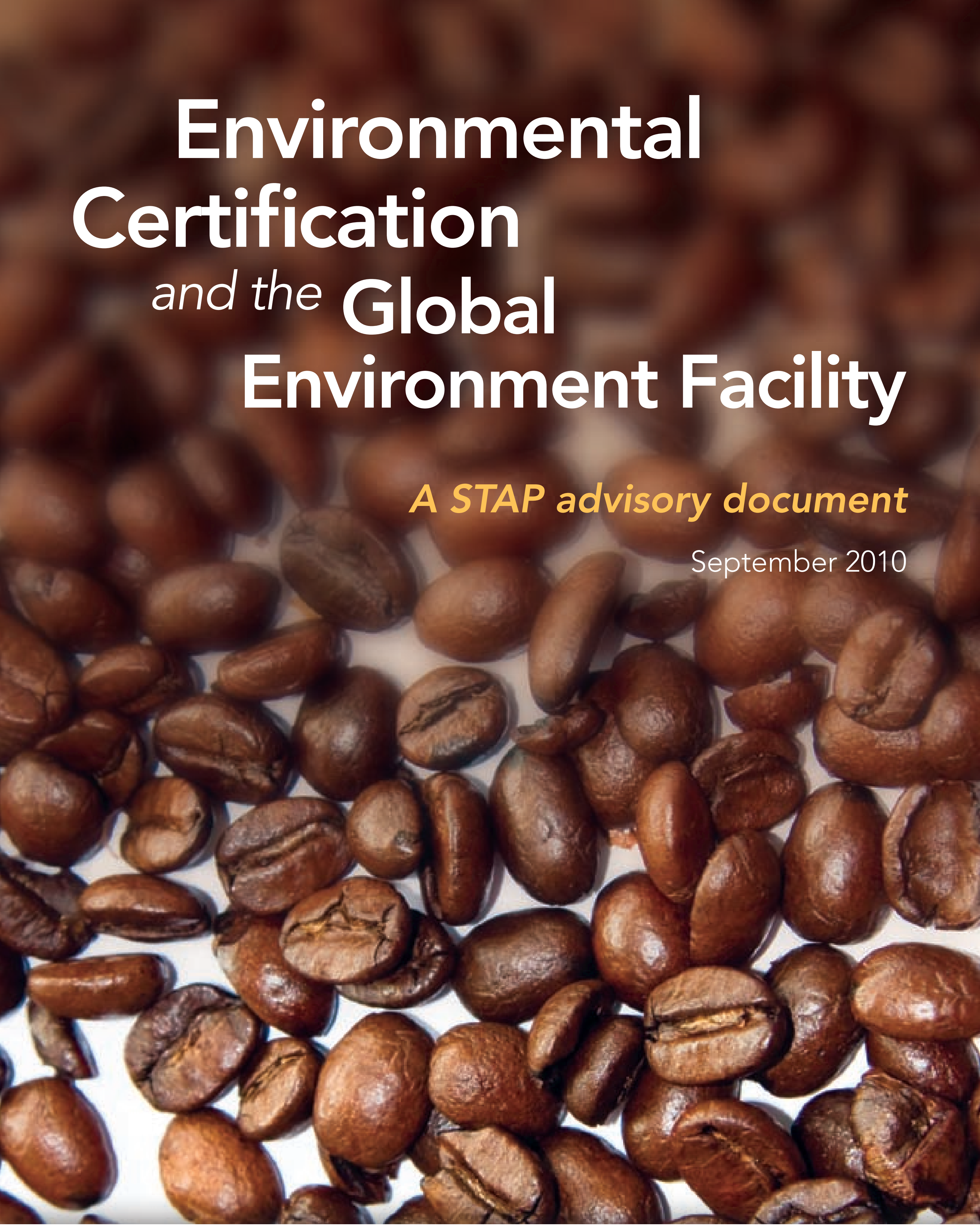
GEF/C.39/Inf.15
October 21, 2010

GEF Council Meeting
November 16-18, 2010
Washington DC

ENVIRONMENTAL CERTIFICATION AND THE GLOBAL ENVIRONMENT FACILITY: A STAP ADVISORY DOCUMENT

(Prepared by STAP)



Environmental Certification *and the* Global Environment Facility

A STAP advisory document

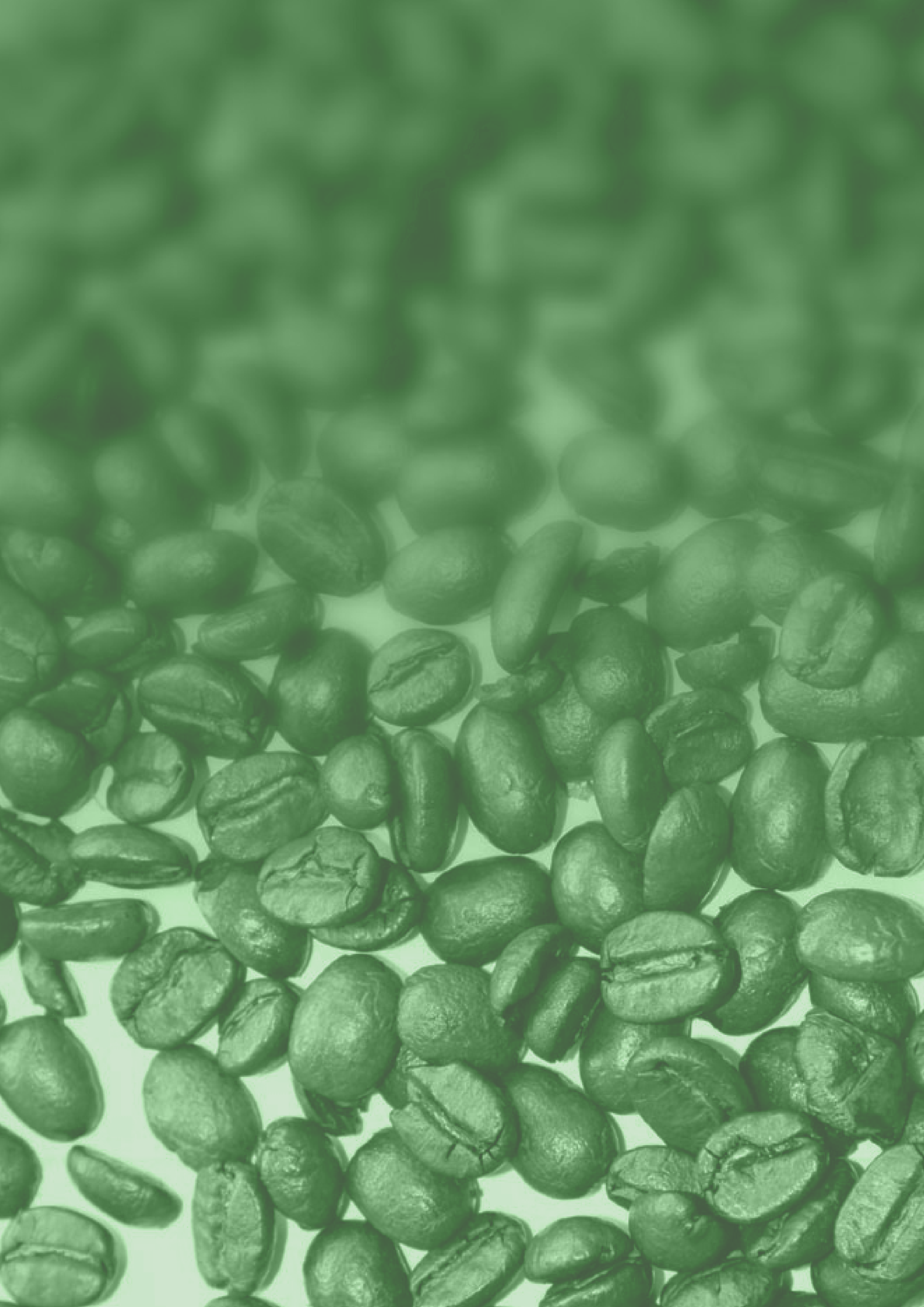
September 2010

Scientific and Technical Advisory Panel

The Scientific and Technical Advisory Panel, administered by UNEP, advises the Global Environmental Facility

gef

UNEP



Environmental Certification *and the* Global Environment Facility

A STAP advisory document

September 2010

Scientific and Technical Advisory Panel

The Scientific and Technical Advisory Panel, administered by UNEP, advises the Global Environmental Facility

gef

UNEP

Environmental certification and the Global Environment Facility:
A STAP advisory document

Prepared on behalf of the Scientific and Technical Advisory Panel (STAP) of the Global Environment Facility (GEF) by

Allen Blackman
Resources for the Future, Washington, DC, and Environment for Development Center for Central America, Turrialba, Costa Rica; blackman@rff.org; and

Jorge Rivera
George Washington University, Washington, DC, and Environment for Development Center for Central America, Turrialba, Costa Rica; jrivera@gwu.edu.

Acknowledgements
We are grateful to David Cunningham and Paul Ferraro of the Scientific and Technical Advisory Panel for their many helpful comments and suggestions and to Adriana Chacón and Jeffrey Ferris for expert research assistance. Remaining errors are our own.

About STAP
The Scientific and Technical Advisory Panel comprises six expert advisers supported by a Secretariat, which are together responsible for connecting the Global Environment Facility to the most up to date, authoritative, and globally representative science.

http://www.unep.org/stap

Foreword

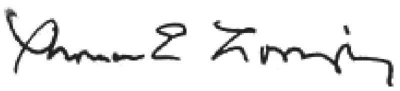
Sustainable certification ("eco-certification") initiatives certify that commercial producers adhere to predefined environmental and social welfare production standards. Such initiatives are common in GEF-funded projects aimed at mainstreaming biodiversity and ecosystem services in production landscapes and seascapes. This advisory document summarizes the evidence base for the effectiveness of certification programs in generating global, national, regional and local environmental benefits. It also summarizes evidence related to the socioeconomic impacts on participants. It was reviewed by two external reviewers, STAP panel members and STAP Secretariat staff.

The key messages and their implications for the GEF include:

1. There are four main threats to eco-certification effectiveness: (i) weak certification standards; (ii) noncompliance with certification standards; (iii) limited participation, which can stem from supply-side or demand-side factors; and (iv) adverse self-selection, whereby actors already engaged in, or intending to engage in, innovative or environmentally-friendly practices disproportionately participate in the program. The first three threats are generally recognized in GEF project designs. However, the threat of adverse self-selection, which has been shown to limit impacts in a wide range of voluntary programs, is typically ignored in project designs. **We recommend that GEF certification project proposals describe design choices to minimize these four threats and specify indicators that will permit one to evaluate the importance of threats (ii) - (iv) during the life of the project.**

2. Despite the abundance of certification programs operating worldwide, only thirty-seven studies have attempted to measure these programs' environmental or socioeconomic impacts. Of these thirty-seven studies, only fourteen make a serious attempt to elucidate the causal impact of certification by eliminating rival explanations of the observed outcomes (e.g., increased incomes) that have nothing to do with certification (e.g., national trends in economic growth). Twelve of these fourteen studies focus on the banana, cocoa or tourism sectors. Ten focus only on Fair Trade or organic certification. Importantly, only four of the fourteen studies examine environmental impacts and only one of these four detected any impact (five out of ten of the socioeconomic studies detected positive impacts). The evidence base provides, at best, weak evidence for the hypothesis that certification has positive socioeconomic or environmental impacts. **GEF agencies proposing a new or expanded eco-certification effort should acknowledge that they are proposing an innovative, but inadequately understood, intervention. Proposals should carefully explain the pathways through which their projects will generate desired environmental, and perhaps socioeconomic, impacts.**

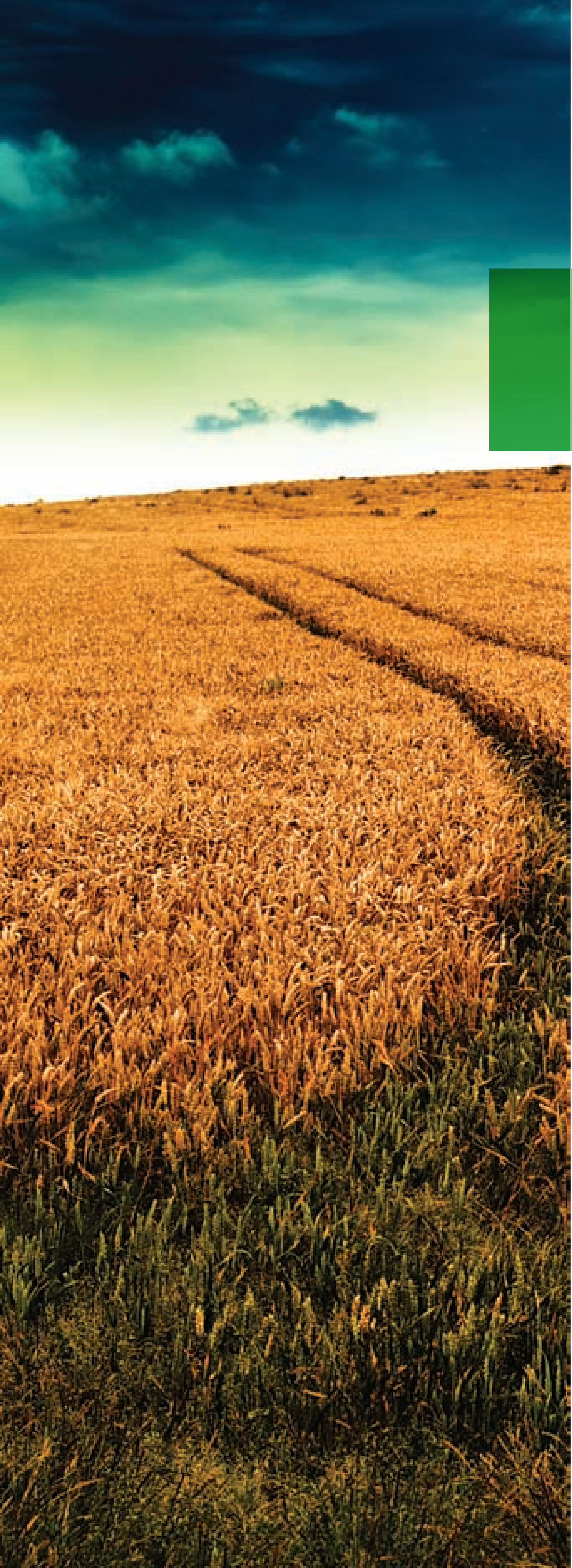
3. Financing of certification initiatives is consistent with the GEF's mandate to increase the supply of global environmental benefits. The limited evidence base does not imply that the GEF should avoid investing in certification programs, nor does it imply that past investments in certification have necessarily failed to yield returns. **However, it does imply that GEF investments in certification should be made in projects that are deliberately designed to evaluate the environmental impacts of the certification program.** Projects must include more than simple monitoring of status and trends of environmental indicators. They must be designed to permit credible inferences to be drawn about whether the program is contributing to changes in the status and trends of the indicators. **Examples of such designs are described in Section 6 of this review.** The information generated by such designs will also contribute to achieving Learning Objective Three of the GEF-5 Biodiversity Focal Area Strategy: Enhancing Impacts through Improved Understanding of the Causal Relationships between Popular Mainstreaming Approaches and Conservation Outcomes.



Thomas Lovejoy
Chair, Scientific and Technical Advisory Panel

Paul Ferraro
Panel Member for Biodiversity (2007-2009)

Table of Contents



Executive Summary

Background

Initiatives that certify that farms and firms adhere to predefined environmental and social welfare production standards are increasingly popular. According to proponents, such sustainable certification initiatives create incentives for farms and firms to improve their environmental and socioeconomic performance. In theory, they can do this by enabling consumers to differentiate among goods and services based on their environmental and social attributes. This improved information facilitates price premiums for certified products or new access to markets for such products. Premiums and access, in turn, create financial incentives for farms and firms to meet certification standards.

Yet certification programs that aim to improve commodity producers' environmental and social performance face important challenges. They must use standards stringent enough and monitoring and enforcement strict enough to ensure poorly performing producers are excluded. In addition, they must offer price premiums high enough, or access valuable enough, to offset the costs of certification and attract a significant number of applicants. Even if these two challenges are met, certification schemes can still be undermined by selection effects. Commodity producers already meeting certification standards have strong incentives to select into certification programs: they need not make additional investments in environmental management or social welfare to pass muster, and can obtain price premiums and other benefits. But certification programs that mainly attract such producers will have limited effects on producer behavior, and thus few additional environmental and social benefits.

Although a fast-growing academic literature examines sustainable certification, we still know little about how it actually affects farms' and firms' environmental and socioeconomic performance. Relatively few studies specifically aim to evaluate the impacts of certification. Among those that do, many rely on crude empirical designs that do not correct for selection effects or other sources of bias.

Objective

The objective of this report is to assess the evidence base on the environmental and socioeconomic impacts of sustainable certification of agricultural commodities, tourism operations, fish and forest products. We do this by identifying empirical studies of sustainable certification impacts, classifying them on the basis of whether they use methods likely to generate credible results, summarizing them, and considering the implications of our findings for future research and GEF funding decisions.

Methods

To identify studies, we searched digital databases, citations in relevant studies, and library catalogues. To be included, studies had to meet three criteria. They had to:

i. analyze certification of agriculture commodities, tourism enterprises, fish or forest products (other types of activities such as manufacturing were omitted);
ii. focus specifically on identifying socioeconomic and environmental impacts of certification (rather than topics such as consumer demand, the drivers of certification, and certification design recommendations); and
iii. present an *ex post* empirical analysis; in other words, an analysis of an actual experience with certification rather than an *ex ante* simulation or discussion.

We categorized the studies that met the above three criteria based on their methodological rigor. To credibly identify certification impacts, a study must construct a counterfactual outcome—an estimate of what environmental or socioeconomic outcomes for certified entities would have been had they not been certified. The impact of certification is defined as the difference between the actual outcome and the counterfactual outcome. A credible counterfactual controls for the selection problem noted above (further explanation is provided in the body of the report). Hence, we categorized studies in the evidence base as follows:

A1. studies that construct a reasonably credible counterfactual and can therefore be considered a test of the causal impact of certification; and
A2. studies that do not construct such a counterfactual.

Overview of the evidence base

We find that the evidence base is limited, comprised of just 37 studies. Of these studies:

- 14 construct a reasonably credible counterfactual and are categorized as A1;
- 18 focus on coffee, 9 on timber, 5 on bananas, 3 on tourism, 1 on fish, and 1 on a portfolio of agricultural products;
- 17 focus on environmental impacts (although of the 14 A1 studies, only 4 focus on environmental impacts); and
- a large share focus on Fair Trade certification.

Impacts

Of the 14 A1 studies in the evidence base, only six find some evidence that certification has positive impacts. One of these six studies tests for an environmental impact and five for a socioeconomic impact. Eight of the remaining 14 studies fail to find that certification had an observable impact. Hence, at best, the methodologically rigorous A1 studies provide very weak evidence for the hypothesis that 'sustainable' certification has positive socioeconomic or environmental impacts.

Implications for the GEF

In the GEF programming document[1], certification is listed as a form of private sector engagement in the International Waters and Biodiversity focal areas, and as a management tool for the Sustainable Forest Management (SFM) and Land Use, Land-Use Change and Forestry program.[2] The GEF-5 strategies for Biodiversity and SFM/LULUCF specifically refer to certification schemes as possible initiatives.[3]

Although hundreds of certification programs are operating worldwide, there are few concrete "lessons learned" for GEF partners seeking to boost impacts of new or expanded eco-certification efforts. Although quite a bit is known about certification processes (e.g.,

1 See GEF_R5_31 GEF_5 Programming Document, May 03, 2010
2 The climate change focal area also lists certification. Although the insights in this STAP document are pertinent to certification programs in energy, we did not attempt to review the evidence base in this area.
3 See GEF/R.5/Inf.21.

how to reduce the transaction costs of participation), far less is known about certification impacts and how to design programs to maximize them. This knowledge gap implies that GEF project designs cannot be justified simply by precedent. Instead, certification proponents must acknowledge that they are proposing an innovative, but poorly understood, intervention and carefully explain the pathways through which their project will generate desired environmental (and perhaps socioeconomic) impacts.

Financing of certification initiatives is consistent with the GEF's mandate to increase the supply of global environmental benefits. The limited evidence base does not imply that the GEF should avoid investing in certification programs, nor does it imply that past investments in certification have necessarily failed to yield returns. However, it does imply that the GEF, as one of the leading funders of eco-certification efforts in developing nations, should only invest in certification projects that are explicitly designed to evaluate the environmental impacts of the certification effort. **In other words, the purpose of eco-certification projects in the GEF portfolio should not simply be to attempt to generate environmental benefits at the project site, but also to catalyze the mainstreaming of biodiversity globally though the generation of credible evidence about what works and under what conditions.** This potential for catalysis is explicitly recognized in Learning Objective Three of the GEF-5 Biodiversity Focal Area Strategy: *Enhancing Impacts through Improved Understanding of the Causal Relationships between Popular Mainstreaming Approaches and Conservation Outcomes.*[4] This learning objective will be accomplished primarily through support of prospective experimental and quasi-experimental project designs.

Specific actions that the GEF can take include:

1. All GEF eco-certification projects should dedicate some of their monitoring resources to tracking the four main threats to effectiveness of certification programs: (i) weak standards; (ii) noncompliance; (iii) limited participation; and (iv) adverse self-selection.

2. All GEF eco-certification projects should contain design elements that are explicitly chosen to increase the ease with which one can infer the program's impact and, if possible, the mechanisms through which impacts are realized. See Section 6 for more details.

4 See GEF/R.5/Inf.21 for full text.

I. Introduction

Initiatives that certify that farms and firms adhere to predefined environmental and social welfare production standards are increasingly popular. For example, over 120 million hectares of forest have been certified by the Pan European Forest Certification Agency, the Forest Stewardship Council, and other organizations (Rametsteiner and Simula 2003). Some 100 certification schemes for tourism have emerged worldwide (Font 2002). And global production of organic, Fair Trade, and other types of certified coffees has grown at between 10% and 20% per year in recent years, a rate far higher than that for other types of specialty coffee (Kilian et al. 2004).

According to proponents, sustainable certification initiatives like these creates incentives for farms and firms to improve their environmental and socioeconomic performance (Giovannucci and Ponte 2005; Rice and Ward 1996). In theory, they can do this by enabling consumers to differentiate among goods and services based on their environmental and social attributes. This improved information facilitates price premiums for certified products or new access to markets. Premiums and access, in turn, create financial incentives for farms and firms to meet certification standards.

Yet certification programs that aim to improve commodity producers' environmental and social performance face important challenges. They must use standards stringent enough and monitoring and enforcement strict enough to ensure poorly performing producers are excluded. In addition, they must offer price premiums high enough or market access valuable enough to offset the costs of certification and attract a

significant number of applicants. Even if these two challenges are met, certification schemes can still be undermined by selection effects. Commodity producers already meeting certification standards have strong incentives to select into certification programs: they need not make additional investments in environmental management or social welfare to pass muster, and can obtain price premiums and other benefits. But certification programs that mainly attract such producers will have limited effects on producer behavior, and few additional environmental and social benefits.

Although a fast growing academic literature examines sustainable certification, we still know little about whether it actually affects farms' and firms' environmental and socioeconomic performance. Relatively few studies specifically aim to evaluate the impacts of certification, and many of these studies rely on crude empirical designs that do not correct for selection effects or other sources of bias.

The objective of this report is to assess the evidence base on the environmental and socioeconomic impacts of sustainable certification of agricultural commodities, tourism operations and fish and forest products. We do this by identifying empirical studies of sustainable certification impacts, classifying them on the basis of whether they use methods likely to generate credible results, summarizing them, and considering the implications of our findings for future research and GEF funding decisions.

The remainder of this report is organized as follows. The next section discusses the methods we used to collect and classify certification studies. The third section discusses the key methodological challenge in evaluating certification impacts—constructing a credible counterfactual outcome. The fourth section provides an overview of the evidence base. The fifth section describes in more detail the studies that comprise the evidence base,[5] and the last section considers the implications for the GEF.

5 An online supplementary appendix contains an annotated bibliography of the evidence base reviewed. See http://www.unep.org/stap

2. Methods

2.1. Literature search

To identify studies of commodity certification, we searched digital databases, citations in relevant studies, and library catalogues. We used the following internet search engines

- Econlit
- Google
- Google Scholar
- Science Direct
- Scirus
- Scopus.

In addition, we searched the digital libraries of the Center for Tropical Agricultural Research and Training *(Centro Agronómico Tropical de Investigación y Enseñanza*, CATIE) in Costa Rica, which houses a collection of unpublished studies of agricultural certification.

In constructing electronic searches, we cast a wide net to identify as many studies as possible. We used a variety of combinations of search terms including "certification", "ecolabel" and "label" sometimes in combination with the names of the sectors on which we focused ("bananas," "cocoa," "coffee," "fish," etc.). We searched for published and unpublished studies in English and in Spanish.

2.2. Criteria for inclusion in evidence base

Studies included in the evidence base had to meet three criteria. They had to:

i. analyze certification of agriculture commodities, tourism enterprises, and fish and forest products (other types of activities such as manufacturing were omitted);
ii. focus specifically on identifying socioeconomic and environmental impacts of certification (rather than topics such as consumer demand for certified products, the drivers of certification, and certification design recommendations); and
iii. present an *ex post* empirical analysis; in other words, an analysis of an actual experience with certification rather than an *ex ante* simulation or general discussion.

Geographical focus was not a criterion; we included studies from industrialized countries as well from developing countries.

2.3. Categorization

We grouped studies that met these three criteria into two categories:

A1. studies that construct a reasonably credible counterfactual and can therefore be considered a test of the causal impact of certification
A2. studies that do not construct a reasonably credible counterfactual.

Section 3 below discusses the reason for this distinction. Methodological issues that caused studies to be classified as "A2" instead of "A1" are detailed in Section 4 below, and in the annotated bibliography.

Studies that did not meet the three criteria for inclusion in the evidence base, but that are somewhat relevant because they shed light on certification impacts, were placed in a third category labeled "B."

2.4. Study summaries

We wrote a brief summary of each study in categories A1, A2 and B. Each summary includes nine types of information:

i. Full bibliographic information (authors, date, title, publication, etc.)
ii. Sector (e.g., bananas, coffee, etc.)
iii. Category (A1, A2, B)
iv. Rationale for categorization
v. Type of certification (e.g., Fair Trade, Forest Stewardship Council, etc.)
vi. Study area (country and region)
vii. Study years (years data collected where applicable)
viii. Method of analysis (a brief description of the methods—in particular, those used to construct a control group where applicable—and any major methodological problems)
ix. Findings about certification impacts.

Appendix A is a compendium of these summaries organized by relevance, category and sector. It is available online at http://www.unep.org/stap/.

2.5. Sectors

We found studies in categories A or B for the following sectors:

i. Bananas
ii. Coffee
iii. Fish
iv. Timber and non-timber forest products
v. Tourism.

In addition, we found studies of certification of beef, pork, biofuels, cacao and other agricultural products, which we include in a catchall "miscellaneous" category.

3. Counterfactual

To credibly identify the impacts of certification, an evaluation must construct a counterfactual outcome, which is an estimate of what environmental or socioeconomic outcomes for certified entities would have been had they not been certified. The impact of certification is defined as the difference between actual outcome and counterfactual outcome.

Unfortunately, most impact evaluations of certification programs use problematic counterfactual outcomes that likely bias their results. One common approach is to use certified entities' precertification outcome as the counterfactual outcome. The implicit assumption is that had certified entities not been certified, their outcomes, on average, would have stayed the same. Obviously, this assumption is violated when outcomes change during the study period because of contemporaneous cofounders, which are factors unrelated to certification that affect outcomes. For example, say that a study of the socioeconomic impacts of Fair Trade coffee certification uses certified growers' precertification household income as the counterfactual outcome and, therefore, measures impact as the difference between average pre-certification and post-certification household income. Furthermore, say that this difference is positive and large so that evaluator concludes that certification raised average household income. This estimate of certification impact would be biased upward—and the evaluators' finding of a causal effect would be misleading—if growers' household incomes rose after certification for reasons that had nothing to do with certification. These reasons might include increases in international prices for coffee, advantageous weather conditions or improvements in processing and marketing.

A second common approach is to use noncertified entities' outcomes as the counterfactual outcome (noncertified entities serve as a comparison group). The implicit assumption is that had certified entities not been certified, their outcomes would be the same, on average, as those of noncertified entities. This assumption is violated when entities with characteristics that affect outcomes select themselves—or are selected by certifiers—into certification, a problem known as selection bias. For example, say that a study of the impacts of organic coffee certification on soil erosion uses a measure of soil erosion on noncertified growers' farms as the counterfactual outcome and, therefore, measures impact as the difference between average soil erosion measures for certified and noncertified households. Furthermore, say that this difference is negative and large so that the evaluator concludes that certification drove reductions in soil erosion. This estimate of certification's impact would be biased upward—and the evaluators' finding of a causal effect would be misleading—if growers with lower soil erosion rates self-selected into organic certification. This might happen if a disproportionate number of growers that had already adopted soil conservation measures sought organic certification, recognizing that they would not have to invest in additional conservation measures to meet certification standards.

Three principal approaches to constructing a credible counterfactual have been used (Ferraro 2009; Greenstone and Gayer 2007; Ferraro and Pattanayak 2006; Frondel and Schmidt 2005; Stern et al. 2005). One requires *ex ante* experimental design of certification projects to facilitate unbiased impact evaluation. For certification projects, this amounts to randomly selecting entities to receive certification from among a group of qualified and interested candidates.[6] The outcome for the randomly constituted (noncertified) control group is then used as the counterfactual outcome for certified entities. This approach requires building evaluation into conservation project design. We discuss this issue in Section 7.

An alternative "quasi-experimental" approach is to use matching, which can be implemented *ex ante* or *ex post*. The idea is to match certified producers with noncertified producers that have very similar, if not identical, observable characteristics that plausibly affect outcomes, and to use outcomes for this matched control sample as the counterfactual outcome. For example, in a study of the soil erosion impacts of organic coffee certification, certified growers would be matched with noncertified growers of similar size, education, and previous history of adopting conservation practices. Measures of soil erosion for this matched control group would be used as the counterfactual. This approach depends on the dual assumptions that no unobservable characteristics of the entities in question (for example, management skill) affect both selection into the certification program and outcomes, and that all noncertified entities in the matched control sample have characteristics that make them suitable for certification. Various methods are available for matching entities when the number of observable characteristics is large.

A second quasi-experimental approach known as instrumental variables methods take advantage of known correlations between certification and "instruments"—characteristics of certified entities that plausibly affect the probability of certification but not the socioeconomic or environmental outcome of interest. These instrumental variables can be used to control for selection bias in a statistical analysis aimed at identifying the impact of certification on socioeconomic or environmental outcomes. For example, a study of environmental impacts of organic certification of coffee growers on farmer income might use distance of the farm to a certifying agency headquarters as an instrument for certification. The drawback of this approach is that credible instrumental variables are generally hard to find.[7]

As discussed in Section 5, the studies in the evidence base on the socioeconomic and environmental impacts of sustainable certifications that construct a counterfactual rely almost exclusively on quasi-experimental matching. Only two studies use instrumental variables, and none use an experimental design.

6 There are many variations on the standard randomized design. The key feature of the experimental design is to create variation in who participates that has nothing to do with the outcomes being monitored. Denying program access to some individuals is not required. See Section 6 for examples.

7 See Section 6 for a description of an approach where the project implementers create their own instrumental variables.

4. Overview of Evidence Base

We find that the evidence base is limited, comprised of just 37 studies meeting the three criteria for inclusion. Of these studies, 14 construct a reasonable counterfactual and have been categorized as "A1." Only these 14 studies can be considered a credible test of the certifications' causal impacts. Most of the studies in the evidence base focus on coffee, timber, and bananas and a disproportionate share examine Fair Trade certification. Although about half of the 37 studies in the evidence base analyze environmental impacts, only four of these are among the "A1" studies that construct a reasonable counterfactual. Among all A1 studies, just six find some evidence that certification has positive socioeconomic or environmental impacts. Hence, at best, the A1 studies provide very weak evidence for the hypothesis that 'sustainable' certification has positive socioeconomic or environmental impacts. The remainder of this section presents a more detailed overview.

4.1. By relevance category

We identified 134 studies that looked (from their titles and abstracts) as if they might meet the criteria for inclusion in the evidence base. Upon acquiring and reading these studies, 75 studies were deemed irrelevant and discarded. Table 1 provides an overview of the remaining 59 studies by relevance category (A1, A2, and B) and sector (bananas, coffee, etc.). Of these 59 studies, 37 meet the three criteria for inclusion in the evidence base listed in Section 2.2 (analyzed certification of agriculture commodities or tourism enterprises; focused specifically on impacts; and presented an *ex post* empirical analysis) and were therefore placed in our "A" category. Twenty-two studies shed some light on certification impacts, but do not focus directly on them and were therefore placed in our "B" category.

Table 1. Count of studies of 'sustainable' certification, by relevance category and sector

	A1	A2	B	Total
	Focused on impact		Not focused on impact but relevant	
	Counterfactual	No counterfactual		
Bananas	3	2	0	5
Coffee	6	12	8	26
Fish	0	1	4	5
Timber	1	8	5	14
Tourism	3	0	3	6
Miscellaneous				
Ag. products	1	0	0	1
Beef & pork	0	0	1	1
Biofuels	0	0	0	0
Cacao	0	0	1	1
Total	**14**	**23**	**22**	**59**

Of the 37 "A" studies included in the evidence base, 14 construct a reasonably credible counterfactual and were therefore categorized as "A1". The remaining 23 studies were categorized as "A2."

Of the 14 studies in the "A1" category, all attempt to identify certification impacts by comparing certified and matched noncertified entities using cross-sectional data. Only three studies in the evidence base, all classified as A2, attempt to identify certification impacts using a before-after comparison (Quispe-Guanca 2007; Ronchi 2002; and Hicks et al. 2008). No studies compare certified and noncertified entities both before and after certification (i.e., "before-after-control-impact," BACI, study design).

4.2. By sector

Of the 37 "A1" and "A2" studies in the evidence base, 18 focus on coffee, 9 on timber, 5 on bananas, 3 on tourism, 1 on fish, and 1 on a portfolio of agricultural products. Of the 14 "A1" studies that construct a reasonably credible counterfactual, 6 focus on coffee, 3 on bananas, 3 on tourism, 1 on timber, and 1 on a portfolio of agricultural products. Finally, of the 23 "A2" studies, 12 focus on coffee, 8 on timber, 2 on bananas, and 1 on fish.

4.3. By environmental focus

Table 2 provides a summary of the number of studies in the evidence base that focus on environmental impacts of certification (some of which also focus on socioeconomic impacts). Of all 37 studies "A1" and "A2" in the evidence base, 17 focus on environmental impacts. Of the 14 "A1", 4 focus on environmental impacts. As discussed in the next subsection, most of the "A1" studies examine Fair Trade certification, a standard that mainly focuses on economic, not environmental, criteria. Finally, of the 23 "A2"studies, 13 focus on environmental impacts.

4.4. By type of certification

Table 3 provides a summary of the types of certification represented in the evidence base. They include: Fair Trade (FT), Organic, Rainforest Alliance (RA), Utz Kapeh, Starbucks, Sustainable Agriculture Network (SAN), C.A.F.E., Dolphin-Safe, Forest Stewardship Council (FSC), Finnish Forest Certification System, Certificate for Sustainable Tourism, and Sustainable Slopes Program. Counting the number of studies focused on each type of certification is problematic because many studies examine more than one type. For example, several examine coffee farmers who are both FT and organic certified. That said, it is clear that a disproportionate share of the studies focus on FT. Of the 14 "A1" studies, seven examine FT. Six of these studies appear in a single edited volume on FT. Of the 23 "A2" studies, 10 focus on FT. It is also clear that a disproportionate share of timber studies examine FSC certification. Of the nine "A1" and "A2" studies of timber, six examine FSC certification.

4.5. Impacts

Table 4 provides a summary of the number of "A1" studies that find certification has an observable positive impact on farms or firms. Of the 14 A1 studies, only six find some evidence that certification

Table 2. Count of studies of 'sustainable' certification, by relevance category, sector, and environmental focus

	A1		A2		A1+A2	
	Impact Any	Environmental impact	Impact Any	Environmental impact	Iimpact Any	Environmental impact
Bananas	3	0	2	1	5	1
Coffee	6	1	12	6	18	7
Fish	0	0	1	1	1	1
Timber	1	1	8	5	9	6
Tourism	3	2	0	0	3	2
Miscellaneous						
Ag. products	1	0	0	0	1	0
Beef & pork	0	0	0	0	0	0
Biofuels	0	0	0	0	0	0
Cacao	0	0	0	0	0	0
Total	**14**	**4**	**23**	**13**	**37**	**17**

Table 3. Studies of 'sustainable' certification, by relevance category, sector, and type of certification

	A1		A2	
	No.	Type certification (no.)	No.	Type certification (no.)
Bananas	3	FT (3)	2	FT (1); RA
Coffee	6	FT (3); Organic	12	FT (9); Organic; RA; Utz; Starbucks; SAN; C.A.F.E.
Fish	0		1	Dolphin-Safe
Timber	1	FSC	8	FSC (5); RA; FFCS;
Tourism	3	CST; SSP	0	
Miscellaneous				
Ag. products	1	FT (1)	0	
Beef & pork	0		0	
Biofuels	0		0	
Cacao	0		0	
Total	**14**		**23**	

C.A.F.E. = Farmer Equity Practices; CST = Certification for Sustainable Tourism; FFCS = Finnish Forest Certification System; FSC = Forest Stewardship Council; FT = Fair Trade; RA = Rainforest Alliance; SAN = Sustainable Agriculture Network; SSP = Sustainable Slopes Program.

Table 4. Count of (A1) studies of 'sustainable' certification that construct a counterfactual by sector, and findings of observable positive impacts on firms and farms

	No.	Positive Socioeconomic Impact	Positive Socioeconomic Impact
Bananas	3	1	—
Coffee	6	2	1
Fish	0	—	—
Timber	1	—	0
Tourism	3	1	0
Miscellaneous			
Ag. products	1	1	—
Beef & pork	0	—	—
Biofuels	0	—	—
Cacao	0	—	—
Total	**14**	**5**	**1**

has positive impacts. One of these six studies tests for a environmental impact and five for a socioeconomic impact. However in two of these five studies of socioeconomic impacts (both of coffee), the authors themselves remark that these impacts are either idiosyncratic or somewhat inconsistent (see discussion in Section 5.2.1). Eight of the remaining 14 studies failed to find that certification had an observable impact. Three of these eight studies tested for environmental impacts, and the rest for socioeconomic impacts. Hence, at best, the "A1" studies provide very weak evidence for the hypothesis that 'sustainable' certification has positive socioeconomic or environmental impacts.

5. Description of Evidence Base

This section briefly reviews the 37 studies that comprise the evidence base on sustainable certification. It is divided into six sections corresponding to the economic sectors represented in the evidence base: bananas, coffee, fish, timber, tourism, and miscellaneous. Each of these sections has three parts. The first is an overview of the count and broad findings of the studies in the sector. Titled, "Causal Impacts," the second subsection presents brief one-paragraph descriptions of each of the "A1" studies that constructs a reasonably credible counterfactual. Titled "Correlations," the third subsection presents a concise discussion of the "A2" studies.

5.1. Bananas

We found five studies that attempt to identify environmental or socioeconomic impacts of banana certification: Fort and Ruben (2008), Ruben and van Schendel (2008), Zúñiga-Arias and Sáenz Segura (2008), Melo and Wolf (2007), and Ruben et al. (2008). All focus on Fair Trade (FT) certification. All but one (Melo and Wolf 2007) are collected in Ruben (2008), an edited volume on FT certification and therefore focus mainly on the impact of certification on growers' socioeconomic status, the main concern of this type of certification. Of the five studies, three—Fort and Ruben (2008), Ruben and van Schendel (2008), and Zúñiga-Arias and Sáenz Segura (2008)—attempt to construct a credible counterfactual, and therefore can be considered tests of certification's causal impact, while two—Melo and Wolf (2007) and Ruben et al. (2008)—simply report on differences in outcomes of certified and non-certified farms.

Overall, these studies do not provide compelling evidence that FT certification boosts banana farmers' socioeconomic status or environmental performance. The last two studies, which simply compare average outcomes of certified and non-certified farmers without controlling for selection effects, find that certified farmers in Ecuador have higher socioeconomic status and better environmental performance. However, two of the first three studies, which use matching techniques to control for selection effects, find that in Ghana and Costa Rica, most socioeconomic indicators were no higher for certified farms than non certified farms. Only Fort and Ruben (2008) find that certification may have an impact. It concludes that FT certification in Peru boosts farm productivity, presumably by generating on-farm investment.

5.1.1. Causal impacts

Fort and Ruben (2008) examine the impact of FT banana certification in northern Peru on farmer household socioeconomic status. They compare average household income and wealth for a treatment sample of 50 farm households that are both FT and organic certified (all affiliated with the same cooperative) and for a matched control sample of 150 farm households (all affiliated with other cooperatives), 110 of which are organic certified but not FT certified and 40 of which are neither FT or organic certified. They use propensity score matching based on nine households characteristics to construct the control group. Comparing the treatment farms to the organic control farms, the authors find that the FT farmers have higher net income and profits mainly because they have higher productivity (not because they receive higher prices for their bananas). The authors hypothesize that FT farms are more productive because of provisions that ensure FT premiums are invested rather than consumed. Comparing FT/organic farms to non-FT non-organic farms, the authors find that FT farmers again have higher incomes, but that in this case, the difference is mainly due to higher banana prices. Note that the authors' finding that certification boosts income and profits implicitly depends on the untestable assumption that the matching procedure controlled for all important factors that account for differences in income and profits in treatment versus control farms. This assumption may not hold, however, because these two samples of farms are drawn from different cooperatives.

Ruben and van Schendel (2008) analyze the impact of FT banana certification in eastern Ghana on worker household socioeconomic characteristics. They compare incomes and expenditures for a treatment sample of 50 worker households affiliated with a cooperative that is FT certified and for 50 worker households affiliated with a cooperative that is not certified. Matching is ad hoc: the authors attempted to construct a control sample with average characteristics (household size, age, highest education level, acres of land owned, and asset value) similar to the treatment sample. The authors find that FT workers receive lower total salaries and have lower total family income than non-FT workers but have to work less hours and receive more fringe benefits. Total expenditures for the two groups, and subjective assessments of job safety, job satisfaction and fairness are not significantly different.

Zúñiga-Arias and Sáenz Segura (2008) examine the impact of FT banana certification in southern Costa Rica on farmer household socioeconomic status. They compare incomes, expenditures, and profits (among other variables) for 58 farm households affiliated with a FT certified cooperative and a matched sample of 55 farm households from a non-FT certified association. They use propensity score matching based on six household characteristics to construct the control sample. They find that income, expenditures, and profits were not significantly different for FT and non-FT households. However, FT households have higher levels of wealth and invest more in education and training. Like Fort and Ruben (2008), the authors attribute this difference to collective decision making about the use of FT premiums. Regarding attitudinal variables, FT farmers have a have a more positive view of their current and future wellbeing and a stronger feeling of belonging to their community.

5.1.2. Correlations

As noted above, two studies that simply compare average outcomes of certified and non-certified banana farmers without controlling for selection effects, find that certified farmers in Ecuador have better environmental performance and higher socioeconomic status. Melo and Wolf (2007) present comparisons of two sets of certified farmers: (i) a random sample of 10 large farms that belong to a producer association certified en mass by Rainforest Alliance, and (ii) a random sample of 13 smaller farms that belong to a producer association certified en mass by FT. Their unmatched control sample is a set of 15 large farms and 9 small ones. Using Likert scale measures of environmental "risks" related land management, water quality, agrochemical management, and waste management, the authors find that certified farms have lower risk indices than non-certified farms. Ruben et al. (2008) compare productivity, income, and other farm characteristics of 57 FT certified farms belonging to a single grower association with those of 63 unmatched neighboring

noncertified farms. They find that FT farmers have higher yields, labor productivity, assets, and credit access, use more organic fertilizer and pest control, and invest more in production, packing, environmental management, and health care—results they attribute to FT rules mandating that premiums be devoted to social and environmental investments.

5.2. Coffee

Although a considerable literature examines the link between coffee certification on one hand and socioeconomic and environmental characteristics of farm households on the other, to our knowledge, only six—Arnould et al. (2009), Blackman and Naranjo (2010), Bolwig et al. (2009), Fort and Ruben (2008), Lyngbaek et al. (2001), and Sáenz Segura and Zúñiga-Arias (2008)—attempt to construct a credible counterfactual, and therefore can be considered tests of certification's causal impact. Most farm-level coffee studies simply compare average characteristics of a sample of certified and non-certified farmers.

Overall, farm-level studies of coffee certification do not provide compelling evidence that certification has positive socioeconomic or environmental impacts. Of the six studies that attempt to construct a credible counterfactual, two—Arnould et al. (2009) and Bolwig and Gibson (2009)—find that certification has significant socioeconomic benefits, and one—Blackman and Naranjo (2010) finds that certification has a significant economic impact. However, Arnould et al. (2009) finds that although certification generates a price premium, it is not consistently correlated with socioeconomic indicators, and Bolwig and Gibson (2009) argue that in their case, these socioeconomic benefits are mainly due to a design anomaly of the certification scheme (see below). The three remaining studies—Fort and Ruben (2008), Lyngbaek et al. (2001), and Sáenz Segura and Zúñiga-Arias (2008)—either find that certification has minimal socioeconomic benefits, or actually generates a net cost.

Even among studies that do not attempt to construct a credible counterfactual, many fail to find a correlation between certification and socioeconomic or environmental benefits. Although Bacon (2005), Barbosa de Lima et al. (2007), Consumers International (2005), and Millard (2006) find that certified farmers receive higher prices, or earn higher profits, or engage in fewer environmental harmful practices than (unmatched) non certified farmers. Jaffee (2008), Martínez-Sánchez (2008), and Quispe Guanca (2007) reach less optimistic conclusions. Calo and Wise (2005) and Kilian et al.

(2008) construct farm budget models suggesting that price premiums for certification are too low for certification to be profitable.

5.2.1. Causal impacts

Arnould et al. (2009) tests for impacts of FT certification on a variety of socioeconomic indicators in communities in Nicaragua, Peru and Guatemala. The authors use a multi-stage method to control for self-selection bias and confounding factors. To construct a matched control group of non FT farmers, they first chose non FT certified communities that were adjacent to the certified communities and comparable to them in terms of climate, geography, and growing conditions (including altitude, infrastructure, and distance to market). Next, they randomly chose farms in these non FT certified communities that met the landholding criteria for FT participation (1-3 hectares per adult household member). Finally, they used the pooled sample of certified and noncertified farmers in each study country to run regressions to explain various farm-level socioeconomic indicators including coffee volume sold, price obtained, educational attainment, and health. The explanatory variables in these regressions include a dummy indicating whether the farm was FT certified along with various farm and farmer characteristics. The authors find that FT certification is positively correlated with coffee volume sold and price obtained, but less consistently correlated with indicators of educational and health status.

Blackman and Naranjo (2010) use detailed agricultural census and geographic information system data on over 6,000 farms in Central Costa Rica to test for the environmental impacts of organic certification. They

compare rates of adoption of four environmentally friendly farm management practices (soil conservation measures, shade trees, windbreaks, and organic fertilizer) and three unfriendly practices (pesticides, chemical fertilizers, and herbicides) for certified farms and for a matched control group of noncertified farms. They use propensity score matching to control for the age and education of the farmer and various physical characteristics of the farm including size, coffee variety, climate, slope, aspect, and distances to population centers. They find that organic certification improves coffee growers' environmental performance. It significantly reduces chemical input use and increases the adoption of environmentally friendly management practices.

Bolwig et al. (2009) use a Heckman selection model to test for the impact of organic certification on farm income using a random sample of 112 certified and 48 noncertified farmers in eastern Uganda. They find that certification boosts net coffee revenue by 75% on average. However, they argue that this revenue effect is not principally due to price premiums offered to certified farmers. Rather it due to the an anomaly of the "contract farming" organic marketing system in their study, which requires participants to process their coffee before selling it, thereby increasing its value added.

Fort and Ruben (2008) use propensity score matching to test for the impact of FT certification on socioeconomic status in central Peru using a sample of 151 farmers from three FT cooperatives and 164 matched farmers from three non-FT cooperatives. Because some FT producers are also organic certified, the authors compared two treatment and control samples: organic FT farmers versus matched organic non-FT farmers, and non-organic FT farmers versus matched non-organic non-FT farmers. A methodological concern is that the matching does not control for important differences between the cooperatives (such as percentage of coffee sold as FT) which almost certainly affect outcomes. In comparing organic FT farmers and matched organic non-FT farmers, the study finds no significant difference in income or investment, although FT farmers have more of certain types of assets. In comparing non-organic FT farmers and non-organic non-FT farmers, the study finds FT farmer have lower income and productivity, but higher levels of some assets and investments. The authors attribute the limited benefits of FT in their study to the "deficient distribution and use" of the FT premiums.

Lyngbaek et al. (2001) use somewhat weak ad hoc matching to identify the socioeconomic impact of organic certification in Costa Rica. They selected 10 matched pairs of small-scale organic and conventional farms in five regions of Costa Rica. Matched conventional farms were located near organic farms and had similar altitude and size. The authors find that average yields on organic farms were lower than on conventional farms and that average net income (excluding fixed certification costs) were similar for both groups, mainly because of price premiums received by organic farmers. However, if certification costs were considered, net income for organic farmers was significantly lower than for conventional farmers.

Sáenz Segura and Zúñiga-Arias (2008) use propensity score matching to test for the impact of Fair Trade certification on socioeconomic status using a sample of 103 farmers from western Costa Rica. A methodological concern is that all FT certified farmers belong to one cooperative and all non-FT certified farmers belong to a second cooperative. As a result, unobserved factors correlated with cooperative membership (not FT certification) may drive the observed differences between FT and non-FT farmers. The authors find that compared to matched non-FT farmers, FT farmers have lower income, profits, and household expenditure and worse perceptions of the functioning of their cooperatives.

5.2.2. Correlations

As noted above, several studies compare certified with unmatched noncertified farmers and find that certified farmers have higher socioeconomic status and/or use more sustainable management practices. Bacon (2005) finds that in a sample of 228 Nicaraguan farmers, organic and FT certified farmers receive higher prices and believe they have more secure land tenure. However, he also finds certified farms were no more insulated from adverse economic impacts of the sharp decline in coffee prices in the late 1990s and early 2000s (the "coffee crisis") than were non-certified farmers. Barbosa de Lima et al. (2009) examine Sustainable Agriculture Network (SAN) coffee certification in Minas Gerais, Brazil. In a sample of 16 farms, half of which were SAN certified, they find that SAN certification is correlated with use of an array environmental practices, including use of less toxic agrochemicals and solid and liquid waste management. Consumers International (2005) analyze environmental and social indicators in a sample of 28 (Fair Trade, organic, Utz Kapeh and Rainforest Alliance) certified farms and 10 noncertified farms. They find certified farms generate higher revenues and use more environmental practices. Finally, Millard evaluates Starbucks and C.A.F.E. Practices certification project in Chiapas, Mexico. He finds that productivity, prices, and profits are higher for participants than nonparticipants.

Several other studies that compare certified with unmatched noncertified farmers find that certified farmers do not do any better in terms of socioeconomic status and environmental performance. Using data from Oaxaca, Mexico on 26 FT and organic certified farms and 25 unmatched noncertified farms, Jaffee (2008) finds that although the certified farms receive higher prices, they do not generate more income or wealth and do not believe they are better off than noncertified neighbors. He suggests that root causes are low premia for FT coffee and high costs of organic certification. Philpott et al. (2007) compares ecological indicators for farms belonging to 3 organic certified, 3 organic and FT certified, and 2 noncertified cooperatives in Chiapas, Mexico. No effort is made to match the three types of cooperatives. They find no differences among the farms in ecological indicators. Similarly, Martínez-Sánchez (2008) compares ecological indicators for 10 certified organic and 10 unmatched noncertified farms in northern Nicaragua. He finds that organic farms do not have significantly different shade levels, bird diversity or bird abundance. Quispe Guanca (2007) uses survey data on changes in environmental management practices before and after (organic, FT, Rainforest Alliance, Utz Kapeh, and C.A.F.E. Practices) certification for a sample of 106 certified farms in Costa Rica. He finds that although all certified farms reduced herbicide use after certification, most did not reduce other agrochemicals.

Finally, two studies use data from field surveys to construct spreadsheet farm budget models for certified and noncertified farms. Calo and Wise (2005) model the returns from organic and FT certification in Oaxaca, Mexico. They find that although FT certification is profitable, price premiums paid to organic farmers generally fail to cover the added costs associated with certification and maintenance (assuming market rates for labor). Focusing on Costa Rica, Kilian et al. (2008) find that with one exception (organic coffee sold in Europe), certification by itself does not generate significant price premiums. However, coffee quality is a prerequisite for a price premia, and certification is a signal of this quality. They also finds that although FT coffee, which establishes a price floor for certified coffee, ostensibly has a high premia, in practice it is much lower since the price floor generates excess supply, that is, not all certified FT coffee can be sold as such.

5.3. Timber

We identified nine studies purporting to evaluate environmental or socioeconomic impacts of forest and forest product certification—mostly Forest Stewardship Council (FSC) certification—by

comparing certified and noncertified entities. Only one (de Lima et al. 2008) attempts constructs a credible counterfactual to disentangle the impact of certification. The other eight either simply compare certified and noncertified entities, or include a certification dummy variable in a regression meant to explain an environmental outcome.

Overall, these studies find little evidence that certification has significant observable environmental or socioeconomic impacts. This findings echo those of a recent review of evidence on the impact of forest certification on biodiversity (van Kuijk et al. 2009). The one study that constructs a counterfactual (de Lima et al. 2008) concludes that the environmental and socioeconomic impacts from FSC certification in Brazil are small.

Eight other studies examine certified and noncertified entities with an eye toward environmental and socioeconomic impacts, but are not rigorous enough to credibly identify these impacts. That said, they do shed light on simple associations between certification on one hand and environmental and socioeconomic characteristics on the other. Regarding environmental characteristics, several of the studies suggest that compared to conventional forest operations, certified operations are more likely to adopt management practices associated with forest conservation. For example, they may be more likely to comply with mandated forest management plans, report violations of environmental law, and adopt wildlife protection, forest fire prevention, and solid waste management practices (Madrid and Chapela 2003). Nevertheless, some of the studies find that certification is not necessarily correlated with actual improved environmental and

conservation outcomes (Kukkonen et al. 2008) and does not prevent large-scale deforestation (Ebeling and Yasue 2009; Nebel et al. 2005).

Regarding socio-economic characteristics, the vast majority of studies that consider price premiums suggest that certification provides zero to negligible premiums (Madrid and Chapela 2003; Morris and Dunne 2003; Owari et al. 2006; and Nebel 2005). That said, several studies find that certification may have indirect socioeconomic benefits including improved: marketing (Ebeling and Yasue 2009; Madrid and Chapela 2003; Owari et al. 2006); quality control in timber operations (Morris and Dunne 2003); relations with regulators (Madrid and Chapela 2003); and overall management (Madrid and Chapela 2003).

5.3.1. Causal impacts

de Lima et al. (2008) examines the impacts of FSC certification in highland natural forests of the Brazilian Amazon region. They compare indicators, drawn from original survey data, of both environmental and socioeconomic impacts for four FSC certified forest associations and two matched noncertified associations. The two noncertified associations were chosen based on three criteria: use of community forest management practices; logging for wood production as the main forest management activity; and land tenure characteristics. The study concludes that the environmental and socioeconomic impacts from certification were small. The authors hypothesize that in their sample, many of the seeming benefits that certification were already being realized through community forest management.

5.3.2. Correlations

As noted above, several studies that do not construct a counterfactual find that although forest certifications may improve environmental management practices, the overall impact on promoting forest regeneration and stemming deforestation is limited. Kukkonen et al. (2008) uses regression analysis to determine whether FSC certification affects environmental management and tree regeneration on a sample of 46 forest tree fall gaps. They find that although FSC certified forest plots used more environmentally friendly practices, tree regeneration was actually lower on certified plots than on conventional ones. Ebeling and Yasue (2009) reports qualitative results about the environmental impacts of FSC certification in Ecuador and Bolivia based on semi-structured interviews with a variety of stakeholders including 13 certified and 16 noncertified timber companies and landowners. They conclude that certification is unlikely to have significant environmental impacts in the many developing countries like Ecuador that have limited governance capacity in the forestry sector. Nebel et al. (2005) uses secondary data on FSC and Smartwood certification in the eastern lowlands of Bolivia to determine what management practices forest operators actually change to get certified. The authors conclude that certification by itself has only resulted in minor improvements in forest management and has not stemmed deforestation. Finally, Thornber et al. (1999) draws on secondary data and existing literature to provide a qualitative overview of socioeconomic and environmental benefits timber certification worldwide. They conclude that the environmental benefits of certification are typically small since most adopters already have superior environmental performance.

Several other studies that do not include a counterfactual generate similarly negative findings about certification's direct socioeconomic impacts, but also make note of longer term and less concrete certification benefits. Madrid and Chapela (2003) presents a qualitative discussion of benefits of unspecified types of certification in communities in Mexico. It concludes that although the direct economic benefits are small or non-existent, indirect benefits include conferring prestige, smoothing relations with external agencies, preserving the option of future business in the event that demand for certified timber increases, and providing an external audit of forestry operations that can be used to detect management inefficiencies. Morris and Dunne (2003) presents an analysis of FSC certification in the South African furniture industry based on interviews with a variety of stakeholders including certified and (unmatched) noncertified producers. They find that although FSC certification does not provide a price premium it helps to preserve existing market opportunities and contributes to quality control (because FSC labels include the location of manufacture and harvest). Markopoulos (1998) analyzes the impact of Rainforest Alliance certification of a community forest management project in Bolivia by comparing environmental and socioeconomic indicators from before and after certification. He finds that certification is correlated with price premiums but has not boosted community incomes. Owari et al. (2006) reports results from a survey of 25 certified and 25 noncertified Finnish wood products companies focusing on the companies' perceptions of certification. They conclude that although certified companies did not receive significant price premiums and did not believe that certification helped improve their financial performance, they viewed it as important for signaling environmental responsibility and maintaining existing market share. Finally, Thornber et al. (1999) (see above) finds that FSC certification may marginalize smaller, local operations that lack the capital necessary to invest in certification.

5.4. Fish

We were able to identify only one empirical case study that focuses specifically on the environmental or socioeconomic impact of fish or shrimp certification—Hicks et al. (2008). This study purports to identify the impact of dolphin-safe certification on fishing practices among U.S.-flagged ships in the Eastern Tropical Pacific. Importantly, dolphin-safe certification became mandatory for U.S-flagged ships in 1990. To identify the environmental impact of the certification mandate, the study uses a dynamic discrete choice model that essentially compares 1990-92 fishing practices with practices from 1980-81. Hence, the latter essentially serve as a counterfactual, that is, an indication of what would have happened absent certification. Results show that U.S. flagged ships did, in fact, change their fishing practices as a result of the certification mandate, switching to methods of targeting tuna that result in killing fewer dolphins. Several issues complicate the policy implications of this study for certification schemes generally. First, it does not control for changes in the fishing fleet between 1980 and 1992 that had little to do with certification. Also, it does not control for self-selection into and out of the Eastern Tropical Pacific U.S. fleet. Finally, the dolphin-safe certification was mandatory, not voluntary. Therefore, it is more akin to a conventional regulatory standard than to a certification scheme. That said, the study suggests, perhaps not surprisingly, that mandatory certification coupled with serious monitoring of compliance boosts the impact of certification.

5.5. Tourism

We identified three studies that focus on environmental or socioeconomic impacts of certification in the tourism sector. All three—Rivera (2002), Rivera and de Leon (2004) and Rivera et al. (2006)—construct a reasonable counterfactual that controls for self selection bias, and therefore can credibly claim to have identified impacts. Overall, the studies paint a mixed picture. Rivera (2002) finds that hotel certification in Costa Rica can generate significant price premiums, and therefore presumably have an economic benefit. However, the other two studies demonstrate that ski slope certification in the United States has failed to improve environmental performance, and may even have generated environmental damage.

5.5.1 Causal impacts

Rivera (2002) examines the economic impact of Certification for Sustainable Tourism program, a Costa Rican program that certifies the environmental performance of hotels based on more than 100 criteria. The study compares pricing and sales of a sample of participating and nonparticipating hotels

using a two-stage Heckman procedure to correct for self-selection bias. A key limitation is that the sample of certified hotels is small (52 hotels). The econometric results suggest that certified hotels with high environmental performance rating show a price increase of about $20 per room per night.

Rivera and deLeon (2004) and Rivera et al. (2006) analyze the environmental impact of the Sustainable Slopes Program, a voluntary program established by the U.S. ski areas' industry association. The studies compare independent third-party environmental performance ratings of participating and nonparticipating ski areas in the western United States using a two-stage Heckman procedure to control for self-selection bias. A key limitation of these two studies is the small sample of certified ski areas (fewer than 100). Results suggest that in the first years of the Sustainable Slopes Program, noncertified ski areas had better environmental performance than certified areas, and subsequently, had equivalent but not superior levels for most environmental performance indicators. The authors attribute this finding to a lack of institutional mechanisms to prevent opportunistic behavior. That is, the program does not involve specific environmental standards, lacks third-party oversight, and does not have sanctions for poor performance.

5.6. Miscellaneous

We identified one empirical study of the impacts of certification in a sector other than the five discussed above (bananas, coffee, fish, timber, and tourism). Becchetti and Costantino (2008) analyze the socioeconomic impact of Fair Trade (FT) certification of a variety of agricultural products (including mango, guava, lemon, sorghum, maize, millet, okra, and red pepper) for producers affiliated with a FT-certified producer association in central Kenya. It explicitly controls for selection bias by estimating a system of two equations—a certification (selection) equation and an impacts (treatment) equation—for each socioeconomic indicator in question (no environmental indicators are included). The first equation regresses a certification dummy onto farm and farmer characteristics and the second regress a socioeconomic indicator onto a participation dummy along with a variable that indicates the number of years the producer has been affiliated with the FT association. The authors find that the number of years of affiliation variable is positive and significant in two of the six selection effects models: for nutritional quality and satisfaction with living conditions. They conclude that FT certification has causal impacts on these two variables.

6. Implications for the GEF

The GEF funds projects that feature sustainable certification in a range of sectors including agriculture, forestry and tourism. These projects purport to expand participation in existing certification systems or to support the development of new systems. Our review indicates that the evidence for the environmental and socioeconomic impacts of such systems is sparse.

Despite the abundance of certification programs operating worldwide, only thirty-seven studies have attempted to document environmental or socioeconomic impacts. Of these thirty-seven studies, only fourteen make serious attempts to elucidate the causal impact of certification programs by seeking to eliminate rival explanations of the observed outcomes that have nothing to do with the certification program. Importantly, only four of the fourteen studies examine environmental impacts and only one detected any impacts (five out of ten of the socioeconomic studies detected positive impacts). Moreover, the extent of insights provided by these fourteen studies is limited. Twelve focus on banana, cocoa or tourism sectors. Ten focus only on Fair Trade or organic certification.

Thus there are few concrete "lessons learned" for GEF partners seeking to boost impacts of new or expanded eco-certification efforts. Although quite a bit is known about certification processes (e.g., how to reduce the transaction costs of participation), far less is known about certification impacts and and how to design programs to maximize them. This knowledge gap implies that GEF project designs cannot be justified simply by precedent. Instead, certification proponents must acknowledge that they are proposing an innovative, but poorly understood, intervention and carefully explain the pathways through which

their project will generate desired environmental (and perhaps socioeconomic) impacts.

Financing of certification initiatives is consistent with the GEF's mandate to increase the supply of global environmental benefits. The limited evidence base does not imply that the GEF should avoid investing in certification programs, nor does it imply that past investments in certification have necessarily failed to yield returns. However, it does imply that the GEF, as one of the leading funders of eco-certification efforts in developing nations, should only invest in certification projects that are explicitly designed to evaluate the environmental impacts of the certification effort. In other words, the purpose of eco-certification projects in the GEF portfolio should not simply be to attempt to generate environmental benefits at the project site, but also to catalyze the mainstreaming of biodiversity globally though the generation of credible evidence about what works and under what conditions. This potential for catalysis is explicitly recognized in Learning Objective Three of the GEF-5 Biodiversity Focal Area Strategy: *Enhancing Impacts through Improved Understanding of the Causal Relationships between Popular Mainstreaming Approaches and Conservation Outcomes*.[8]

6.1. What actions can the GEF take to build the evidence base?

First, all GEF eco-certification projects should dedicate some of their monitoring resources to monitoring threats to effectiveness. There are four main threats: (i) weak certification standards; (ii) noncompliance with certification standards; (iii) limited participation, which can stem from supply-side or demand-side factors; and (iv) adverse self-selection, whereby actors already engaged in, or intending to engage in, innovative or environmentally-friendly practices disproportionately participate in the program. Every GEF certification project proposal should describe design choices aimed at minimizing these four threats, and measurable indicators of these threats over the life of the project. Such indicators, particularly of adverse self-selection, can help provide indirect evidence of program effectiveness. For example, indicators demonstrating that coffee farms enrolling in a shade-coffee certification program have the same characteristics as farms switching to environmentally unfriendly sun-coffee systems constitute indirect evidence that the program is having a positive environmental outcome.

Second, all GEF eco-certification projects should contain design elements that are deliberately chosen to increase the ease with which one can measure the program's impact and, if possible, identify the pathways through which project generates these impacts. Learning Objective Three of the GEF-5 Biodiversity Focal Area Strategy echoes this statement:

> *"As a leader in supporting innovative incentive-based and information-based mainstreaming approaches, the GEF has observed an increase in the number of funded projects using certification, PES and ecosystem service valuation. Thus, the GEF has an opportunity to contribute the evidence base of these approaches by supporting work to answer the following question, "How do certification, PES and transfers of information about the distribution and values of ecosystem services affect conservation and sustainable use outcomes, and in what circumstances are they likely to be most effective?" This learning objective will be accomplished primarily through support of prospective experimental and quasi-experimental project designs."*

In other words, the effort to evaluate program effectiveness should not be an *ex post* exercise. Rather, it should be planned alongside the certification effort itself and should be built into the project design. A review of a sample of eight Project Identification Forms with eco-certification components suggests that projects are not taking full advantage of the opportunities to expand the evidence base through deliberate project design. Several recent studies discuss design principles for environmental project evaluation (Ferraro 2009; Greenstone and Gayer 2007; Ferraro and Pattanayak 2006; Frondel and Schmidt 2005; Stern et al. 2005). In what follows, we briefly discuss typical recommendations as they apply to sustainable certification programs

The first, and most important, step in an *ex ante* evaluation design is to identify and collect data on the key factors that affect the outcomes to be measured. These factors may mimic or mask program effectiveness. For example, in the case of a certification program aimed at preserving and expanding tree cover on coffee farms, administrators would want to identify and collect data on the factors that affect growers' decisions about tree cover, including the availability and cost of sun coffee technologies and changing global market conditions that affect growers' ability to afford these technologies. This step would help evaluators isolate the effect of the certification program on shade cover by eliminating rival explanations for such changes.

8 See GEF/R.5/Inf.21 for full text.

Second, certification programs should collect outcome data for both participants and nonparticipants, ideally from before and after certification is assigned. Collecting *ex post* data from participants is generally straightforward and low cost. More difficult—but critically important—is collecting data from nonparticipants and *ex ante* (baseline) data from both participants and nonparticipants. Collecting data from a comparison group of nonparticipants makes it easier to eliminate rival explanations for changes in impact indicators, such as movements in commodity or input prices affecting all producers.

Third, projects should implemented in a way that enables evaluators to select a valid comparison group. By "valid" we mean very similar to the participants in terms of characteristics that affect the impact indicators (e.g., farm productivity, farmer education, access to markets and extension agents, entrepreneurial talent, etc.). Only such a comparison group will generate a credible counterfactual—that is, a good estimate of what participants' outcome would have been absent certification. One way to construct such a comparison group is to identify a sample of farms or firms interested in, and eligible for, certification and to use a lottery to randomly select a subsample for certification.[9] Those not selected are assigned to the comparison group. Random assignment makes it more likely that the nonparticipant group will be similar to the participant group. It is often most practical in projects that have more eligible applicants than can be certified in the first few years of operation. Since some applicants must be put off, administrators can randomly choose which are enrolled immediately and which are enrolled later.

A slightly more technical strategy for implementing certification projects in a way that facilitates a valid counterfactual is to permit any and all farms or firms to obtain certification, but to only deliberately encourage a randomly selected subset to seek certification. In practical terms, this amounts to varying where and when the project is promoted (in a way that is unrelated to the monitored outcomes) rather than spending the program's marketing budget uniformly across time and space. This random variation enables evaluators to construct an instrumental variable that can be used to isolate the impact of the program (Duflo et al. 2006).

Similarly, one can set eligibility rules in ways that create a group of nonparticipants that are similar to the participants, a technique called "discontinuity design." For example, pilot programs often determine eligibility based on location or scores that are a function of observable characteristics. If the precise spatial boundaries or threshold scores that determine eligibility are chosen based on factors that have nothing to do with the monitored outcomes (i.e., "as if random"), the participating and nonparticipating entities close to the boundary or threshold are likely to be similar in terms of characteristics that affect outcomes.

Techniques for creating random variation in program participation unrelated to monitored outcomes may face resistance from program administrators, patrons, and beneficiaries who are unfamiliar their purpose and value. If such resistance, or other logistical issues, prevent the use of these strategies, administrators can at least use field knowledge to select comparison nonparticipants and collect data from both participants and comparison nonparticipants on the most important characteristics that affect outcomes. These data can be used to ensure that comparison nonparticipants are similar to participants in terms of these characteristics and, if they are not, to adjust for the differences statistically. This approach, which is common in the literature cited in the body of this report, is less certain to generate a valid counterfactual than those discussed above, but is far better than many alternatives including simply comparing outcomes of participants before and after certification, or comparing outcomes of participants and unmatched nonparticipants.

Fourth, after a control group is constructed and where practical, a second layer of variation can be introduced to create additional knowledge about certification drivers and impacts. This could be accomplished by, for example: randomly assigning different types of certification across applicants to gauge their relative impacts (e.g., Rainforest Alliance and Bird Friendly for coffee producers); randomly varying the amount and type of certification subsidies (financial and technical) provided to producers to gauge the effectiveness of these subsidies; and randomly varying certification requirements or auditing systems in minor ways across selected applicants.[10]

The kinds of innovative program designs described in this section are more likely to be incorporated into GEF projects when project personnel are aware of the essential principles of impact evaluation and are encouraged to collaborate with third-party experts to create appropriate project designs. They are also more likely if the GEF explicitly encourages these designs and provides incentives for their application. Incentives are justified because of the learning spillovers these project designs provide to the entire GEF portfolio. These spillovers are a global public good.

9 More complicated designs using stratification and pre-matching or randomization at the unit of groups, like villages or cooperatives, are also feasible; see Duflo et al. (2006).

10 S*uch designs can be done even when there is no non-participant control group. The design will allow the implementer to compare impacts of one approach to certification to another approach. The only downside is that their impacts are no different, one cannot be sure if they are both equally effective or equally* ineffective.

REFERENCES

Bibliographic information about the certification studies reviewed in this document is included in an annotated bibliography available at http://www.unep.org/stap.

Darnall, N. and S. Sides. 2008. Assessing the performance of voluntary environmental programs: Does certification matter? *Policy Studies Journal* 36: 95-117.

Duflo, E., R. Glennerster and M. Kremer. 2006. Using randomization in development economics research: A toolkit. MIT Department of Economics Working Paper No. 06-36, Cambridge, MA.

Ferraro, P. 2009. Counterfactual thinking and impact evaluation in environmental policy. In M. Birnbaum and P. Mickwitz (eds.) *Environmental Program and Policy Evaluation: New Directions for Evaluation 122*, 75–84.

Ferraro, P. and S. Pattanayak. 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology, 4*(4), 482–488.

Frondel, M., and C. Schmidt. 2005. Evaluating environmental programs: The perspective of modern evaluation research. *Ecological Economics 55*(4): 515–526.

Giovannucci, D. and S. Ponte. 2005. Standards as a new form of social contract? Sustainability initiatives in the coffee industry. *Food Policy* 30: 284–301.

Greenstone, M., and T. Gayer. 2007. *Quasi-experimental and experimental approaches to environmental economics. http://papers.ssrn.com/sol3/papers.cfm?abstract_id_1001330*

King, A. and M. Lenox. 2000. Industry self-regulation without sanctions: The chemical industry's Responsible Care program. *Academy of Management Journal* 43(4): 698-716.

Rice, R. and J. Ward. 1996. Coffee, conservation and commerce in the western hemisphere: How individuals and institutions can promote ecologically sound farming and forest management in northern Latin America. Smithsonian Migratory Bird Center: Washington, DC.

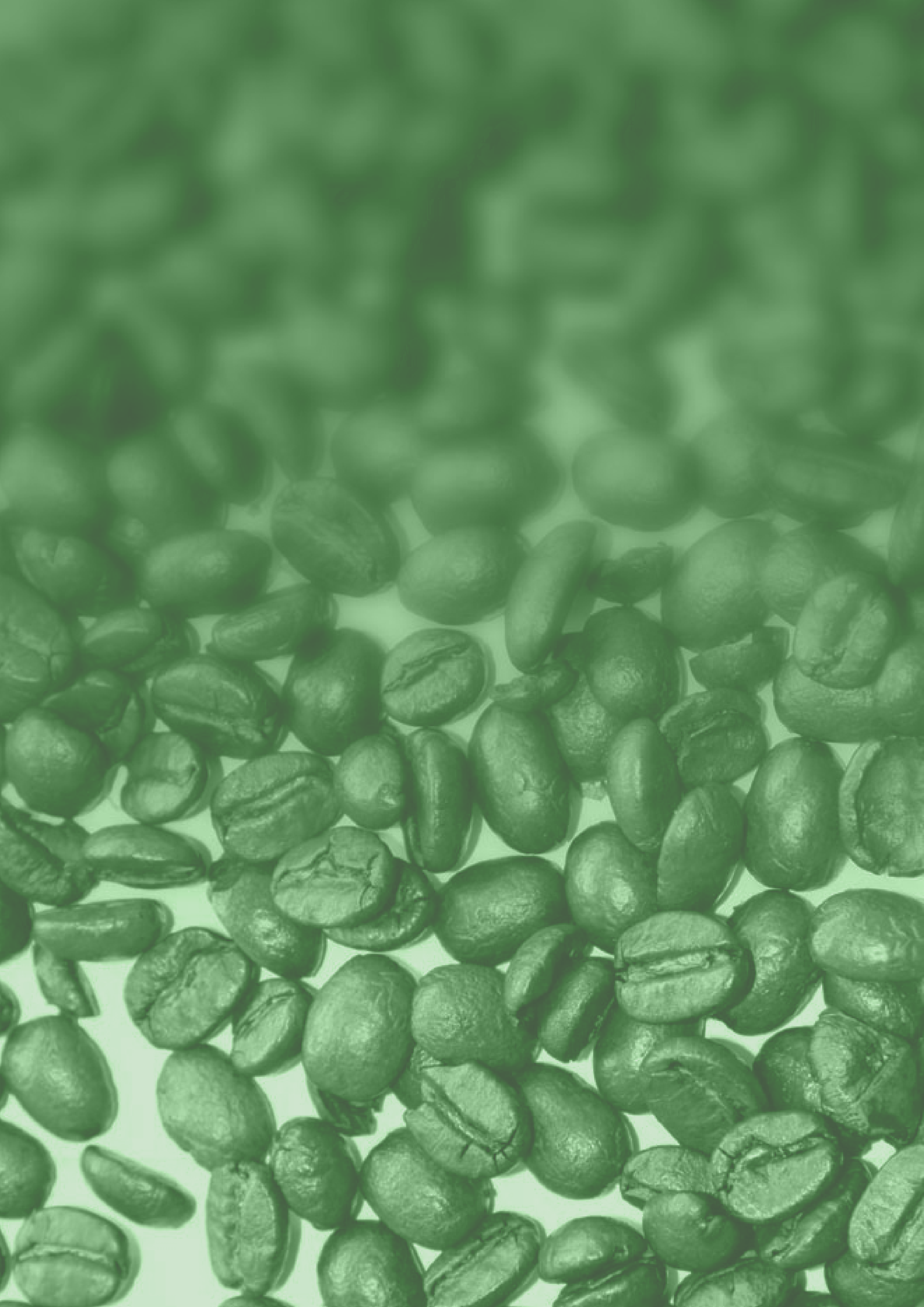
Stem, C., R. Margoluis, N. Salfasky and M. Brown. 2005. Monitoring and evaluation in conservation: A review of trends and approaches. *Conservation Biology 19*(2): 295–309.

van Kuijk, M., J. Putz and R. Zagt. 2009. Effects of forest certification on biodiveristy. Tropenbos International: Wageningen, Holland.

Appendix A.

Environmental and Socioeconomic Impacts of 'Sustainable' Certification: An Annotated Bibliography

Available online at **http://www.unep.org/stap**



Scientific and Technical Advisory Panel

The Scientific and Technical Advisory Panel, administered by UNEP, advises the Global Environmental Facility

gef

UNEP

www.unep.org/stap