WORKING PAPER 5

Impact Evaluation

BY MUHAMMAD IBNU, S.P., M.M., M.Sc., Ph.D

FAKULTAS PERTANIAN
UNIVERSITAS LAMPUNG
2018

Impact evaluation

Definition

Interventions (e.g., certifications, standards, partnerships, governmental policies and other programs) have particular aims and beneficiaries. However, it will not be clear whether such interventions really work and give intended impact to the beneficiaries, unless their impacts are evaluated. Impact evaluation, therefore, is obviously needed “To help policy makers decide whether programs are generating intended effects; to promote accountability in the allocation of resources across public programs; and to fill gaps in understanding what works, what does not, and how measured changes in well-being are attributable to a particular project or policy intervention” (Khander et al., 2010). Consequently, impact evaluation should provide an appropriate methodology and a framework to comprehend whether the recipients actually benefit from the programs—and not from other reasons or confounding factors that also influence the outcome (Ferraro, 2009).

Impact evaluation has taken different definitions for the last twenty years; as a result, methodological development - on how to measure it – also continue following these changes (White, 2006). Some organizations, such as the International Initiative for Impact Evaluation (3ie), the World Bank, the US Environmental Protection Agency (EPA), and The Global Social Venture Competition (GSVC), as well as the Organization for Economic Co-operation and Development (OECD), have provided impact evaluation with their own definitions.

The International Initiative for Impact Evaluation (3ie) (2008) defines rigorous Impact Evaluations as:

> “Analyses that measure the net change in outcomes for a particular group of people that can be attributed to a specific program using the best methodology available, feasible and appropriate to the evaluation question that is being investigated and to the specific context.”

According to the World Bank’s DIME Initiative:

> “Impact evaluations compare the outcomes of a program against a counterfactual that shows what would have happened to beneficiaries without the program. Unlike other forms of evaluation, they permit the attribution of observed changes in outcomes to the program being evaluated by following experimental and quasi-experimental designs.”

Similarly, according to the US Environmental Protection Agency (2013) impact evaluation is:

> “A type of evaluation that determines what difference the program has made. This is achieved by comparing the observed outcomes with an estimate of what would have happened in the absence of a particular program.”

The Global Social Venture Competition (GSVC) defines impact evaluation as:

> The assessment of "the portion of the total outcome that happened as a result of the company's activity, above and beyond what would have happened anyway or the next best alternative solution. In social science, one needs what is called a "counterfactual" to compare to the experimental state in order to discern the effect of the dependent variable from among all other factors that could be causing a change".

According to the Organization for Economic Co-operation and Development (OECD) (2012) Impact evaluation is:

> "An assessment of how the intervention being evaluated affects outcomes, whether these effects are intended or unintended. The proper analysis of impact requires a counterfactual of what those outcomes would have been in the absence of the intervention."

Based on numerous definitions that exist in the last two decades, White (2006) summarizes the most common definitions of impact evaluation:

- "An evaluation which looks at the impact of an intervention on final welfare outcomes, rather than only at project outputs, or a *process evaluation* which focuses on implementation;"
- "An evaluation concerned with establishing the counterfactual, i.e. the difference the project made (how indicators behaved with the project compared to how they would have been without it);"
- "An evaluation carried out some time (five to ten years) after the intervention has been completed so as to allow time for impact to appear;"
- "An evaluation considering all interventions within a given sector or geographical area."

Counterfactual Problems

One of the most common meanings of impact evaluation define impact as the gap between actual outcome and counterfactual outcome (White, 2006; Blackman and Rivera, 2010; Ruben et al. 2010; Ruben and Zuniga 2010; Ruben and Fort, 2011). Thus, following this definition, impact evaluation is designed to answer the question: how would outcomes, such as farmers' well-being, have changed if the interventions had not been undertaken? This query encompasses counterfactual analysis, which is "a comparison between what actually happened and what would have happened in the absence of the intervention" (White, 2006). The principle of counterfactual examination is the exclusion of possible competing explanations of the observed outcomes (Ferraro, 2009).

However, the main difficulty in impact evaluation is that the counterfactual cannot be directly observed (Khandar et al. 2010). Therefore, the challenge of an impact assessment is to establish a "convincing" and rational comparison group for participants (Khandar et al. 2010;

White, 2006; Blackman and Rivera, 2010; Ruben et al. 2010; Ruben and Zuniga 2010; Ruben and Fort, 2011). Ideally, researchers would like to compare the same household when it is with and without an intervention. In other words, if a treatment is given to a household, the impact of the treatment can be measured by evaluating the same household when it is free of treatment. However, a household, at the same time, cannot be in both the treated and the control group. A household cannot have "two simultaneous existences" (Khandar et al. 2010).

There are some approaches to address the problems of the counterfactual. One common approach is "with-and-without comparison" (Khandar et al. 2010). In this approach, nonparticipant's outcome is used as counterfactual outcome. This means that the nonparticipants serve as a control group. This method has an implicit assumption that if certified farmers had not been certified, their outcomes would be the same, on average, as those of noncertified farmers (Blackman and Rivera, 2010). However, this approach can be deceptive and just construct counterfeit counterfactual; therefore, it will lead to a selection bias. Selection bias is a problem that commonly occurs when comparing two identities. This bias can happens when entities select themselves or chosen by certifiers into certification programs (Khandar et al., 2010; Blackman and Rivera, 2010).

Blackman and Rivera (2010) give an example say that a study which evaluates the impacts of organic coffee certification on soil erosion. In this study, soil erosion on noncertified growers' farms serve as the counterfactual outcome. The impact, therefore, is measured as "the difference between average soil erosion measures for certified and noncertified households" (Blackman and Rivera, 2010). However, there is a chance that some coffee growers have adopted soil conservation before they involve in certification. They voluntarily join the certification recognizing that they would not have to invest in extra conservation requirement in order to fulfill the certification standards. Therefore, say the result shows the certification reduces soil erosion; the findings would be misleading because farmers already practiced soil conservation prior to the certification.

Another common approach to solve the problem of the counterfactual is to use "before-and-after comparison" (Frondel and Schmidt, 2005; Blackman and Rivera, 2010). This approach is called "reflexive method" of impact where, for example, farmers' outcome before certification serves as control outcome. While this approach can be very useful in evaluation national policies in which the whole populations participate and there is no need for a control group, this method may be problematic because some other external effects probably influence the impact of certification during the process of implementation. For example, say a study of socioeconomic impact of Fair Trade certification on coffee farmers. This study evaluates the impact based on the difference of household income on average between pre-certification and post-certification. This research uses certified farmers' household income before certification as the counterfactual. In addition, say that household income post-certification is higher than that of pre-certification, so the evaluator may conclude that certification raise average farmers' household income. However, there is also a chance that certification-unrelated factors influence the rise of household income post-certification. These factors may include an increase in coffee international prices, an improvement in handling, trading system and cooperative's role, good weather conditions, and others.

Some scholars argue that there are two primary approaches for constructing a more reliable counterfactual (Ferraro 2009; Greenstone and Gayer 2007; Ferraro and Pattanayak 2006; Frondel and Schmid, 2005). The first approach is an experimental design. In an experiment, variation is induced by "controlling how data are collected" (Ferraro, 2009). The most common way to collect data in this way is through randomization, especially for the program that has one phase and limited treatment variations (Ferraro, 2009). The most important is, in this approach, the probability of each of population member of being selected as participant is the same. In addition, the participants (i.e., the treatment group) and the non-participant (i.e., the control group) statistically should be identical, except the experience with the treatment. Hence, any differences of outcome between the two groups can be attributed to the treatment (Greenstone and Gayer 2007), and in an ideal experiment these differences may yield an unbiased evaluation of the true program impact (Frondel and Schmidt, 2005). Moreover, randomization has been used in a development setting for reasons that its cost is not higher than any other surveyed- based impact study; therefore, it may deal with budget constraint. In addition, there is no necessity to take the whole population as participants; samples, in most cases, are sufficient to represent the eligible population (White, 2006).

Although randomized experiments have a potential advantage in reducing selection bias through the randomization (Khandar et al., 2010), it also has limits to apply in the setting of development evaluation (White, 2006; Ferraro, 2009). The first limitation is the lack of opportunity to perform randomization. This is because most of evaluation schemes are ex-post so that the evaluator chances to control the collected data by randomization has passed. The second drawback is randomization perhaps has great difficulties in dealing with complex interventions. In fact, many development programs are so complex in which require multiple comparisons to address their heterogeneity. The third weakness is the evaluators retain control in the experiment. In some occasions, this leads to the classic problem of selection bias. Some potential participants may not want to participate so that researchers do some intervention for the inclusion of the participants such as lobbying them or giving a kind of incentives (i.e., present or money). Consequently, the purity of the randomization design is contaminated.

Random assignment, as critical element of experimental design, is not a panacea to all selection-bias-related problems (Bawden and Sonenstein, 1992). Many times true experimental designs are not feasible for political, financial, legal, practical, or ethical reasons" (Ferraro, 2009). For example, in health sector, it is unlikely and unethically to do experimental design for treating children with malaria, while at the same time do nothing to other children for a reason of establishing control group. Moreover, Bawden and Sonenstein (1992) argue that another limitation of experimental designs is the construction of "laboratory condition" by which make evaluation results are difficult to generalize.

The second approach, as an experimental design alternative, is to use a quasi-experimental design. This approach can be used to overcome experimental design limitations in revealing causal correlation among factors for impact evaluation (Greenstone and Gayer, 2007). In this approach, a "truly comparable" non-participant is compared to participants. This appropriate comparison provides a balance for all measurable factors, just as the ideal experiment could (Frondel and Schmidt, 2005). The approach focus on correcting improper comparison between different types of groups (i.e., participants and non-participants) and it may also reduce selection bias (Ruben and Fort, 2011; Ruben and Zuniga, 2010; Greenstone and Gayer, 2007). At least there are three ways to conduct such a fair comparison ((Bawden and Sonenstein, 1992). Firstly, comparable non-participants are randomly selected from community. Secondly, using database, individuals who possess similar characteristics with participants are selected. Finally, individuals chosen from other match comparison site but have similar characteristics with participants and eligible to participate in the program.

In the case of coffee certification, certified farmers are matched with non-certified farmers. Balance sample of both groups should be taken in order to able to provide net effects of certification on farmers (e.g., on production, revenue and spending, capital and savings, and attitude and perception) (Ruben et al. 2010; Ruben and Zuniga 2010; Ruben and Fort, 2011). The non-certified farmers serve as the control group in which the outcome of this group serves as the counterfactual outcome. The two types of farmers necessarily have very close similarity in term of observable characteristics (e.g., education, land size, certification history and so forth). Thus, both types of farmers are actually eligible for certification. However, in this design, evaluators must be careful that there is no hidden or unobservable characteristic that may influence the certification and its outcome. For example, knowledge and management skill for both groups should also be similar. This is unlikely to compare inexperience farmers with those having much experience in coffee cultivation. However, in addition to matching, quasi-experimental methods can apply differencing, instrumental variables and the pipeline approach for constructing credible counterfactuals. These quasi-experimental approaches, most of them, usually rely on the quantitative methodology of multivariate regression analysis (White, 2006).

The Threats of Validity

The same as all impact evaluation design, experimental and quasi-experimental design are subject to critiques whether they can represent valid approaches (Ferraro, 2009; Greenstone and Gayer, 2007). The critiques involve the issues of internal validity, external validity and construct validity. Internal validity refers to whether "it is possible to validly draw the inference that the difference in the dependent variables is due to the explanatory variable of interest"; External validity refers to whether "an experiment's or quasi-experiment's results can be generalized to other contexts"; and construct validity refers to whether "the researcher correctly understands the nature of the treatment" (Greenstone and Gayer, 2007).

Confounding and selection bias are the common threats to internal validity. Confounding refers to spurious relationships in which external variables, other than the observed parameters, influence the outcome of the programs (Ferraro, 2009). Selection bias primarily occurs when participants are non-randomly selected from the eligible population, and the standards defining selection are having correlation with outcomes. This bias is also including self-selection and endogenous program selection bias. Selection bias occurs when participants who already possess some characteristics that match to a program, voluntarily participate or intentionally selected to take part in the program. Endogenous program selection bias happens when some participants are chosen to participate in the intervention because they are expected to obtain more benefit from it (Khandar et al, 2010; Ferraro 2009; Greenstone and Gayer 2007).

The threats of external validity can be grouped into two general classes: (1) those in a relation with population (What population can be anticipated to act in the same manner as did the sample experimental subjects?), and (2) those related to the experimental environment (in what condition - settings, dependent variables, treatments and experiments – can the similar results be projected?). The aforementioned classes connect to two categories of external validity: population validity and ecological validity (Bracht and Glass, 1968). The threats of population validity can be attributed to the limitation of population taken as samples and biases in communities such gender, racial and cultural bias. In addition, Hawthorne effects (i.e., the awareness of participants to participate in an experiment), Rosenthal effects (i.e., the awareness of participants about expectations upon them), specific situation (time, location, scope, extent of measurements and etcetera), pre-test and post-test effects can contribute to ecological validity (Bracht and Glass, 1968; Gorard, 2003; Murnane, R. J., and J. B. Willet. 2011). Therefore, general suggestions for dealing with this validity are to establish a research methodology that (1) able to evaluate correlation between measurements and their representable concepts, and (2) to recognize and amend the measurement errors while "testing non-observational propositions" (Calder, et al., 1982).

Contamination (or contagion) potentially gives a threat to construct validity (White, 2006). Contamination may come from two causes. The first is contagion as attributed to spillover effects. These effects often happen when treatment groups and controls group are geographically very close or neighboring. Sometimes interventions given to the participants also reach the non-participants. For example, information regarding a good fertilizer delivered to certified farmers, by word of mouth, is received by non-certified farmers as well. Hence, there is a dilemma between ensuring similarity (of the treated groups and the comparison groups) and avoiding contamination; distant of both groups, therefore, also need to consider. The second source of contamination is the overlapping influences of other interventions in addition to the one being studied. For example, an individual or a group of coffee farmers probably subject to both Fair Trade and Utz Kapeh certification at the same time. Hence, the impact of a single intervention to a certain extent will be more difficult to evaluate, and evaluators need to be very cautious about this.

Example of Impact Evaluation Studies

The following studies are example of studies that utilized a matched control group for establishing a more credible counterfactual.

1. Ruben et al. (2009) evaluated the direct and indirect impact of Fair Trade on farmers' life. They conducted the research in Peru and Costa Rica and surveyed two different commodities of producers – coffee and banana farmers. The authors evaluate the influence of Fair Trade at the level of producers' household by comparing the changes of several characteristics (i.e., revenue, capital and investments). They found that despite giving benefit to farmers and strengthening their organizations (i.e., indirect effect), Fair Trade only give "fairly modest" net revenue to farmers.

   Ruben et al. (2009) compare the effect of Fair Trade on farmers involved in Fair Trade certification and on those followed conventional practice. They use the non-certified farmers as a control group. This control group is matched for growers' land use, household revenue, expenses, credit usage, and risk attitudes conditions. In addition, Ruben et al. (2009) also used Propensity Score Matching (PSM) techniques for addressing biases frequently happened in selection; thus, providing a control group that has relevant characteristics resemble the treatment group.

2. Ruben and Zuniga (2010) evaluate the impact of various types of standards on farmers' wellbeing, and the role of these standards in value chains upgrading. By surveying 315 farmers, who cultivate coffee in Northern Nicaragua, the authors compare the effect of Fair Trade, Rainforest Alliance and Café' Practices labels in coffee sector. They find that the farmers under Fair Trade received higher practices than independent farmers, but private labels nevertheless better than Fair Trade in terms of productivity and quality. In addition, according to Ruben and Zuniga (2010), Fair trade can be supportive for "initial market corporation" whereas private labels give more encouragement in quality improvement.

   Ruben and Zuniga (2010) provide two categories of control groups for comparison the effect of Fair Trade and other different certification schemes (i.e., Rainforest Alliance and Café' Practices) on farmers' welfare. The first control group is the farmers cultivating coffee under Rainforest Alliance and Café' Practices, and the second group is the farmers practicing conventional method. The authors uses a balanced sample for each the treatment group (i.e., farmers under Fair Trade scheme) and the control groups. In order to avoid bias selection of sample characteristic, the researchers also perform Propensity Score Matching (PSM) technique in providing "Probit functions for the likelihood of receiving a particular certification." Therefore, Ruben and Zuniga (2010) use a methodological study that has potential ability in correcting selection bias while assessing and comparing the impacts of various certifications.

3. Ruben and Fort (2011) studies "The impact of Fair Trade certification for coffee farmers in Peru." They find that there is no significant difference, in term of a real price, between the Fair Trade-certified farmers and the conventional producers; thus, the certification simply provides "modest direct income and production effects" to the farmers. However, the certification encourages substantial changes in several aspects beyond the farmers' welfare such as improvement of cooperative, input use, capital and belongings, and risk attitudes.

   Ruben and Fort (2011) use Propensity Score Matching (PSM) technique in their study. This technique allows the researchers to identify a group of conventional farmers who are match, in all applicable pre-treatment attributes, to the certified farmers. Therefore, the technique could provide adequate comparison between the Fair Trade-certified farmers (i.e., as the experiment group) and the non-certified farmers (i.e., as the control group).

4. Blackman and Naranjo (2010) evaluate "The environmental impacts of organic coffee certification in central Costa Rica." The authors find that that organic-coffee certification could develop coffee farmers' ecological performance by inducing some changes in framers' practices. The certification encourages farmers to considerably decrease chemical substances for their farming inputs; while at the same time, farmers are motivated to intensify the use of some "environmentally friendly management practices."

   Blackman and Naranjo (2010), for their study, collected samples of 36 certified organic farms and 2,567 uncertified farms. The data came from three sources: a national census of Costa Rican coffee producers provided by the National Statistics and Census Institute (*Instituto Nacional de Estadística y Censos*, INEC) in collaboration with the Costa Rican Coffee Institute (*Instituto del Café de Costa Rica*, ICAFE), data of GIS complied from a variety of sources, and farmer data list of the Association of Organic Producers of Turrialba (*Asociacion de Productores Organicos de Turrialba*, APOT). To address "self-selection bias", Blackman and Naranjo (2010) use Propensity Score Matching (PSM) technique in their analyses. This technique is useful to construct comparable observed characteristics between the non-certified famers and the certified farmers. In addition, by ensuring the control groups have similar features and circumstances to the experiment groups, Blackman and Naranjo (2010) measures the impact of the certification as "the average treatment effect on the treated (ATT)." This means that the so-called impacts are the differences between the percentage of certified farmers and the percentage of non-certified farmers using the same management practice (i.e., organic farming).

5. Bolwig et al. (2009) studies "The economics of smallholder organic contract farming in Tropical Africa." The authors surveyed a total of 160 farmers in Uganda participating in Kawacom (U) Ltd.'s Sipi organic coffee contract farming scheme. They find that, with several factors are in controlled, participation in contract farming provide more optimistic income for farmers than do application of organic farming methods. The scheme participation could increase the farmers' net revenue by average of 75%, whereas the

impact of organic techniques application may generate the producers' revenue at around 9% averagely.

In this study, the selection of both designs the experiment and control group is based on the two-stage random sampling method. While the experiment group was randomly selected from a list of recorded farmers given by Kawakom, the control group was unsystematically chosen from a list of farmers provided by village leaders. The agro-ecological condition of the two groups was (almost) similar to one another. In addition, Bolwig et al. (2009) use "A standard OLS regression and a full information maximum likelihood (FIML) estimate of Heckman selection models" in providing matched characteristics for comparing the organic with the non-organic farmers. These models are robust, especially in dealing with small sample sizes. Hence, overall, Bolwig et al. (2009) use methods that have a purpose of reducing the biases which frequently occurs as results of inappropriate techniques in selecting samples and in comparing objects.

6. Lyngbaek et al. (2001) conducted a study on "Productivity and profitability of multistrata organic versus conventional coffee farms in Costa Rica". They compare farmers practicing organic farming with those performing conventional method in term of output, profitability, producer-defined limitations as well as objectives and study main concern. The authors compare ten pairs of organic and conventional farms. The conventional farms act as a control group and are matched to the organic farms for a "biophysical and socioeconomic" characteristic. Because of their proximity to the experiment groups, the conventional farms could provide a fair evaluation of the organic practices. Therefore, the assessment of organic certification impacts on the coffee producers could be more reliable.

   The study found that the production of five organic farms was higher than the conventional farm; however, the other three organic farms' yields were 22 % lower than that of conventional plantations. In addition, Lyngbaek et al. (2001) argues that organic certification cost would increase the cost of organic farming. The certification cost makes the organic-price premium could not cover the total cost spent by producers; and this total cost is much higher than the conventional cost. Therefore, the net income received by organic farmers is actually lower than the income received by conventional producers.

References:

Bawden, D. L., and F. L. Sonenstein. 1992. Quasi-Experimental Designs. Chidren and Youth Services Review. Vol. 14, pp. 137.144. 1992

Blackman, A., and J. Rivera. 2010. The Evidence Base for Environmental and Socioeconomic Impacts of "Sustainable" Certification. Resources for the future, Washington, DC.

Blackman, A., and M. A. Naranjo. 2010. Does eco-certification have environmental benefits? Organic coffee in Costa Rica. Working paper. Washington, DC: Resources for the Future.

Bolwig, S., P. Gibson, and S. Jones. 2009. The economics of smallholder organic contract farming in tropical Africa. *World Development* 37(6): 1094–104.

Bracht, G. H., and G. V. Glass. 1968. The ExternalV alidity of Experiments. American Educational Research Journal Volume 5 Number 4 November

Calder, B. J., L. W. Phillips and A. M. Tybout. 1982. The Concept of External Validity. Journal of Consumer Research, Vol. 9, No. 3 pp. 240-244

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): 482–88.

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

Frondel, M., and C. Schmidt. 2005. Evaluating environmental programs: The perspective of modern evaluation research. *Ecological Economics* 55(4): 515–26.

Gorard, S. 2003. Quantitative Methods in Social Science. Continum. London, United Kingdom.

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

International Initiative for Impact Evaluation (3ie) (2008) Principles for Impact Evaluation. 3ie, New Delhi. Available at http://www.3ieimpact.org/media/filer/2012/04/20/principles-for-impact-evaluation.pdf

Khandker, S. R., G. B. Koolwal, H. A. Samad. 2010. Handbook on impact evaluation: quantitative methods and practices. The International Bank for Reconstruction and Development/ The World Bank.

Lyngbaek, A., R. Muschler, and F. Sinclair. 2001. Productivity and profitability of multistrata organic versus conventional coffee farms in Costa Rica. *Agrofestry Systems* 53: 205-213.

Murnane, R. J., and J. B. Willet. 2011. Methods Matter. improving Causal Inference in Educational and Social Science Research. Oxford University Press, Inc. New York, USA.

Ruben, R. and G. Zuniga. 2010. How standards compete: comparative impact of coffee certification schemes in Northern Nicaragua. Supply Chain Management: An International Journal 16/2 (2011) 98–109.

Ruben, R. and R. Fort. 2011. The Impact of Fair Trade Certification for Coffee Farmers in Peru. World Development (2011), doi:10.1016/j.worlddev.2011.07.030.

Ruben, R., R. Fort, and G. Zuniga. 2010. Measuring the Impact of Fair Trade on Development. Development in Practice, Volume 19, Number 6, August 2009.

The Global Social Venture Competition (GSVC). 2012. 2012 GSVC Social Impact Assessment Guidance1. Available at http://www.gsvc.org/docs/2012-GSVC-SIA-GuidanceFinal.pdf.

The Organisation for Economic Co-operation and Development (OECD). n.d. Outline of Princioles of impact evaluation. Access at 31 July 2013. Available at http://www.oecd.org/development/evaluation/dcdndep/37671602.pdf.

US Environmental Protection Agency.2013. Program Evaluation Glossary. Accessed on 31 July, 2013. http://www.epa.gov/evaluate/impact-eval/index.htm.

White, H. 2006. Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank, World Bank, Washington, D.C., p. 3.

World Bank .n.d. The Development IMpact Evaluation (DIME) Initiative, Project Document, World Bank, Washington, D.C. Available at http://siteresources.worldbank.org/INTDEVIMPEVAINI/Resources/DIME_project_document-rev.pdf