate subjectivity from the assessment process. But, by specifying models and variables, program administrators can better manage biases and thereby increase the replicability and comparability of assessments.

2.3.1 What is the dependent variable?

We have established that a policy intervention—represented by treatment variables in a model—is what causes additional activities to be proposed. What then represents the "effect" or dependent variable that is predicted or explained by a model? In the context of environmental-related offsets, the instinctive answer is often some measure of a change in pollution. However, this question actually requires more than instinct.

[18] Policy interventions could be tailored to specific classes. For example, only activities in less-developed countries could receive additional support in the form of free legal services.
[19] For example, a model may only be applicable for proposed activities where a measure of annual solar insolation is within a specified range.
[20] Standardized approaches are also referred to as benchmarks, multi-project baselines (OECD/IEA 2000), or performance standards (Hayashi, Müller et al. 2010) by some authors. Under the framework of this article, benchmarks and performance standards are just one type of model.

The aim of an additionality assessment model is to answer the following question: Will a given change in the value of the treatment variable(s), due to the removal of a policy intervention, change the outcome of a decision making process? To do so, it is first helpful to identify a limited set of alternative baseline scenarios for the class of activities (e.g., project types in a given context) being assessed. Each alternative scenario is an outcome of a decision making event, with each alternative outcome being functionally equivalent (i.e., delivering equivalent products or services). One of these alternatives must be a duplicate of the proposed activity absent the policy intervention. If the potential outcomes for a class of activities are so uncertain that a reasonable number of scenarios cannot be specified, then there will likely be policy options, other than an offset mechanism, that would better address that class of activities.

For a given class of activities, the dependant variable for an assessment model can then be specified in one of two forms. It can select one baseline scenario, out of the predetermined list of alternatives. Or it can be the likelihood of one or more alternative scenario being the correct baseline. The former type of dependent variable would be considered a deterministic classifier (i.e., it would classify a choice into one of a given set of alternatives), while the latter would be probabilistic.

If a model uses a deterministic classifier as its dependent variable, then it is not necessary for the model to be able to predict precise effect of the full range of control variable values. Instead, it would only need to capture the threshold effects of specific control variable values that correspond to a change in the classification from one alternative scenario to another. As such, a deterministic classifier model can easily be structured as a decision tree algorithm.

Within emission offset programs, additionality is commonly thought of as a deterministic variable—either a proposed activity (e.g., project) is additional or it is not. Some authors, though, have suggested additionality could or should be represented by a probabilistic variable that could then be used to discount the number of credits issued to an activity. For example, Michaelowa (2008), Schneider (2009b), Meyers (1999), and Tanwar (2007) have suggested that the emission reduction credits issued to CDM projects be discounted based on the likelihood that a project is actually additional (i.e., by the uncertainty in the additionality assessment).[21] However, such an approach presents some challenges. Such models would need to not only assign one or more baseline scenarios to a proposed activity, but they would also need to provide estimates of the probability that each scenario is the correct one.

Of the two options, the use of a deterministic dependent variable leads to a model that can more readily be tested and falsified. To falsify a probabilistic model, it would be necessary to test it on a sufficiently large and representative sample of cases to demonstrate that the probability estimates it produces are not reliable.

2.3.2 Role of control variables

Control variables are needed in these models to capture variation among proposals within a class so that they can be classified as additional or non-additional as well as assigned a baseline. This variation between proposed cases may include economic, technological, regulatory, demographic, physical, and other social factors, each of which would be represented by one or more control variables. Non-additional proposals would then be those that have control variable values within certain ranges indicating the proposal is itself classified as the baseline scenario.[22]

Control variables are also used to determine whether a proposal belongs within a class of activities. In other words, the model is only applicable to proposals that have control variable values within a given

[21] Referred to as "fractional additionality" in EPRI (2008).
[22] For example, a control variable could be the distance from the site of the proposed project to the nearest electrical transmission line.

range (e.g., variables representing geographies, climate zones, or levels of economic development). Other control variables and the remaining variation in those used to determine applicability are then used to classify proposals to one or more baseline scenarios and thereby assess additionality.

3 Applying the framework

The administrators of major GHG emission offset programs, such as CDM, have relied on a process for determining additionality that largely relies on a subjective case-by-case adjudication.[23] Facts and interpretations are argued over and the eligibility of proposed project activities are judged based on information presented, much of which is unsubstantiated (Schneider 2009a). This approach has been criticized for a variety of reasons, including whether it is scalable as offset markets grow. In response, several authors have called for moving towards more objective metrics and standardized approaches to assessing additionality and baselines (Kartha, Lazarus et al. 2004; Schneider 2007; OQI 2008; OQI 2009a; Schneider 2009a; Hayashi, Müller et al. 2010).

To standardize a process, one needs a fairly precise understanding of what the process is for. This section discusses how the framework described in the previous section can be used to develop standardized approaches to additionality and baselines. The kind of models needed by program administrators will necessarily entail compromises between accuracy and the costs to develop and apply them. But without grounding in evidence-based investigations and a framework for the resulting assessment models, all that will have been accomplished through standardization is increased uniformity without addressing issues of credibility, accuracy, or transparency.

This section explores the following questions with respect to additionality and baseline models, recognizing that further work by researchers and offset program administrators will be necessary:

- What types of policy interventions and treatment variables should be used?
- Models should assume what theory of behavior?
- How should models be built and maintained?

Clearly, if a model is overly complex or relies on variables that cannot be cost-effectively measured with acceptable certainty then it is of little use. Therefore, this article is concerned with models that have a realistic chance of being applied in the administration of real-world offset programs.

3.1 Policy intervention

A key step in developing an offset program is recognizing a type of policy intervention for the assessment of additionality and baselines. Policy makers and program administrators have a variety direct interventions they can make (see Table 4) as well as different ways of conceptualizing a policy intervention (see Table 1).

Table 4. Relevance of categories of environmental policy interventions to emissions offset policy

Categories of environmental policy interventions	**Relevance to emissions offset policy and intervention types from Table 1**
1. Economic subsidies for an activity;	Relevant under Types *A*, *B*, and *C*. Expected revenue from sale of issued offset credits. Unlike a fixed publicly funded subsidy, value is a function of market forces (i.e., supply of and demand for credits).

[23] Referred to as a project-specific approach in WRI/WBCSD (2005).The offset programs operated by the Regional Greenhouse Gas Initiative and the Climate Action Registry have instituted performance benchmark approaches, and therefore are, to a partial degree, exceptions.

2. Measures to reduce transaction costs, which can reduce the cost of a known activity for an actor and/or add to the list of known options for a given decision or behavioral choice;	
a. Information or educational campaigns;	Relevant under Types *B* and *C*. An offset program could directly deploy informational campaigns that influence behavior or cause other institutional actors to deploy such programs.
b. Changes to existing rules (e.g., remove regulations that discourage or increase the costs of certain activities);	Relevant to Types *B* and *C*. Legislation that creates offset program could also include or lead to other policy or legal changes that directly or indirectly influence behavior.
c. Providing new infrastructure (e.g., telecommunication infrastructure)	Relevant to Types *C* and potentially *B*. The offset program could directly provide new infrastructure that influences behavior (e.g., build new electrical transmission lines), which would be relevant to Type *B*. More likely, it could indirectly influence the path of infrastructure development, which would be relevant to Type *C*.
d. Research and development (R&D) funding;	Relevant to Type *C* and potentially *B*. The offset program could directly provide new funding for R&D that changes technology options in the future, which would be relevant to Type *B*. More likely, it could indirectly influence the path of R&D investment, which would be relevant to Type *C*.
e. Development of voluntary technical standards	Relevant to Type *B* and *C*. Standards could result in lower transaction costs as well as influence the path of technology development.
3. Enforced government mandates or internal organizational policies;	Offset policies are typically designed as a market-based policy instrument, not a mandate.[24]
4. Reputation-based programs;	Relevant under Type *B* and *C*. An offset program could directly recognize the superior or inferior environmental performance of actors or cause other institutional actors to deploy such programs.
a. Recognition programs for superior performing actors; and	
b. Shaming programs for poorly performing actors.	

Once these decisions are made, then the recognized type of policy intervention needs to be specified with one or more treatment variables that can be objectively measured or coded by experts using a elicitation protocol (Morgan and Henrion 1990).[25] It is preferable to define the type of policy intervention as narrowly as possible without missing significant effects that are likely to alter actor behavior. Program

[24] Theoretically, an offset policy could mandate behaviors and still issue credits for only those activities that are deemed to have occurred because of the presence of the mandate (i.e., some activities may take place even in the absence of the mandate). The effect would be to subsidize those activities that were caused by the mandate and not subsidize those that were not while still ensuring the activity was universally implemented.

[25] Expert elicitation is the synthesis of opinions of experts on a subject where there is uncertainty due to insufficient information. An elicitation protocol involves a structured approach to questioning experts so as to obtain estimates of parameter values and uncertainties with minimum bias.

administrators can then quantify—or develop rules for quantifying—the value of each treatment variable when the policy intervention is absent.

Returning to the types of policy interventions in Table 1, Type *A*—related to the economic subsidy effect resulting from the expected value of offset credits (i.e., price signal)—is the most reduced form of a policy intervention. Offset policy is typically thought of and designed as a mechanism that issues credits with that have some market value. The treatment variable would be the expected revenue to an activity from the sale of offset credits. The variable is an "expected" value because the actual number of credits issued, as well as their future selling price, will be uncertain.

If an offset program actively provides other types of direct support to activities that is intended to influence behavior (e.g., in-kind project development assistance), then it may be appropriate to include other treatment variables in the model, in keeping with Type *B*. Taking from the literature on transaction costs, these other treatment variables may be coded in terms of their economic value (Williamson 2005). Analogously, non-financial barriers to a class of activities can also be represented in models as transaction costs using control variables, thereby building relationships in the model with treatment variables (Schneider 2007).

A practical way forward is for offset policy makers and program administrators to limit what type of policy intervention and treatment variables they recognize to Type *A* (i.e., expected economic value of offset credit revenue) except where significant program resources are dedicated to other categories of policy interventions that would fall under Type *B* (i.e., other measures directly taken by offset program or policy). This approach recognizes that there may be some equilibrium (i.e., spillover) effects (i.e., Type *C*) that are omitted; however, specifying these effects with treatment variables and incorporating them into models would be challenging and highly uncertain. Limiting the number of treatment variables both simplifies the modeling and results in more conservative additionality assessments. The implications of Type *D* (i.e., emergence of climate change as an issue for policy or decision making consideration) make modeling, and the entire concept of offsets, impractical and should not be given serious consideration.

3.2 Theory of behavior

Taking as a given that altruism is not a reliable theory of behavior for an additionality and baseline model, the choice of theory reduces to either pure or bounded rationality. To completely reject the rational actor theory one would have to believe that: 1) actors are primarily irrational and do not maximize their own utility, or 2) that a model based on pure rationality misses some key factors that significantly affect behavior in a given context. The first belief can be discarded. To further explore the second, Table 5 describes the relevance of the bounded rationality effects listed in Table 3 to additionality and baselines.

Table 5. Bounded rationality biases in the context of additionality and baselines

Bounded rationality biases	Relevance to additionality and baseline models
Certainty effect: actors are biased towards outcomes they perceive as less uncertain;	Although two scenarios may offer the same net benefits when risks are accounted for, actors may be biased towards choosing the scenario in which the benefits and/or costs are more certain. A stronger or weaker policy intervention may be required to change behavior than would be predicted by purely rational actor theory.
Reflection effect: actors seek risk when all choices involve losses, while they avoid risk when all choices yield gains;	The set of choices (scenarios) will typically involve both losses and gains; therefore, this effect has limited relevance.
Loss-aversion or endowment effect: actors' decisions are influenced more by a potential loss than an equal-sized potential gain;	A change in behavior may involve the exchange of an existing set of benefits for another set. Although both may offer equal net benefits, an actor may weigh the loss as being greater than the gain. A stronger policy intervention (i.e., larger change in treatment variables) may be required than would be predicted by purely rational actor theory.
Isolation effect: actors decisions focus only on perceived differences in choices and ignore elements that choices have in common, thereby assuming they are identical;	Assuming actors have perfect or near perfect information about the costs and benefits associated with each scenario and that they accurately interpret the information may lead to errors in the model. A stronger policy intervention than would be predicted by a purely rational actor theory may be necessary to overcome problems of poor information.
Habitual, behavioral inertia or status quo effect: actors will continue to use existing decision making heuristics even after they no longer offer optimal results.	A stronger policy intervention may be required to change existing behaviors than would be predicted by purely rational actor theory.

A practical way forward for offset policy makers and program administrators is to use a purely rational "economic" actor theory of behavior as the basis for their models, except in contexts where it is expected that specific biases are likely to be significant and where those biases can be analytically incorporated into a model.[26] This way forward does not assume that actors (e.g., business managers, facility operators, project developers or investors) should be treated as having perfect information, perfect foresight, zero un-monetized transaction costs, unlimited access to capital, an exclusive focus on financial factors, operate in perfectly competitive markets, or be ideal in other ways. However, for many classes of activities, economic rationality—which can still account for political and other factors through the inclusion of transaction costs—is an appropriate assumption for emission reduction activities, given the significant long-term investments often involved.

In some contexts, more realistic models can be developed if likely actor biases are incorporated. For example, biases can be taken into account if an emission reduction project involves many distributed actors making small investments (e.g., residential energy efficiency), in contrast to larger investment decisions made by a single large firm.

[26] Greiner and Michaelowa (2003), Shogren and Taylor (2008), and McFadden (1999) make similar recommendations for CDM and in the broader context of environmental policy and economic analysis, respectively. The assumption of economically rational actor is also functionally adopted in the CDM's investment analysis guidance (CDM 2009).

3.3 Creating and maintaining models

There are a number of desired policy outcomes resulting from the use of the kind of models for assessing additionality and baselines (i.e., standardized approaches) described here, including:

- public confidence and credibility with stakeholders;
- accuracy of additionality and baseline assessments (i.e., minimize error rates);
- predictability (i.e., low uncertainty) for actors (e.g., project investors) in the outcome of the process;
- objectivity in variable value measurements and coding as well as model algorithms (e.g., decision points in decision tree are not based on subjective judgments);
- option for administrative flexibility to address unique circumstances;
- incorporation of processes for continual improvement of models;
- minimized transaction costs (e.g., for data collection, model application, quality assurance);
- resistance to manipulation (i.e., gaming);
- high transparency in the technical and empirical justification for the development and application of models; and
- ability for stakeholders to test, challenge, and falsify models.

Many of these outcomes involve unavoidable tradeoffs, such as flexibility versus predictability or accuracy versus cost. It is important to recognize, however, that some frameworks for assessing additionality and baselines can be worse in achieving all outcomes, while others can be better. As stated earlier, one of the main points of this article is that it is possible to better, although not perfectly, achieve all of the above outcomes with a more rigorous understanding of and approach to additionality and baselines.

After clearly specifying the policy intervention recognized by an offset program, the steps for creating and maintaining models can be summarized as follows:[27]

1) clearly specify the class of activities (e.g., type of project proposals, contexts, and actors) to which a model is applicable;
2) identify the dependent, treatment, and control variables using evidence-based investigations of the class of activities;
3) elaborate a limited set of alternative baseline scenario candidates for the class;
4) create a model to classify a given case into one or more scenario candidates; and
5) test the model and revise (i.e., maintain) based on new evidence and changes in the applicable context.

These steps are ideally completed in combination with a rigorous investigation of the relevant class of activities, including input from technical experts and use of case studies. An analogy for this type of investigative work would be studies performed for rulemaking, in which regulators develop a deep understanding of the technological, economic, and social issues associated with a class of activities in a given context prior to elaborating a regulatory standard.

Are these steps likely to require significant investment of time and resources by offset program administrators? For many classes of activities, the answer will be yes. If quality of standardized approaches is to be assured, programs will require significant upfront and ongoing investments related to methodology development, monitoring, reporting, and verification (OQI 2008; Hayashi, Müller et al. 2010). However, if coordinated, multiple offset programs could collaborate to share the cost of model development and maintenance, and in doing so ensure that offset credits are based on common definitions, meet common quality criteria, and are fungible.

[27] Taken, with modification, from George and Bennett (2004).

3.3.1 Specifying a class of activities

One of the advantages of offsets as a policy mechanism is that they can harness markets to search out new low-cost mitigation opportunities and promote innovation. Therefore, offset programs are valuable in applications where policy makers do not, *a priori*, know where activities should be implemented or how to implement them. Yet, standardized approaches do require program administrators to clearly specify the classes that are likely to be implemented under their program as well as other applicability criteria for assessment models.

Each class of activities will generally require its own model and applicability criteria. Some individual classes of activities may be globally uniform in their technological, economic, and social characteristics. In such situations a single model may be appropriate for that class in all contexts. In most other situations, though, it will be more appropriate to limit a class to activities, for example, in a single country, in urban areas, in regions with low precipitation, or where a reliable electricity supply is available.[28] The purpose of this step is to define a class's boundaries such that all the cases and actors in it are sufficiently homogeneous to assume that the same independent variable values will predict accurately (King, Keohane et al. 1994).[29] If a class is deemed insufficiently homogenous then it can be broken up into more than one class with separate models. Although it has been used by leading GHG offset programs (e.g., CDM combined tool for additionality and baselines[30]), the desired policy outcomes listed above are unlikely to be best met by a single globally uniform assessment tool (i.e., model) applicable to all classes of activities.

Once a class of activities is specified, a limited number of alternative scenario candidates can be elaborated along with procedure for quantifying the performance (e.g., emissions) of each candidate. This quantification can be a function of empirical data from representative cases, expert judgment, and/or *ex post* metrics (e.g., scaled to production output in the observed project scenario).

3.3.2 Specifying variables

If not directly measured then treatment and control variables can be estimated using a proxy variable that is highly correlated with the original variable or an expert elicitation protocol that controls for bias. For all variables, it is fundamental to the credibly of the offset programs that the process of collecting data and coding variable values be transparent.

For the dependent variable in most additionality and baseline models, a practical way forward is to use a deterministic dependent variable that classifies a proposal with one baseline scenario (Figure 3). For simple or narrowly specified classes of activities where there are only two alternative baseline scenario candidates, it may be possible to reduce the model to a simple performance standard.

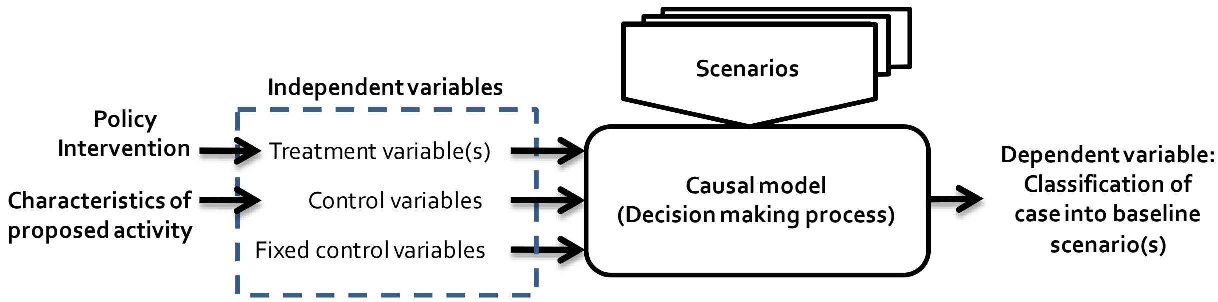
---

[28] This step is similar to defining applicability conditions for CDM project methodologies, but should be based on objective control variable value ranges that clearly differentiate cases that are part of a class from those that are not. It is also similar to the defining of boundaries for estimating market penetration rates (e.g., common practice analysis), as discussed by Kartha, Lazarus et al. (2005).

[29] Referred to as "causal homogeneity" by King, Keohane et al. (1994).

[30] See http://cdm.unfccc.int/methodologies/PAmethodologies/approved.html

Figure 3. Illustration of causal model incorporating scenario inputs and dependent variable as classifier



However, for classes in which a model is expected to have a high predictive uncertainty, the use of probabilistic dependent variable may be considered. The resulting probability estimates could then be used to discount the number of credits issued to an activity. The important caveat to such an approach is that probability estimates need to be specific to a given proposal (i.e., a function of that proposal's control variable values) not common to the entire class. If the additionality of an entire class of activities is highly uncertain, and it is not possible to estimate this uncertainty separately for each proposal, then the entire class is likely to be inappropriate for inclusion in an offset program.[31]

Within a class, there may be the expectation that two or more alternative scenario candidates could be the correct baseline scenario with equal probability. For these classes, models and dependent variables can then be designed to allow the classification of more than one scenario as the likely baseline scenario. The performance metric (e.g., emissions) against which credits are then calculated could be the more conservative of the likely baseline scenarios or some combination or average.[32]

The ultimate focus of investigations used to develop models is to identify easily measured and verified control variables—representing characteristics of each specific proposal versus the entire class—that can be used to assign baseline scenarios to proposed activities and thereby distinguish additional cases from non-additional cases.

The behavior of actors for some classes of activities may be well-predicted by a traditional financial investment analysis (e.g., discounted cash flow analysis),[33] in which case a standardized form of such an analysis would be an example of the type of model described here. It would need to include standardized variable values representing the typical financial characteristics of the entire class (versus each individual proposed activity or project) including typical capital costs, operating costs, revenue sources, and a default hurdle rate of return. Other context dependent variables could include Institutional Investor country rankings—as suggested by Griener and Michaelowa (2003)—or average ratings on related financial instruments by credit rating agencies—as suggested by Shrestha and Timilsina (2002). The model's probability of error would then be informed by the difference in the estimated IRR for the proposed activity and the predicted baseline scenario (i.e., a smaller change in the IRR would be an indicator of increased likelihood of a false positive or false negative) (Au Yong 2009). More generally,

[31] As noted by Castro and Michaelowa (2010), Bushnell (2010), and Kollmuss, Lazarus, et al. (2010); applying discounting to an entire class of activities is more likely to exclude truly additional activities because they will have higher costs than non-additional activities. See Kollmus, Lazarus, et al. (2010) for discussion of discounting applied to GHG emission offset programs.
[32] Effectively, the CDM combined and build margin approach for electricity generation projects uses an average of two historical-based scenarios (Kartha, Lazarus et al. 2004).
[33] Financial analysis is already used by a number of GHG offset programs, such as by CDM (Chomitz 1998; Rentz 1998; Meyers 1999; Greiner and Michaelowa 2003; Sutter and Parreño 2007; Au Yong 2009).

the overall likelihood of model error will increase as the ratio of offset credit revenue to the total revenue of the project decreases (Bode and Michaelowa 2003).

3.3.3 Creating models

There are several scientifically credible approaches available for informing the model creation process and collecting default data that is representative of a class of activities. Many of these approaches can also be used to test models and include, but are not limited to:

- engineering calculations and physical process modeling;
- economic and other statistical and regression modeling;
- discrete choice surveys and analyses;
- formal expert elicitation;
- statistical surveys and market research studies;
- case studies;
- field experiments utilizing control groups; and
- laboratory experiments (e.g., utilizing techniques from behavioral and experimental economics).

Chomitz (1998) and Gustavsson, Karjalainen et al. (2000) discuss the use of case studies with separate treatment and control groups for baseline assessments.[34] Such cases can be drawn from historical records, observed in the field, or created through experimentation. By comparing cases, with and without a policy intervention while holding all other factors constant (or at least controlling for these other factors), it is possible to investigate whether a policy intervention causes behavior change. Methods for designing and analyzing case studies and control groups for policy interventions are well-developed (Khandker, Koolwal et al. 2010).

These approaches include both those that use larger numbers of case studies to achieve statistical representative samples (King, Keohane et al. 1994) as well as those that focus on in-depth studies of a small number of cases (George and Bennett 2004). The number of cases does not necessarily have to be large to enable valid inferences to be made, especially where the dependent variable is narrowly defined (e.g., a deterministic classifier). A combination of approaches that involves a few in-depth studies, to inform the creation of a model, and then a larger sample of cases to test the model is likely to be a reasonable way forward.

Offset program administrators can also use randomized trials with a control group in the creation and testing of models. Both laboratory experiments using techniques from behavioral economics (Weber and Camerer 2006; Kagel and Roth 2007) and social science field experiments for investigating the efficacy of policy interventions (Greenberg, Linksz et al. 2003; Harrison and List 2004) are recognized methods of causal inference and can produce valuable information to inform and justify offset program models.

When selecting case studies or designing experiments, program administrators need to ensure there is sufficient variation in the dependent variable as well as treatment and control variables to make statistically valid causal inferences. Variation in the dependent variable means that the investigation should include cases where the type of proposed activity is implemented as well as cases where alternatives were implemented. The variation in control variables should cover the range of values corresponding to the defined boundaries of the relevant class of activities. And variation in treatment

[34] Chomitz (1998) outlines two approaches for gathering information to assess additionality and baselines. He refers to these two approaches as comparison groups (i.e., control groups) and simulation (i.e., financial or behavioral models). However, to predict additionality, a model (i.e., simulation) of some form is always needed. Case or "comparison" studies can be used to inform and test a model, but are not a separate and distinct approach that can substitute for an assessment model.

variables means that the investigation includes cases where the policy intervention is present as well as where it is absent. The latter group of cases is referred to as a control group.

For some classes of activities, it may be challenging or impractical to find concurrent cases to include in a control group (Meyers 1999).[35] This challenge will, in part, be a function of how the policy intervention is defined. Interventions of Type *B* or *C* will cause greater difficulties (refer to Table 1). Investigators can also consider the use of historical cases, where data are available and historical conditions are deemed sufficiently representative, or they can utilize experiments and other techniques such as discrete choice surveys (Hoyos 2010). This work will need to address issues of systematic error in variable measurement, case study selection bias, unidentified control or treatment variables, and case study independence.[36]

To facilitate the development of an offset market, the development of standardized approaches may focus on classes of activities that are expected to be the most common and have readily available case studies. Even once a model is developed and tested, a process for appeals and exceptions will be needed as well as a feedback for the findings from this process to be incorporated as models improvements.[37]

For activities that are actually a bundle of many smaller activities (e.g., residential lighting retrofits under programmatic CDM), it may be preferable to utilize a probabilistic dependent variable. For these classes the model is predicting the collective behavior of many individual actors; probability estimates could then represent the fraction of free riders in the population and be used for discounting credit issuance.[38]

For most classes of activities, there will be an unavoidable element of expert judgment involved in the creation of models due to the challenge of locating ideal case studies in sufficient numbers to provide statistically unambiguous results. Program administrators will also need to make judgments regarding the conservativeness of additionality assessments and baselines. Further, program administrators will have to make judgments regarding how additionality and baseline models address existing or future government policies that promote behaviors contrary to the objectives of the offset program (e.g., subsidies that promote the consumption of fossil fuels or encourage deforestation). As asked by Chomitz (1998), "[s]hould baselines be evaluated under prevailing prices and policies, or in a hypothetical distortion-free policy environment?"[39] Ultimately, the resolution of this issue is a matter of broader political and offset policy design considerations.[40] However, the development of standardized approaches will be hampered if policy makers decide that an existing policy be excluded from the definition of alternative baseline scenarios. Excluding existing policies increases the complexity of models because they must then model behavior under conditions in which both the offset program's policy intervention and the other excluded policy are absent. Such a decision will also decrease the likelihood that representative case studies can be identified, forcing administrators to rely to a greater extent on other investigation and testing approaches (e.g., lab experiments).

3.3.4 Testing models

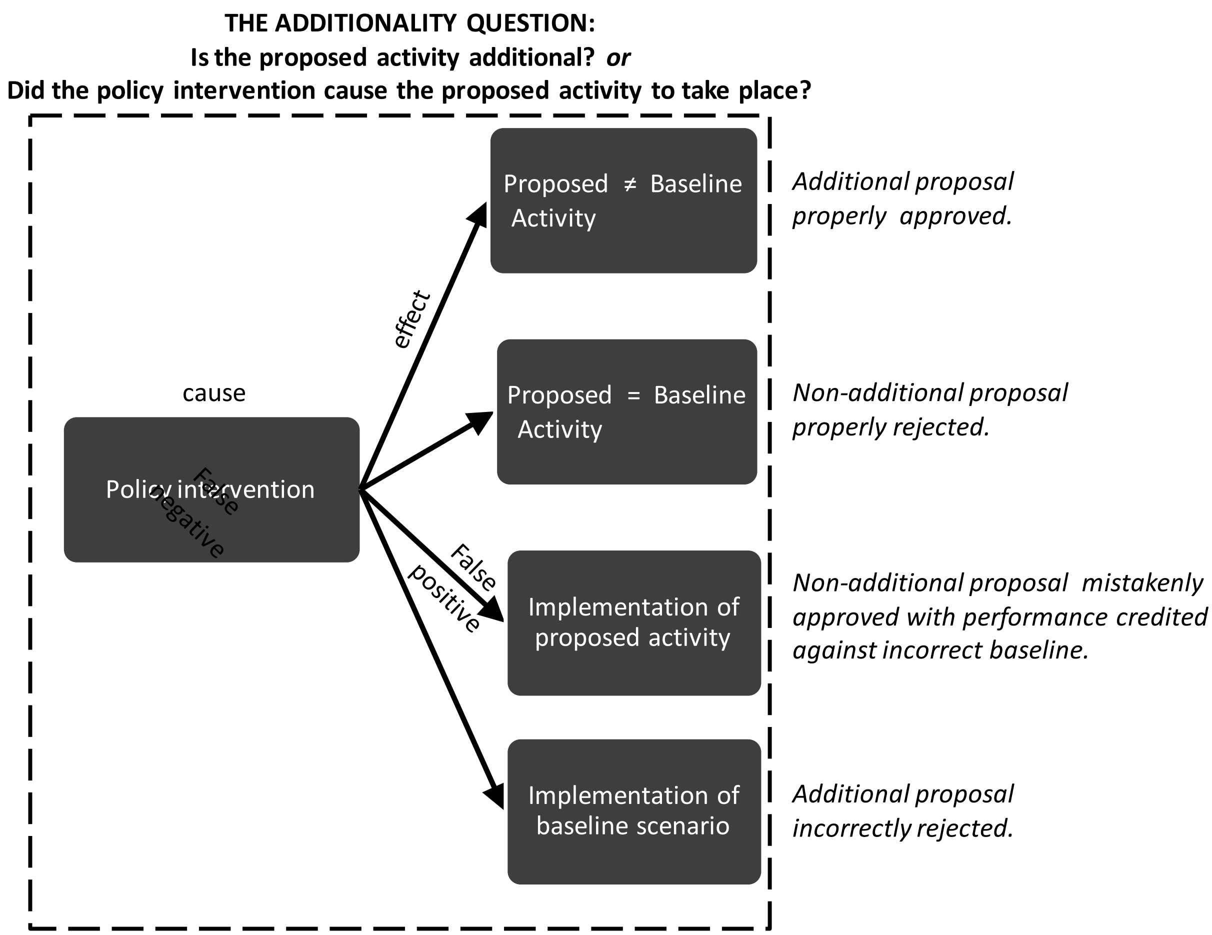
For a given class of activities, there will be variability in the behavior of actors that may not be captured by a model, thus leading to model error. However, through testing, offset program administrators can better understand the sources and magnitude of model error and then make improvements. Testing can

[35] As an offset program expands, it will also become increasingly challenging to locate representative cases to include in a control group.
[36] See King, Keohane et al. (1994) and George and Bennett (2004) for guidance on using case studies.
[37] Rules, such as a fee for appealing, may be introduced to discourage unjustified utilization of bespoke assessments.
[38]See the U.S. EPA's Conversation Verification Protocol for an example of models used to discount for free riders (Vine and Sathaye 1999).
[39] Also see He and Morse (2010) for an example from wind power CDM projects in China.
[40] For example, see issues related to the handling of E+/E- issues under the CDM (PDF 2009; PDF 2010).

also lead to the conclusion that model error is both high and intractable, and provide a basis for future proposals within that class being rejected from inclusion in the offset program.

In contrast to "project-specific" determinations and arbitrarily set performance benchmarks and market penetration thresholds, a key advantage of using the framework for standardized approaches described here is that the resulting models are subject to falsification. Model falsification can occur on the basis of a false positive (type I error), where the model says a proposal is additional when in reality it is not, or false negative (type II error), where the model says a proposal is not additional when in reality it is (Figure 4) (Trexler, Broekhoff et al. 2006). Table 6 summarizes the information provided by testing a model that uses a dependent variable that is a deterministic classifier against evidence from case studies or experiments. As illustrated in the table, it is possible to falsify a model based on a false positive case, but it is more difficult to do so with false negatives. Testing is, therefore, more useful to ensure a model is conservative in its assessment of additionality than it is to ensure it is unbiased.[41] Because it is easier to improve the accuracy of a model with respect to false positives, policy makers may have to accept that even rigorously tested models will be more prone false negative errors (i.e., falsely reject proposals that are truly additional).

Figure 4. Additionality cause and effect relationships include false positive and negative errors



[41] Testing for false positives will need to consider the effect of early adopters and other atypical actors (i.e., outliers) within certain classes of activities before a model is deemed completely falsified.

Table 6. Testing a deterministic dependent variable model with case studies

True status of case	With policy intervention		Without intervention (control group)	
	Proposal implemented	**Proposal not implemented**	**Proposal implemented**	**Proposal not implemented**
False positive: model says case is additional but it is not	No new information	Model may need improvement, but case does not falsify model	*Falsification case*	Case supports model
False negative: model says case is not additional but it is	Not applicable, proposal is rejected, therefore full policy intervention is not applied to the case		Case supports model	Case supports false negative claim, but does not falsify model

Finally, models cannot be validated by testing them against the same set of cases that were used to develop the model. And part of rigorous testing is to consider alternative policy interventions and theories that might better explain the observed behavior (e.g., an investor may be willing to take a loss on an investment to gain experience in a new market).[42]

3.4 Crediting periods and other temporal aspects of additionality and baselines

In the context of GHG emission offset programs, the temporal issues related to additionality and baselines have been dealt with through the use of a predefined crediting period during which an activity, once approved, is deemed to remain additional through the crediting period. However, how long a specific activity or a class of activities is actually additional is a function of factors such as capital lifetime of equipment and technological and market changes. Baseline scenarios inherently include predictions of these factors in an alternate future where the policy intervention is absent.

Instead of a fixed crediting period of multiple years, the additionality and baseline for an approved activity could be more frequently reassessed as new evidence is collected, and then adjusted if evidence indicates that earlier predictions and assumptions were incorrect. However, prior knowledge that additionality and baselines will be reassessed on an ongoing basis will itself have an effect on the behavior of actors. It will increase their uncertainty and hence reduce the effect of the policy intervention. In other words, the act of reassessing, if applied retroactively to approved proposals, will reduce the ability of an offset program to produce behavior change.

The use of a predetermined crediting period eliminates or reduces this uncertainty by promising a fixed eligibility period. They also reduce the cost of administering offset programs, which would otherwise need to repeatedly assess the additionality and baseline for each approved activity.

My focus has been on the assessment of additionality and baselines when the original determination of eligibility for a proposed activity is made. Reassessing proposals later in time in the light of new evidence is important for improving the quality of models and adjusting the length of fixed crediting periods. However, to reduce uncertainty to program participants and the burden on program administrators, a practical way forward is to continue to rely on fixed crediting periods and only apply improved models

[42] It may be possible to apply a *reductio ad absurdum* argument on a macroeconomic scale, in which one asks the theoretical question: if the policy intervention was removed would all of the activities in the class claiming to be additional have happened? For examples, see Wara and Victor (2008) in relation to CDM wind energy projects in China and Gillenwater (2008) and OQI (2009b) in relation to the renewable energy certificate voluntary market in the United States.

and rules to future proposals rather than retroactive changes to the baselines of previously approved activities.[43]

If an offset program recognizes a policy intervention of Type *B* (i.e., *A* + Measures directly taken by offset program or policy) or Type *C* (i.e., *A* + *B* + Indirect market spillover effects), then other temporal factors will also need to be considered. For example, for the identification of baseline scenario candidates, does the program assume the policy intervention never existed? Or should it assume that the historical effects of the policy intervention are a given and only the future effects of the policy intervention are absent in the baseline scenario? Another situation is if a new technology is deemed to be additional (i.e., on a positive list) because its invention is recognized to be the result of the policy intervention (i.e., Type C). Then, how long should this technology continue to be considered additional?

4 Discussion and conclusion

Media reporting on GHG offset programs often focus on the challenge of "proving" the additionality of proposed activities (i.e., projects).[44] The approach frequently used by reporters, and some researchers, is to ask project developers or investors whether they would have undertaken the project "anyway" (IR 2008; FOE 2009; Haya 2009).[45] This type of questioning, though, is a poor form of inquiry for reasons discussed above (e.g., asymmetric information and cognitive biases). Asking interviewees to predict causal effects in complex situations is unlikely to produce reliable conclusions (Chomitz 1998). More importantly, it does not lead to a model that can predict behavior in other cases and have its reliability tested.

Historically, GHG offset programs that have focused on assessing the intentions of individual actors (e.g., project developers and investors) have put the cost, burden, and uncertainty of the assessment process onto those actors as well as third-party auditors (i.e., validation/verification bodies). A process that relies on the presentation of an argument by project proponents followed by a case-by-case adjudication by program administrators requires little investment to set up. But it is more dependent on ad hoc rationalizations, and is therefore less consistent across assessments. In long-term for offsets programs to achieve scale and greater cost-effectiveness, they will likely need to shift from this judicial approach to one akin to a regulatory standard setting based on causal inference investigations and carefully elicited expert judgment. This shift to more standardized approaches will require greater upfront investments by GHG offset programs in model creation, testing, and maintenance.

The case-by-case assessment of a project developer's motivations and intentions may also be creating an unwelcome perverse incentive (i.e., moral hazard). Consider two near identical offset activity proposals. The only difference between the two proposals is that the actor behind the first proposal is of exceptional management and technical ability and the actor behind the second has below average ability. The first proposed activity is predicted to be profitable even without the policy intervention, and is therefore determined to be non-additional by an offset program assessing additionality on a case-by-case basis. However, because of the second actor's poor management practices, the second proposed activity is more costly, is unprofitable, and therefore deemed additional by the offset program using the same assessment process. If the additionality of these two cases is judged independently, then it is possible to have a system that promotes poorly run businesses over those that are better run. This problem can be avoided if processes for addressing additionality and baselines are built upon standardized models of a representative actor in a given context that assume all actors are of average competency.

[43] Revisions may be warranted, however, for proposals where fraud (e.g., submission of knowingly false information) is found.
[44] For example, see Monbiot (2006), Harvey and Fidler (2007), and Mukerjee (2009).
[45] Unfortunately, the recognized policy intervention is often not explicitly or precisely defined in their questioning.

Will the models developed using the framework described here be perfect? Certainly not. But an informed approximation is not the same as a groundless guess. Additionality and baseline models entail imperfect predictions of actual behavior, and for a given model with given estimators, there is a trade-off between false positive and negative error rates. Yet, models and estimators can be revised and improved. And so, in contrast to the conclusions of Trexler, Broekhoff et al. (2006)—that any reduction in false positive error rates will be lead to an equivalent increase in false negative error rates—it is possible to better manage model uncertainty and reduce, although not eliminate, *both* false positive and false negative errors by developing better models.

For causal inference models to be scientifically credible, it must be possible for others to present evidence through their own testing that a model is wrong. Given conflicting evidence, offset programs then need processes to consider whether and when to revise, replace, or eliminate a specific model.

Although standardized approaches should be as transparent and objective as possible, some subjectivity in predicting environmental, economic, and technological trends associated with baseline scenarios is unavoidable. But, if the assumptions behind these judgments are made explicit, then it is possible to improve models as new evidence is gathered. If standardized approaches are not grounded in good causal inference, then they will simply result in the replacement of a subjective case-by-case process with subjective judgments made at the class (e.g., project-type) level.

Applying concepts and the recommendations discussed here to offset policy, the generic definitions for additionality and baseline provided in Part 1 can now be tailored as follows:

Additionality is the property of an activity being *additional*. A proposed activity is *additional* if policy interventions are deemed to be causing the activity to take place. The policy interventions are recognized to be the expected economic value of offset credit revenue as well as other measures directly taken by an offset program. The occurrence of additionality is determined by assessing whether a proposed activity is distinct from its baseline (see below).

A **baseline** is a prediction of the quantified amount of an input to or output from an activity resulting from the expected future behavior of the actors proposing, and affected by, a proposed activity in the absence of one or more policy interventions, holding all other factors constant (*ceteris paribus*). These actors are assumed to be economically rational except in contexts where it is expected that specific behavioral biases are likely to be significant and where those biases can be analytically incorporated into a deterministic model used for assessing additionality and assigning a baseline. The conditions of a baseline are described in a baseline scenario.

Some authors have posited that offsets are fundamentally "a fiction" because they are too dependent on subjective judgments (Wara and Victor 2008; Millard-Ball and Ortolano 2010). A more thoughtful evaluation, though, is based on the question of whether offset program administrators can create models for additionality and baseline assessments for a given class of activities that are good enough to provide net benefits greater than the alternative policies available to achieve a given objective. For some classes of activities in some contexts, it may be impractical to develop a sufficiently reliable and evidence-based predictive model. These classes can then be excluded from offset programs and addressed with other policy mechanisms.

In conclusion, offset policy makers and program administrators should think in terms of policy interventions, theories of behavior, and objective models and variables when developing standardized approaches. Their continued failure to recognize a well-defined policy intervention and associated theory of behavior leaves the additionality and baseline assessment process to the vagaries of politics and *ad hoc* subjective rationalizations.

The next step is for researchers and policy makers to apply the framework presented here to develop and test models for specific classes of activities (e.g., project types in a given context) as well as investigate lessons learned from existing offset programs, such as the Climate Action Registry, where more standardized approaches are being developed.

Acknowledgements

I would like to acknowledge the useful input and review comments received from Derik Broekhoff, Bindiya Patel, Tim Stumhofer, Mike Burnett, Axel Michaelowa, Mark Trexler, Hoi Wen Au Yong, and Michael Oppenheimer. The author is solely responsible for any errors or omissions.

References

Andreoni, J., 1990, "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal* **100**(401): 464-477.

Asuka, J. and K. Takeuchi, 2004, "Additionality reconsidered: lax criteria may not benefit developing countries." *Climate Policy* **4**: 177-192.

Au Yong, H. W., 2009, Investment Additionality in the CDM. *Technical Paper*. Edinburgh, Ecometrica Press.

Begg, K. and D. Van der Horst, 2004, "Preserving Environmental Integrity in standardised baselines: The role of additionality and uncertainty." *Mitigation and Adaptation Strategies for Global Change* **9**(2): 181-200.

Bennett, K., 2010, "Additionality: The Next Step for Ecosystem Service Markets." *Duke Environmental Law & Policy Forum* **20**: 417-438.

Bernow, S., S. Kartha, et al., 2001, "Cleaner generation, free-riders, and environmental integrity: clean development mechanism and the power sector." *Climate Policy* **1**: 229-249.

BIS, 2009, Research to improve the assessment of additionality, UK Department for Business, Innovation & Skills. BIS Occasional Paper No.1.

Bode, S. and A. Michaelowa, 2003, "Avoiding perverse effects of baseline and investment additionality determination in the case of renewable energy projects." *Energy Policy* **31**(6): 505-517.

Brady, H. E. and D. Collier, Eds., 2004, *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Oxford, UK, Rowman & Littlefield Publishers.

Bushnell, J. B., 2010, The Economics of Carbon Offsets. Cambridge, MA, National Bureau of Economic Research. No. 16305**:** 14.

Castro, P. and A. Michaelowa, 2010, "The impact of discounting emission credits on the competitiveness of different CDM host countries." *Ecological Economics* **70**(1): 34-42.

CBO, 2009, The Use of Offsets to Reduce Greenhouse Gases. *Economic and Budget Issue Brief*. Washington, D.C., Congressional Budget Office**:** 8.

CDM, 2009, "Guidelines on the Assessment of Investment Analysis." Clean Development Mechanism, Executive Board. EB 51 Report, Annex 58, (Version 03), from http://cdm.unfccc.int/EB/051/eb51_repan58.pdf.

Chomitz, K. M., 1998, Baselines for Greenhouse Gas Reductions: Problems, Precedents, Solutions. Washington, DC, Development Research Group, World Bank. Draft for discussion (rev. 1.4).

Collier, D., J. Seawright, et al., 2004, The Quest for Standards: King, Keohane, and *Verba's Deisigning Social Inquiry*. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. H. E. Brady and D. Collier. Oxford, UK, Rowland & Littlefield Publishers**:** 21-50.

Cyert, R. and J. March, 1963, *Behavioral Theory of the Firm*. Oxford, Blackwell.

DeCanio, S. J., 1993, "Barriers within firms to energy-efficient investments." *Energy Policy* **21**(9): 906-914.

EP, 2008, Additionality Guide: A standard approach to assessing the additional impact of interventions. London, English Partnerships. Method Statement (Third Edition).

EPRI, 2008, Overview of Different Approaches for Demonstrating Additionality of Greenhouse Gas Emissions Offset Projects. *Background Paper for the EPRI Greenhouse Gas Emissions Offset Policy Dialogue Workshop 2*. Washington, DC, Electric Power Research Institute.

Epstein, L. and G. King, 2002, "The Rules of Inference." *The University of Chicago Law Review* **69**(1): 1-133.

Ferraro, P. J. and S. K. Pattanayak, 2006, "Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments." *PLoS Biology* **4**(4): 0482-0488.

Fischer, C., 2005, "Project-based mechanisms for emissions reductions: balancing trade-offs with baselines." *Energy Policy* **33**(14): 1807-1823.

FOE, 2009, A Dangerous Distraction: Why offsets are a mistake the U.S. cannot afford to make. Washington, DC, Friends of the Earth**:** 28.

Fox, C. R. and A. Tversky, 1995, "Ambiguity Aversion and Comparative Ignorance." *Quarterly Journal of Economics* **110**(3): 585–603.

George, A. L. and A. Bennett, 2004, *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA, MIT Press.

Gillenwater, M., 2008, "Redefining RECs--Part 1: Untangling attributes and offsets." *Energy Policy* **36**(6): 2109-2119.

Gneezy, U. and A. Rustichini, 2000, "A Fine is a Price." *Journal of Legal Studies* **XXIX** 1-17.

Grafton, R. Q., L. H. Pendleton, et al., 2001, A Dictionary of Environmental Economics, Science, and Policy. Cheltenham, UK, Edward Elgar**:** 362.

Greenberg, D., D. Linksz, et al., 2003, *Social experimentation and public policymaking*. Washington, D.C, Urban Institute Press.

Greiner, S. and A. Michaelowa, 2003, "Defining Investment Additionality for CDM projects--practical approaches." *Energy Policy* **31**(10): 1007-1015.

Gustavsson, L., T. Karjalainen, et al., 2000, "Project-based greenhouse-gas accounting: guiding principles with a focus on baselines and additionality." *Energy Policy* **28**(13): 935-946.

Harrison, G. W. and J. A. List, 2004, "Field Experiments." *Journal of Economic Literature* **XLII**: 1009-1055.

Hartley, J. E., 1996, "Retrospectives: The origins of the representative agent." *Journal of Economic Perspectives* **10**: 169-177.

Harvey, F. and S. Fidler, 2007, Industry caught in carbon smokescreen. *Financial Times*. London. 25 April.

Haya, B., 2009, Measuring Emissions Against an Alternative Future: Fundamental Flaws in the Structure of the Kyoto Protocol's Clean Development Mechanism. Berkeley, CA, Energy and Resources Group, University of California, Berkeley. Working Paper ERG09-001.

Hayashi, D., N. Müller, et al., 2010, Towards a more standardised approach to baselines and additionality under the CDM: Determining nationally appropriate performance standards and default factors. Zurich, Perspectives GmbH.

He, G. and R. K. Morse, 2010, Making Carbon Offsets Work in the Developing World: Lessons from the Chinese Wind Controversy. Stanford, CA, Stanford University, Program on Energy and Sustainable Development. Working Paper #90**:** 38.

Holland, P., 1986, "Statistics and Caual Inference." *Journal of the American Statistical Association* **81**: 945-960.

Hoyos, D., 2010, "The state of the art of environmental valuation with discrete choice experiments." *Ecological Economics* **69**(8): 1595-1603.

IR, 2008, Bad Deal for the Planet: Why carbon offsets aren't working...and how to create a fair global climate accord. Berkeley, CA, International Rivers.

Kagel, J. H. and A. E. Roth, Eds., 2007, *The Handbook of Experimental Economics*. Princeton, NJ, Princeton University Press.

Kahneman, D., 2003, "Maps of bounded rationality: psychology for behavioral economics." *The American Economic Review* **93**(5): 1449-1475.

Kahneman, D. and A. Tversky, 1979, "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* **47**(2).

Kartha, S., M. Lazarus, et al., 2004, "Baseline recommendations for greenhouse gas mitigation projects in the electric power sector." *Energy Policy* **32**(4): 545-566.

Kartha, S., M. Lazarus, et al., 2005, "Market penetration metrics: tools for additionality assessment?" *Climate Policy* **5**: 147-165.

Khandker, S. R., G. B. Koolwal, et al., 2010, *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington, D.C., World Bank.

King, G., R. O. Keohane, et al., 1994, *Designing Social Inquiry: Scientified Inference in Qualitative Research*. Princeton, NJ, Princeton University Press.

Kollmuss, A., M. Lazarus, et al., 2010, Discounting Offsets: Issues and Options. Somerville, MA, Stockholm Environment Institute. Working paper WP-US-1005.

Kotchen, M. J., 2009, "Voluntary Provision of Public Goods for Bads: A Theory of Environmental Offsets." *The Economic Journal* **119**(537): 883-899.

Kragt, M. and J. Bennett, 2012, "Attribute Framing in Choice Experiments: How Do Attribute Level Descriptions Affect Value Estimates?" *Environmental and Resource Economics* **51**(1): 43-59.

Lindblom, C. E., 1959, "The Science of "Muddling Through"." *Public Administration Review* **19**(2): 79-88.

McFadden, D., 1999, "Rationality for Economists?" *Journal of Risk and Uncertainty* **19**(1): 73-105.

Menges, R., C. Schroeder, et al., 2005, "Altruism, Warm Glow and the Willingness-to-Donate for Green Electricity: An Artefactual Field Experiment." *Environmental and Resource Economics* **31**(4): 431-458.

Meyers, S., 1999, Additionality of Emission Reductions From Clean Development Mechanism Projects: Issues and Options for Project-Level Assessment, Lawrence Berkeley National Laboratory. LBNL-43704.

Michaelowa, A., 2008, Discounting of CERs to avoid CER import caps. Cambridge, UK, Climate Strategies.

Millard-Ball, A. and L. Ortolano, 2010, "Constructing carbon offsets: The obstacles to quantifying emission reductions." *Energy Policy* **38**: 533-546.

Monbiot, G., 2006, Selling Indulgences: The trade in carbon offsets is an excuse for business as usual. *Guardian* London. 18 October.

Morgan, M. G. and M. Henrion, 1990, *Performing Probability Assessment Uncertainty*, Cambridge University Press.

Mukerjee, M., 2009, Is a Popular Carbon-Offset Method Just a Lot of Hot Air? . *Scientific American*. New York. June.

Müller, B., 2009, Additionality in the Clean Development Mechanism: What and Why?, Oxford Institute for Energy Studies. EV 44.

Nadel, S., j. Thorne, et al., 2003, Market Transformation: Substantial Progress from a Decade of Work. Washington, DC, American Council for an Energy-Efficient Economy. A036.

OECD/IEA, 2000, Emission Baselines: Estimating the Unknown. Paris, International Energy Agency.

OQI, 2008, Ensuring Offset Quality: Integrating High Quality Greenhouse Gas Offsets Into North American Cap-and-Trade Policy. Portland, OR, Offset Qualitative Initiative.

OQI, 2009a, Assessing Offset Quality in the Clean Development Mechanism. Portland, OR, Offset Quality Initiative.

OQI, 2009b, Maintaining Carbon Market Integrity: Why Renewable Energy Certificates Are Not Offsets. Portland, OR, Offset Quality Initiative.

PDF, 2009, Letter to Head and Members of the CDM Executive Board, Subject: Chinese wind and E+/E- policy, dated 13 November 2009. London, Project Developer Forum.

PDF, 2010, Letter to Head and Members of the CDM Executive Board, Subject: EB 55 decisions on E+/E- policy guidance and Chinese Tariffs for Renewable Energy Projects, dated 27August 2010. London, Project Developer Forum.

Reeson, A. and J. Tisdell, 2010, "The Market Instinct: The Demise of Social Preferences for Self-Interest." *Environmental and Resource Economics* **47**(3): 439-453.

Rentz, H., 1998, "Joint implementation and the question of `additionality'--a proposal for a pragmatic approach to identify possible joint implementation projects." *Energy Policy* **26**(4): 275-279.

Rossi, P. H., M. W. Lipsey, et al., 2004, *Evaluation: A Systematic Approach*. Thousand Oaks, CA, Sage Publications.

Schneider, L., 2007, Is the CDM fulfilling its environmental and sustainable development objectives? An evaluation of the CDM and options for improvement. Berlin, Öko-Institut**:** 75.

Schneider, L., 2009a, "Assessing the additionality of CDM projects: practical experiences and lessons learned." *Climate Policy* **9**: 242-254.
Schneider, L., 2009b, "A Clean Development Mechanism with global atmospheric benefits for a post-2012 climate regime." *International Environmental Agreements* **9**: 95-111.
Shogren, J., G. Parkhurst, et al., 2010, "Two Cheers and a Qualm for Behavioral Environmental Economics." *Environmental and Resource Economics* **46**(2): 235-247.
Shogren, J. F. and L. O. Taylor, 2008, "On Behavioral-Environmental Economics." *Review of Environmental Economics and Policy* **2**(1): 26-44.
Shrestha, R. M. and G. R. Timilsina, 2002, "The additionality criterion for identifying clean development mechanism projects under the Kyoto Protocol." *Energy Policy* **30**(1): 73-79.
Simon, H., 1957, A Behavioral Model of Rational Choice. *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York, Wiley.
Sugiyama, T. and A. Michaelowa, 2001, "Reconciling the design of CDM with inborn paradox of additionality concept." *Climate Policy* **1**(1): 75-83.
Sutter, C. and J. Parreño, 2007, "Does the current Clean Development Mechanism (CDM) deliver its sustainable development claim? An analysis of officially registered CDM projects." *Climatic Change* **84**(1): 75-90.
Tanwar, N., 2007, "Clean development mechanism and off-grid small-scale hydropower projects: Evaluatin of additionality." *Energy Policy* **35**: 714-721.
Trexler, M. C., D. J. Broekhoff, et al., 2006, "A Statistically-Driven Approach to Offset-Based GHG Additionality Determinations: What can we learn?" *Sustainable Development Law & Policy* **VI**(2): 30-40.
van den Bergh, J. C. J. M., A. Ferrer-i-Carbonell, et al., 2000, "Alternative models of individual behaviour and implications for environmental policy." *Ecological Economics* **32**(1): 43-61.
Venkatachalam, L., 2008, "Behavioral economics for environmental policy." *Ecological Economics* **67**(4): 640-645.
Vine, E. and J. Sathaye, 1999, "The Monitoring, Evaluation, Reporting and Verification of Climate Change Projects." *Mitigation and Adaptation Strategies for Global Change* **4**(1): 43-60.
Vine, E. L. and J. A. Sathaye, 2000, "The monitoring, evaluation, reporting, verification, and certification of energy-efficiency projects." *Mitigation and Adaptation Strategies for Global Change* **5**(2): 189-216.
Wara, M. W. and D. G. Victor, 2008, A Realistic Policy on International Carbon Offsets. Stanford, CA, Program on Energy and Sustainable Development, Stanford University. Working Paper #74.
Weber, R. and C. Camerer, 2006, "“Behavioral experiments” in economics." *Experimental Economics* **9**(3): 187-192.
Williamson, O. E., 2005, "Transaction cost economics and business administration." *Scandinavian Journal of Management* **21**(1): 19-40.
WRI/WBCSD, 2005, The Greenhouse Gas Protocol: The GHG Protocol for Project Accounting. Washington, DC and Switzerland, World Resources Institute and World Business Council for Sustainable Development.