ABSTRACT

BAUCH, SIMONE CAROLINA. Evaluation of Conservation Policies in the Brazilian Amazon. (Under the direction of Dr. Erin O. Sills).

The debate about how to conserve nature while promoting sustainable development is ongoing and seemingly never-ending. A key question in this debate is how feasible is it to integrate conservation and people in tropical forest regions. The debate over this approach has been acrimonious and characterized by sweeping generalizations about the negative effects of protected areas on local people, the complete failure of integrated conservation and development, and the lack of sustainability of externally funded projects. In this dissertation, I examine some of these claims by applying rigorous empirical methods to rich panel datasets that I have compiled for this purpose, seeking lessons from the nuances and variation in outcomes, with the goal of moving beyond the generalizations.

The three essays in this dissertation focus on different dimensions of the relationship between protected areas (PAs) and human welfare in the Brazilian Amazon, which has the largest contiguous remaining area of tropical forest in the world, half in protected areas. First, I evaluate the impact of PAs on the incidence of malaria, a reemerging and endemic disease with 500 thousand annual cases in the Brazilian Amazon alone. Malaria is prevalent in disturbed forest environments and its association with deforestation has been well documented. My contribution is to assess the link between malaria and policy variables, specifically different types of PAs and roads. Since some of the key policies of interest do not vary across time in my dataset, I use an innovative fixed effects vector decomposition method to include these time-invariant variables in a panel model. I find that strict protection

PAs and Indigenous reserves decrease malaria, as expected since these areas have a lower population density and indigenous peoples have natural resistance to the disease. Roads (both official and unofficial) increase malaria, indicating that infrastructure development tends to increase the disease incidence.

The second and third essays focus on a particular sustainable use protected area, the Tapajós National Forest. Specifically, I evaluate participation in and effects of community-based microenterprises supported through an ICDP (Integrated Conservation and Development Program). Chapter 2 evaluates the program in terms of the impacts of participation during the external financing phase using a panel dataset formed by household observations in 1997 (2 years before the ICDP was implemented) and 2006 (the year funding ended). I control for the potential confounding effects of program placement and selection by applying propensity score matching to evaluate difference-in-difference outcomes. Participation in the microenterprises results in higher cash income, but this is not reflected in greater asset wealth or other indicators of well-being. Participation also not have consistent effects on agricultural cropping area and size of the cattle herd, the direct environmental targets of the ICDP. Considering that $1.4 million was invested in the ICDP, the lack of impact at the household level is a significant finding.

In my third essay, I evaluate the retention of households in the microenterprises and decisions about agricultural cropping area in the Tapajós National Forest between 2006 and 2008. I hypothesize that social networks have a significant role in determining who stays in the projects and who decreases land in agriculture (two measures of success relevant to the

external funders). Estimation results show that social networks do have an important influence, and that the different microenterprises have heterogeneous effects (with some more likely to retain household participants, and some more likely to encourage reductions in agricultural areas). Overall, my results offer cautionary lessons about the difficulty of conserving nature while promoting sustainable development and illustrate the importance and need for well-designed empirical research.

Evaluation of Conservation Policies in the Brazilian Amazon

by
Simone Carolina Bauch

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Forestry and Environmental Resources

Raleigh, North Carolina

2010

APPROVED BY:

_____	_____
Subhrendu K. Pattanayak	Christopher Timmins
_____	_____
Walter Thurman	Erin O. Sills Chair of Advisory Committee

DEDICATION

To Papi and Mami.

BIOGRAPHY

Simone Carolina Bauch was born in Alagoinhas, Bahia, Brazil on June 25th 1980. While growing up in Brazil and in Nicaragua (and traveling a lot), her father, a forester, taught her by example to focus on sustainable development and environmental, always encouraging her to do what she truly likes.

After graduating from High School, Simone earned a BS in Forest Engineering at the University of São Paulo, graduating with honors and receiving several awards. To continue her education, Simone went to Virginia Tech where she completed a Master of Science in Forest Economics in July 2004. She studied the rents and costs to logging in the Brazilian Amazon. She then completed the core year classes for a PhD in economics and after passing her prelims took off to work at Imazon – the Amazon Institute for People and the Environment – in Belém, Brazil. She coordinated market research, community commercialization and worked closely with her advisor to be, Dr. Erin Sills. Having been admitted for the PhD program in Forest Economics at North Carolina State, she moved to Raleigh in August 2006, while maintaining close contact with Imazon and continued to spend 3 months a year there. During her PhD Simone was awarded a Forestry and Environmental Resources Department Assistantship, a BECA grant from IIEB in Brazil, a Latin American and Caribbean Environmental Economics Program research grant and the Cooper Graduate Fellowship for Academic Excellence in Social Science. While at NCSU, Simone was the president for the International Society of Tropical Foresters, the NCSU representative for the

Working Group on the Environment in Latin America, the president of the Brazilian Student Association and a council member for the Latin American Association of Environmental and Natural Resource Economists (ALEAR).

Currently Simone is working with CIFOR (the Center for International Forestry Research) on a Global Comparative Study on REDD+ and in August 2010 she will start working as a post-doc at the Norwegian University of Life Sciences (UMB) in Ås, Norway, working on the same project with CIFOR but in an academic setting.

ACKNOWLEDGEMENTS

"Good company in a journey makes the way seem shorter." Izaak Walton

I want to thank all of you that have guided, helped, cheered, supported and advised me through this journey.

First, I want to thank my advisor, Dr. Erin Sills for her insight into research and interest on environment and development (especially in Brazil), but also for the inexhaustible energy and patience. Dr. Subhrendu Pattanayak deserves appreciation for his guidance and ability to foresee upcoming issues. I'm thankful for Dr. Chris Timmins' ability to encourage and specially link seemingly unrelated topics. I am thankful for Dr. Wally Thurman's ability to simplify discussions and analyzes and bring us back to Earth.

I would also like to thank all researchers at Imazon that supported me through this journey, for all their help during the Summer months and patience when my dissertation kept me from focusing on other things. I am grateful to all the households that participated in my research both in the Tapajós National Forest, in Gurupá, Porto de Moz and islands surrounding Belém. The gift of their time made it possible to accomplish my degree and hopefully contribute to their well being. I also thank the government officials that made my life easier and all the researchers I interacted with in the endeavor to collect and analyze data. I would also like to thank IIEB for their financial support through the BECA program, LACEEP for

their research grant and the Forestry and Environmental Resources Department for the Research Assistantship.

This quest would have been very tough, mostly impossible, without friends. Those going through similar experiences and could relate or guide me like Shubho, Rodrigo, Nia and Yun did. Those that were in very different walks of life and helped remind me of the bigger picture out there. For those in the USA, in Brazil and spread through the world that always tried to figure out where I was. To my family and their unconditional support through rough

times and difficult episodes. And finally, to my boyfriend, Arvind, for his never-ending patience and for always making life more interesting.

TABLE OF CONTENTS

LIST OF TABLES vii
LIST OF FIGURES x
1. Introduction 1
1.1 References 6
2. Development, Deforestation and Disease: Policy Implications for Malaria 9
2.1 Abstract 9
2.2 Introduction 11
2.3 Methods 18
2.3.1 Data 18
2.3.2 Statistical methods 21
2.4 Results 26
2.4.1 Descriptive statistics 26
2.4.2 Correlations 29
2.4.3 Regression results 29
2.4.4 Robustness check 31
2.5 Discussion 31
2.6 Conclusion and policy implications 33
2.7 References 34
APPENDICES 47
Appendix 2.A Protected Areas in Brazil 48
Appendix 2.B The FEVD model 57
Appendix 2.C FEVD results "controlling" for spatial correlation 64
Appendix 2.D Modeling deforestation 72
Appendix 2.E Clustered quantile regression models for malaria 74
3. Revisiting ICDPs: the Impact of Community Based Enterprises in the Brazilian Amazon 81
3.1 Abstract 81
3.2 Introduction 82
3.2.1 Study site 86
3.2.1.1 Agriculture in the National Forest 87
3.2.1.2 ProManejo 88
3.3 Data 90
3.4 Empirical strategy 93
3.4.1 Matching (Based on Heckman and Navarro, 2004) 95
3.4.2 Panel based fixed-effect estimators (DID) (Based on Ravallion, 2001 and Albouy, 2005) 99

3.5 Defining treatment, controls and outcomes..... 103
3.6 Results 107
3.6.1 Communities with versus without CBEs 107
3.6.2 Percentage of households participating in CBEs 108
3.6.3 Households in CBEs versus non participants 109
3.7 Conclusion 112
3.8 References 114
APPENDICES 145
Appendix 3.A Timeline of relevant events between the communities and Ibama at the Tapajós National Forest 146
Appendix 3.B Sample attrition in TNF database between 1997 and 2006 149
Appendix 3.C Propensity score estimation 152
4. Microenterprises after the funding ends: Two Steps Forward and One Step 161
4.1 Abstract 161
4.2 Introduction 163
4.3 Literature review 166
4.4 Motivation and empirical model 171
4.5 Data and context 175
4.6 Empirics 183
4.7 Conclusions and policy implications 196
4.8 References 199
APPENDICES 211
Appendix 4.A Literature review on social network studies in development applications 212
Appendix 4.B Theoretical model 220

LIST OF TABLES

Table 2.1 Literature relating deforestation to malaria (N is the sample size, K is the number of covariates) 42

Table 2.2 Descriptive statistics for the municipalities in the Brazilian Amazon (N=378) 43

Table 2.3 Correlations between variables 44

Table 2.4 Results of third stage of FEVD and robustness checks of log of malaria cases 46

Table 2.C.1 Results of first stage of Plümper and Troeger (2007) – fixed effects model of cases of the difference in the log of malaria cases per km2 unweighted and weighted by population/ total number of malaria tests 64

Table 2.C.2 Results of third stage of Plümper and Troeger – Clustered OLS of log of malaria cases per km2, unweighted and weighted by population/ total number of malaria tests 66

Table 2.C.3 Results of third stage of Plümper and Troeger – Clustered OLS, weighted and unweighted models of log of malaria cases omitting deforestation 70

Table 2.D.1 Annual deforestation as a function of policy variables and covariates..... 72

Table 2.E.1 Results of the clustered quantile regression for the log positive cases of malaria in the 2002-2006 period, weighted by municipality area 76

Table 2.E.2 Results of the reduced form model of clustered quantile regression for the sum of positive cases of malaria in the 2002-2006 period, weighted by municipality area (without deforestation) 78

Table 2.E.3 Results of the clustered quantile regression for the log positive cases of malaria in the 2002-2006 period, weighted by municipality area, accounting for spatial correlation 80

Table 3.1 Land use in the five population areas within the Tapajós National Forest (Source: Cohenca, 2007) 121

Table 3.2	Projects supported by Component IV of ProManejo in the Tapajós National Forest (highlighted projects refer to CBEs evaluated in this chapter)…….....	124
Table 3.3	Descriptive statistics and T tests ……..	129
Table 3.4	Outcome for treated and control communities in 1997 …….....	133
Table 3.5	Outcomes of treated and control communities in 2006…….….....	134
Table 3.6	Descriptive statistics for 1997 and 2006 by current participants in the CBEs.....	138
Table 3.7	Descriptive statistics for 1997 and 2006 by having ever participated in CBEs…….....	140
Table 3.8	Matching results for 2006 participants versus never having participated in CBEs.....	142
Table 3.9	Matching results for having ever participated in CBEs versus not…….....	143
Table 3.10	Logit of CBE dropouts (Likelihood ratio test is significant at 2% significance level and pseudo R2 is 25%, sample size of 145 households)…….....	144
Table 3.B.1	Correlation between leaving the sample and 1997 household characteristics….....	149
Table 3.B.2	Multivariate probit of sample attrition (model significant at 1%) …….	150
Table 3.C.1	Results of the propensity score logit for current participation…….....	153
Table 3.C.2	Treament and control characteristics before and after matching (based on matching with 2 nearest neighbors)…….....	155
Table 3.C.3	Treament and control characteristics before and after matching (based on kernel matching)…….....	157
Table 3.C.4	Propensity score logit for having ever participated in the CBEs …….....	158

Table 3.C.5 Treatment and control characteristics for households who ever participated in CBEs before and after matching (based on matching with two nearest neighbors)……..... 159

Table 3.C.6 Treatment and control characteristics for households who ever participated in CBEs before and after matching (based on matching with two nearest neighbors)……..... 160

Table 4.1 Descriptive statistics of household variables…….…….……..... 206

Table 4.2 Results of OLS regressions of household characteristics that explain the number of links for each type of interaction matrix…….……..... 207

Table 4.3 Household models of decision to participate in microenterprises (projectpd as LHS variable).....….…..... 208

Table 4.4 Household models of decision to decrease land in agriculture (agdec as LHS variable)…….. 209

Table 4.5 Bivariate probit for simultaneous decision of whether to participate in microenterprises and whether to decrease land in agriculture (agdec and projectpd)……..... 210

Table 4.A.1 Literature review of social networks in development applications in developing countries……..... 212

LIST OF FIGURES

Figure 2.1	Malaria, rainfall and policies in the Brazilian Amazon.....	41
Figure 3.1	Map of communities in the Tapajós National Forest	122
Figure 3.2	Satellite image of the Tapajós National Forest (limits in yellow), communities (dots) and deforested areas (pink and white)…….....	123
Figure 3.3	Timeline of interventions and surveys in the Tapajós National Forest….....	128
Figure 3.4	Bias in double difference estimates for a targeted problem (from Ravallion, 2001)……	132
Figure 3.5	Difference in cash income between 1997 and 2006 for control and treatment.....	135
Figure 3.6	Difference in liquid assets between 1997 and 2006 for control and treatment…….....	135
Figure 3.7	Difference in total assets between 1997 and 2006 for control and treatment….....	136
Figure 3.8	Difference in heads of cattle between 1997 and 2006 for control and treatment….....	136
Figure 3.9	Difference in land in agriculture between 1997 and 2006 for control and treatment communities…….....	137
Figure 3.10	Common support graph for current participants (using households that never participated as controls)	142
Figure 3.11	Common support graph for having ever participated in CBEs…….....	143

1. Introduction

The debate about how to conserve nature while promoting sustainable development is ongoing and seemingly never-ending. The underlying concern is how feasible is it to integrate conservation and people. While initially conservation strategies revolved around separating these two elements, the 1980s and early 1990s witnessed an apparent sea-change in government attitudes towards the management of common pool natural resources in India and elsewhere (Lele, 2000). 'Joint management', 'co-management', or 'shared management' of natural resources with local communities became and remains a dominant recommendation for conservation policy (Poffenberger and McGean, 1996). Brazil has been in the lead in developing and testing new models and approaches to implement this recommendation in human-inhabited protected areas over the past two decades (Allegretti 1990)[1]. The debate over this approach (reviewed in Adams and Hutton 2007) has been acrimonious and characterized by sweeping generalizations about the negative effects of protected areas on local people, the complete failure of integrated conservation and development, and the lack of sustainability of externally funded projects. In this dissertation, I examine some of these claims by applying rigorous empirical methods to rich panel datasets that I have compiled for this purpose, seeking lessons from the nuances and variation in outcomes, with the goal of moving beyond the generalizations.

[1] An example of support for people-centered conservation change was the approval of the PPG7 (the Pilot Program to Conserve the Brazilian Rain Forest) which has invested $428 million since 1992.

The relationship between people and protected areas is nowhere more fundamental than in the Brazilian Amazon, which has the largest remaining contiguous area of tropical forest (3.3 million square kilometers), over half currently in officially designated protected areas (Nepstad et al. 2009). In 2000, Brazil adopted a new National Conservation Unit System that divided protected areas (PAs) into two broad categories: strict protection (which allows only for indirect use) and sustainable use (which allows sustainable management of its resources). The most emphasis has been placed on creating and implementing sustainable use PAs, which allow for resident indigenous populations and/ or commercial forest management. Reasons include their lower political cost of implementation, their perceived utility as a means to resolve disputed land tenure, and foreseen profits of selective logging while maintaining forest cover. There is great variety across these protected areas in terms of both their resident populations (e.g., indigenous vs. traditional rural) and implementation strategies (e.g., integrated conservation and development, co-management, or no active intervention).

While there is a rich case study literature on the impacts of protected areas and their programs, there have been relatively little empirical analysis employing relevant econometric and program evaluation methods (Ferraro and Pattanayak 2006). For example, at a regional scale, there have been a few rigorous empirical assessments of the effects of protected areas on deforestation (Andam et al, 2008, Gaveau et al, 2007, Pfaff et al, 2009, Nelson and Chomitz, 2009, Nepstad et al, 2006), but similar methods have rarely been employed to examine other impacts (notable exceptions are Gaveau et al, 2009 and Andam et al, 2008).

Likewise, there is a large and growing literature on integrated conservation and development projects in and around protected areas (Garnett et al, 2007), but few studies that attempt to tease out causality or model household decisions that determine the long-term sustainability of the projects. This dissertation addresses this gap at both a local and a regional scale.

In the first essay of my dissertation, I evaluate the impact of protected areas on the incidence of malaria, a reemerging and endemic disease with 1.9 million annual cases in the Brazilian Amazon. Malaria is prevalent in disturbed forest environments and its association with deforestation has been well documented. My contribution is to assess the link between malaria and policy variables, specifically different types of protected areas (PAs) and roads. I compile a rich dataset comprised of municipality observations on policies and covariates such as health care, climate and socio-economic characteristics. Since some of the key policies of interest do not vary across time in my dataset, I use an innovative time-series cross section estimation method that allows me to include time-invariant variables into a fixed effects approach. I find that sustainable use PAs have the unintended consequence of increasing malaria prevalence, while strict protection PAs and indigenous reserves do not have this effect. The effects on roads are not significant and might be due to the fact that despite dispersing the human population into areas of active deforestation they also function as means of access to health care. Protected areas are not bad though, as a similar specification without controlling for deforestation shows that in this case malaria would be even higher. This suggests that protected areas do work as conservation policies but also helps explain another dimension of regional opposition to them.

In my other two essays, I focus on a particular sustainable use protected area, the Tapajós National Forest. Specifically, I evaluate participation in and effects of community-based microenterprises supported through an ICDP (Integrated Conservation and Development Program). This ICDP sought to engage resident households in sustainable forest management in order to draw them out of slash-and-burn agriculture that is considered to degrade the environmental quality of the PA. Chapter 2 evaluates the program in terms of the impacts of participation during the external financing phase using a panel dataset formed by household observations in 1997 (2 years before the ICDP was implemented) and 2006 (the year funding ended). The program developed the microenterprises in only some communities, and households in those communities had the choice of whether to participate. I control for the potential confounding effects of program placement and selection by applying propensity score matching to evaluate difference-in-difference outcomes. Participation in the microenterprises results in higher cash income, but this is not reflected in greater asset wealth or other indicators of well-being. Area used for agricultural crops and the size of the cattle herd, the direct environmental targets of the ICDP, are also not significantly different for participants (in terms of land in agriculture and limited in cattle). While the overall ICDP may have had important impacts that affected all communities and households in the National Forest, my results suggest that the community micro-enterprises in particular did not achieve the hoped-for impacts on welfare or resource use. This is important because this is a commonly employed strategy in sustainable use protected areas. Considering that an investment of $1.4 million was made in the ICDP, it is important to show that results are not as significant as promoted.

Externally funded projects always have end dates, and thus it is critically important to evaluate household decisions once external funding ends in order to understand long-term impacts. This is the focus of my third essay, in which I evaluate the retention of households in the microenterprises and decisions about agricultural cropping area in the Tapajós National Forest between 2006 and 2008. Since operation of the microenterprises was transferred to the communities in that time period, I hypothesize that social networks have a significant role in determining who stays in the projects and who decrease land in agriculture (two measures of success relevant to the external funders). The literature on social networks has been rapidly expanding. However, my comprehensive literature review of empirical studies of the role of social networks in developing countries did not identify any previous studies of exit from programs or disadoption of technologies, and confirmed that many have the same data constraints (i.e. no panel data on social networks). Estimation results show that social networks do have an important influence, and that the different microenterprises have heterogeneous effects (with some more likely to retain household participants, and some more likely to encourage reductions in agricultural areas).

To inform the debate surrounding conservation policies in the Brazilian Amazon. I have sought to compile rich panel datasets and apply state-of-the-art empirical methods to assess causality, understand mechanisms such as social networks, and contribute to the public debate on key issues such as environmental health. The fact that my results are somewhat surprising in some cases should be taken as indication of the importance and need for well-designed research. I hope the chapters ahead provide support for this argument. Enjoy!

1.1 References

Allegretti, M.H. 1990. Extractive reserves: An alternative for reconciling development and environmental conservation in Amazonia. In: Anderson, A.B. (ed.). Alternatives to deforestation: Steps toward sustainable use of the Amazon rain forest.

Adams, W.M., Hutton, J. 2007. People, Parks and Poverty: Political Ecology and Biodiversity Conservation. Conservation and Society 5(2): 147-183.

Andam, K.S., Ferraro, P.J., Pfaff, A., Sanches-Azofeifa, G.A., Robalino, J.A. 2008. Measuring the effectiveness of protected area networks in reducing deforestation. Proceedings of the National Academy of Sciences 105(42):16091-16094.

Andam, K.S., Ferraro, P.J., Holland, M.B. 2008. Do Protected Areas Increase Poverty? Evidence from Costa Rica. Presented at AAEA 2008, Orlando, FL.

Ferraro, P.J., Pattanayak, S.K. 2006. Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments *PLoS Biology* 4(4) Available : http://biology.plosjournals.org/perlserv?request=get-document&doi=10.1371/journal.pbio.0040105

Garnett, S.T., Sayer, J., du Toit, J. 2007. Improving the effectiveness of interventions to balance conservation and development: a conceptual framework. Ecology and Society 12(1): 2. [online] URL: http://www.ecologyandsociety.org/vol12/iss1/art2/

Gaveau, D.L.A., Wandono, H., Setiabudi, F. 2007. Three decades of deforestation in southwest Sumatra: Have protected areas halted forest loss and logging and promoted re-growth? Biological Conservation 134: 495-504.

Gaveau, D.L.A., Linkie, M., Suyadi, Levang, P., Leader-Williams, N. 2009. Three decades of deforestation in southwest Sumatra: Effects of coffee prices, law enforcement and rural poverty. Biological Conservation 142: 597-605

Lele, S. 2000. Godsend, Sleight of Hand, or Just Muddling Through: Joint Water and Forest Management in India. Natural Resources Perspective #53.

Nelson, A., Chomitz, K.M. 2009. Reducing Tropical Deforestation. A Global Analysis of the Impact of Protection Status. The World Bank IEG, Evaluation Brief 7.

Nepstad, D., Schwartzman, S., Bamberger, B., Santilli, M., Ray, D., Schlesinger, P., Lefebvre, P., Alencar, A., Prinz, E., Fiske, G., Rolla, A. 2006. Inhibition of Amazon Deforestation and Fire by Parks and Indigenous Lands. Conservation Biology Volume 20, No. 1, 65–73

Nepstad, D., Soares-Filho, B.S., Merry, F., Lima, A., Moutinho, P., Carter, J., Bowman, M., Cattaneo, A., Rodrigues, H., Schwartzman, S., McGrath, S.G., Stickler, C.M., Lubowski, R., Piris-Cabezas, P., Rivero, S., Alencar, A., Almeida, O., Stella, O. 2009. The End of Deforestation in the Brazilian Amazon. Science 326:350-351.

Pfaff, A., Robalino, J., Sanchez-Azofeifa, G.A., Andam, K.S., Ferraro, P.J. 2009. Park Location Affects Forest Protection: Land Characteristics Cause Differences in Park Impacts Across Costa Rica. B.E. Journal of Economic Analysis & Policy 9(2) #5.

Poffenberger M, McGean B. 1996. Village Voices, Forest Choices: Joint Forest Management in India. New Delhi: Oxford Univ. Press. 356 pp.

2. The road less traveled: Does forest conservation reduce malaria in the Brazilian Amazon?

2.1 Abstract

While tropical deforestation has attracted global concern due primarily to loss of biodiversity and carbon stocks, there is also an increasing recognition that human health depends on ecosystem – including tropical forest - health (MEA, 2005). Key policy recommendations for reducing deforestation, including expansion of protected area networks and restriction of road networks, face regional opposition due to the perception that they would impose local costs in order to produce global public goods. While there are potentially local benefits of such policies, these are less well documented. For example, one key question is the net public health effects of conservation policy levers such as establishing different types of protected areas (PAs) or slowing expansion of the official or unofficial road network. Literature suggests a multidimensional approach to eliciting this link in which socio-economic and biological factors are considered jointly to elicit the effect of policies and thus guide decision makers in incorporating health into conservation decisions. This is a critical issue in the Brazilian Amazon, which suffers from endemic malaria, the world's highest deforestation rates, persistent poverty, and an expansive protected area network. Among migrants to the region, there is a widespread belief that the forest is the source of malaria, in contrast to recent research that finds less malaria where the forest is relatively undisturbed (Pattanayak, 2006). One way to reconcile these views would be if it is the active interface of humans and forest fosters malaria, and this would in turn raise serious questions about integrating PAs

designated for sustainable use of the forest into a landscape patchwork of human use and conservation.

I address this issue by estimating a Time Series Cross Section model of malaria as a function of key policy levers (various types of roads and protected areas), while controlling for ecological and social covariates. My dataset is unique in that it combines municipal-level panel data on malaria and key covariates with time-invariant but highly differentiated policy variables (i.e., different types of protected areas and roads). To fully exploit these data, I employ the model presented by Plümper and Troeger (2007) to incorporate invariant and rarely changing variables into the usual Time Series Cross Section estimation. The results show that strict protection PAs (which do not allow people living within its boundaries) decrease malaria incidence in a municipality, while sustainable use PAs (which allow for people residing within its boundaries and sustainable management of resources) do not affect malaria. Both official as well as unofficial roads positively (and significantly) affect malaria as they allow for people to disperse into areas of disturbed forest. Inmigration to the Amazon from people outside the region increases malaria (indicating that people from areas with no malaria have lower resistance) while migration by people from the Amazon elsewhere in the region decreases the disease prevalence. Despite the positive relation between sustainable use PAs to malaria, there is evidence that they fulfill their conservation objectives and that by reducing deforestation there is less malaria than there would be otherwise. These results demonstrate the need to incorporate the health dimension into the decision making of conservation policies.

Keywords: malaria, deforestation, protected areas, migration, roads, Amazon, fixed effects, TSCS

2.2 Introduction

The natural environment threads the very fabric of human lives in remote forest areas of developing countries. Yet these forests are being cut fast and furiously. One explanation is that we do not fully appreciate the mechanisms and magnitude of ecosystem benefits of forest conservation (Pattanayak and Butry, 2005). This lack of understanding is showcased by the poor quality of information on linkages between ecosystem degradation and human health (Pattanayak and Wendland, 2007). A recent first approximation suggests that 25% of the global burden of disease is attributable to poor environmental policies, including those impacting deforestation (WHO, 2006). Such estimates partly reflect attempts to step back from an excessively bio-medical and molecular view of health to a broader focus on community and environmental prevention strategies (Parkes et al, 2003). It also reflects greater attention to climate change related ecosystem services that highlight the positive role of tropical forest conservation for regional stakeholders and even global health security (Turner and Daily, 2008; Daily and Ehrlich, 1996a,b). However, the desire to make the case for forest conservation has led to broad claims about the benefits of forest conservation for human health with no or few empirical analyses (Pattanayak and Yasuoka, 2008; Pattanayak et al, 2009). For example, the claim that 'deforestation causes malaria' directly contradicts local wisdom that relates tropical diseases with tropical forests. More importantly, it is imperative to link such empirical ecosystem valuation studies to specific policies because the

evidence base for policy design in developing countries is particularly weak and scarce research funds must deliver practical advice to policy markers (Pattanayak and Pfaff, 2009). Thus, I address these challenges by building a unique panel data set on malaria, deforestation, and conservation policies in Brazil and applying a novel empirical methodology.

Rapid reductions in biodiversity and carbon stocks has re-ignited concerns about deforestation, which is one of the substantial human drivers of environmental change and therefore key to many of the climate change mitigation and adaptation debates. Current world deforestation levels hovered around 7.3 million hectares per year in the 2000-2005 period (down from 8.9 million hectares per year in the 1990-2000 period) (FAO, 2006). Out of this total, South America lost 4.3 million hectares of forests annually in the 2000-2005 period, and Brazil was responsible for almost all of this amount. National estimates from Brazil show a yearly deforestation average of over 1.8 million hectares in the 1988-2007 period. Although there are many drivers of deforestation, we can roughly think of these as factors that determine the (1) private and local benefits of agricultural use, and (2) direct and indirect costs of clearing forest. Here we briefly discuss some of these:

- Roads "cause" deforestation by decreasing transportation costs and therefore increasing profitability of agriculture relative to forest management, as predicted by the von Thünen model (Pfaff, 1999; Pfaff et al, 2007; Schaefer et al, 2005; Chomitz, 2007). As such, transportation costs determine the reach of economic activities by decreasing the costs and increasing the returns to both timber extraction and agriculture.

- Land tenure: The lack of clear property rights for forested land increases deforestation by reducing the returns to long-term forest management and increasing the returns to clearing land (and therefore more clearly establishing land tenure). To control the supply of land available for deforestation, a popular policy response is the creation of protected areas (PAs). In this public policy scenario, access to an area is limited through enforcement of boundaries and restrictions on economic and human activities with the intention of avoiding impact on the existing ecosystem. This policy has been applied in the United States since 1916 (Hays, 1987) and adopted in most countries. There are over 100,000 protected areas worldwide, covering over 12% of the Earth's land surface (Chape et al, 2005). In the specific case of Brazil, there are two major categories of land use restrictions: strict protection (which only allows indirect use and research activities) and sustainable use (which allows for resident human populations and the development of sustainable economic activities) and cover a total of 209 million hectares in the Brazilian Amazon alone (the equivalent to 43% of the land) (Barreto et al, 2008).
- Commodity prices: The most basic prediction of economics is that supply increases with increases in output prices (or decreases in input prices). This suggests that agricultural commodity prices are a fundamental driver of deforestation. For example, the lagged prices of both soy beans and beef have good predictive power over the next period's deforestation in Brazil (Pattanayak et al, 2009). Also, increases in cattle herds (Margulis, 2004) as well as agricultural operations (Simon and Garagorry, 2005) have been linked to deforestation. Soy expansion has been specifically linked to deforestation (due to strong international demand, government incentive, and favorable climatic conditions) (Morton et al, 2006). Input prices

reflect international market conditions, access (fundamentally driven by roads, as discussed above), land tenure or supply, national economic conditions and policies (e.g., driving wage rates), and any subsidies offered by the government (e.g., tax breaks, protection against imports, support for phytosanitary, transport, and storage) which artificially increase profitability of this sector leading to similar effects as an increase in prices.

Policies for reducing deforestation include restricting road expansion and expanding protected area network. But there is regional opposition to these policies, which appear to impose local costs - e.g. restricted access to public services and economic development - in order to produce global public goods. A major question, and the one addressed by my chapter, is if the health costs of deforestation outweigh the conventional economic benefits to the region (Pattanayak et al, 2009).

Malaria and deforestation are endemic to many tropical areas and central elements of a vicious cycle of rural poverty in many developing countries (Pattanayak et al, 2006). While these problems are not new they have gained significance because malaria is a reemerging tropical disease that kills more people today than 40 years ago (Wilcox and Colwell, 2005; Greenwood et al, 2005) and global deforestation is still quite high (FAO, 2006). There is a thin literature linking malaria to deforestation (see table 2.1) and several studies lack multivariate analyses with controls for confounding factors. Another shortcoming of most studies is the lack of attention to the mechanisms through which deforestation affects malaria (Pattanayak and Yasuoka, 2008). The positive correlation found in these studies suggests that

there could be “win-win” policy interventions to reduce both malaria and deforestation. In this chapter, I estimate these effects and present empirical evidence on the paths through which policies affect malaria by using municipality level data for the Brazilian Amazon to estimate time series cross section (TSCS) models of malaria as a function of policies and covariates.

The worldwide malaria burden is substantial. There are 300 to 500 million cases of malaria yearly and between one to three million deaths attributed to this disease (Sachs and Malaney, 2005). In the Brazilian Amazon there are approximately 500 thousand cases of malaria per year, with most of it at an endemic level of 0.1 cases per km^2 per year and some epidemic breakouts with cases up to 60,000 cases in one year (see figure 2.1A for the distribution of malaria cases in 2006). The incidence of malaria is highly dependent on environmental factors and land use, since the vector population (mosquitoes of the Anopheles genus) is directly affected by these (Yasuoka and Levins, 2007). 70-90% of the risk of malaria is attributed to environmental factors (Smith et al, 1999) (see figure 2.1B for distribution of rainfall in the Brazilian Amazon). Also, it is important to consider that infectious diseases are part of a broader human ecology which considers human social systems, economic activities, interactions with the environment and lifestyles that affect the key domains of interaction that affect infection and disease risk (Wilson, 2001). As such, malaria transmission and control have clear links to ecosystem changes that result from natural resource policies on land tenure, road building and agricultural subsidies (Pattanayak and Yasuoka, 2008).

Consider at least four potential pathways through which forest management and deforestation can affect malaria infection (Pattanayak and Yasuoka, 2008). First, deforestation changes the ecology of a disease vector and its options for hosts; favoring the development of *Anopheles* larvae. Second, deforestation can affect climate at local, regional, and even global scales. Third, deforestation is often just a first step in a chain of land-use changes. Fourth, deforestation is accompanied by migration and other behavioral changes that may enhance the spread of malaria (Daily and Ehrlich, 1996a). Finally, ecosystem change can play a role in the antibiotic resistance that has become a major concern for several plasmodium species. These links depend on the species of *Anopheles sp.* present in the area and land use implemented post deforestation as well (Yasuoka and Levins, 2007). As argued in literature, it is "considered critical to incorporate deforestation and forest management into malaria research, not only because of the potential linkages enumerated earlier, but also because of linkages through the human causes and consequences of malaria and deforestation" (Pattanayak et al, 2006 p.50).

Assessing the relation between deforestation and malaria is also an important public policy concern because the indirect, albeit significant effect of development and conservation can indirectly generate health benefits. For example, policies such as roads can increase deforestation significantly (Pfaff, 1999; Pfaff et al, 2007; Schaefer et al, 2005; Chomitz, 2007), while also reducing malaria through improved accessibility or increasing it due to their role in dispersing population through the deforestation frontier. Figure 2.1C shows the distribution of official and unofficial roads in the Brazilian Amazon. Another policy example

is the implementation of protected areas (PAs) which can reduce malaria by limiting the area where development occurs and thus, the interaction of humans and the disease vector. Alternatively certain types of PAs that allow for local population to remain in close contact to forest and vectors (the so called "sustainable use" PAs) can also increase malaria. Figure 2.1D shows the network of protected areas and indigenous reserves in the Brazilian Amazon as of 2002. Therefore, Pattanayak et al (2006 p.46) call for malaria control policy options to be enhanced "by the application and dissemination of transdisciplinary research that integrates economics, ecology, and epidemiology to examine the critical nexus of malaria, deforestation, and economic development."

This chapter seeks to identify the impact of conservation policies on malaria, while controlling for ecological, economic and social factors thought to explain malaria prevalence in the Brazilian Amazon. Specifically, I consider the impacts of three types of protected areas and road expansions. I control for biological characteristics that determine the natural behavior of the vector and socio-economic characteristics that affects access to health care. Finally, I consider how malaria is affected both by deforestation and by the subsequent land uses that replaces forest. my major bio-statistical challenge is to construct a rich panel data set and employ credible estimation methods, while accounting potentially major sources of confounding. The typical fixed effect strategy throws out the "baby with the bathwater" because policy variables and ecological factors typically do not change over data time window. I apply an innovative 'fixed effect vector decomposition" strategy (a type of hierarchical linear model) that uses as much of the cross-section variations as I can to

estimate coefficient on time invariant policies. By examining one specific mechanism (health) through which PAs impact local populations, my chapter contributes to a small but growing econometrics literature that examines development impacts of PAs using secondary statistics.

2.3 Methods

2.3.1 Data

The Brazilian Amazon is an appropriate setting to test the link between malaria and conservation policies. First, much of the Brazilian Amazon is affected by endemic malaria (with an average of 612 cases per municipality per year over the 2003-2006 period), has historically had one of the highest absolute levels of deforestation in the world (between 1998 and 2008, annual losses of forest cover have averaged about 1.8 million hectares). The Brazilian government has been actively pursuing a conservation and sustainable development agenda for the region, including a protected area program that has attracted international attention largely due to the enormous biodiversity of the region. Finally, there are reasonably good quality secondary data at the municipal level publicly available in Brazil.

The data set used in the analysis is comprised of yearly (2003-2006) municipality level observations for malaria prevalence, deforestation, agriculture (perennial and annual) area, head of cattle, health program intensity; monthly data on rainfall and temperature for the same period; and one-time (cross-sectional) observations for GDP, population density, protected areas (sustainable use and strict protection), and paved roads for 378 municipalities

in the Brazilian Amazon. Most of the data come from publicly available secondary sources, such as the IBGE (Brazilian Institute of Geography and Statistics), IPEA (Applied Economics Research Institute), INPE (National Institute of Spatial Research), and IBAMA (Brazilian Institute of the Environment and Renewable Natural Resources). Besides these public sources of data, I incorporate yearly deforestation rates per municipality and unofficial road density calculated by Imazon (Amazon Institute of People and the Environment) and yearly climatic data from Conservation International.

The resulting dataset is quite comprehensive in terms of social, economic and climatic covariates affecting both deforestation and malaria, and detailed in terms of key policy variables, distinguishing between two types of roads (unofficial and official) and three types of protected areas: sustainable use and strict protection protected areas and indigenous reserves. Official roads refer to those planned by the government and encompass the main highways that link urban centers in the region. These roads are often constructed in completely forested areas and drive deforestation (Pfaff, 1999). The informal road network is comprised by roads opened by economic agents such as loggers and farmers to facilitate access to land. This results in a much denser network around human settlements. Thus, these different types of roads might have different effects on malaria incidence, reflecting their differential impact on access to intact forest (official roads) or disturbed forest (unofficial roads).

The protected area (PA) system in Brazil (National System of Conservation Units - SNUC) includes into two broad categories):

- Sustainable use PAs allow people to live within their borders and practice sustainable economic activities (thus a disturbed forest environment);
- Strict protection PAs have conservation as their primary objective, as such, only indirect uses (such as research) are allowed.

Third, while not a specific category of PA, indigenous reserves encompass large tracts of intact forest where indigenous groups live. These peoples have control over the land use within its boundaries.

While the dataset is very rich, it has two important limitations. First, there are only one-time observations on key policy variables (as well as some covariates). This reflects both the difficulty of generating these variables (such as unofficial road network) and the relatively limited change in some variables over the four year period (e.g. existence of different types of protected areas by municipality). A second limitation is the municipal scale of the data, which masks significant variation in conditions within some very large municipalities in the Brazilian Amazon.

2.3.2 Statistical Methods

A key advantage of panel data is the possibility of applying estimation methods that control for unobserved characteristics of the municipalities that might otherwise bias estimates of the effects of observed variables on malaria incidence (Plümper and Troeger, 2007). With my dataset, a key challenge for implementing these methods is the presence of time-invariant variables which can bias the estimates and render them inconsistent. Here, I first define these time-invariant variables, then review three common estimation methods for panel data including whether and how time-invariant variables are treated, and finally describe the modeling approach applied in this chapter.

There are two categories of time-invariant variables. The first is comprised by variables that are permanently time invariant by definition. In the case of this chapter, altitude is such an example. The second category covers variables that are time invariant for the period under analysis. This can be due to data constraints (as is the case of variables such as GDP and % rural population, which are obtained from the census carried out once every 10 years) or due to the rare nature of events (such as creation of new protected areas). By increasing the number of periods and/or the number of cases, it would be possible to render these variables time-variant.

The first standard approach to panel data is to pool them and run an OLS regression, adding the lagged dependent variable plus unit and period dummies to the battery of RHS variables, and calculating "panel corrected standard errors" (Plümper et al, 2005). Pooling

has at least two advantages compared to either a pure cross section of units or a pure time-series. First, it increases the number of observations and consequently the degrees of freedom, making it possible to estimate more fully specified models. Second, pooling makes it possible to control for exogenous shocks common to all observations (by controlling for time effects) and to reduce the omitted variable bias (by controlling for unit effects). In the pooled approach, time-invariant variables are treated the same as other independent variables. The key short-coming of this approach is that it does not recognizge that the observations are usually not independent. There are four potential violations of OLS standard assumptions in pooled panel data analysis (Wooldridge, 2002):

1) errors tend to be autocorrelated (serial correlation of errors) – i.e. they are not independent from one time period to the other;
2) errors tend to be heteroskedastic – i.e. they tend to have different variances across units;
3) errors tend to be correlated across units due to common exogenous shocks (contemporaneous correlation of errors); and
4) errors may be non-spherical in both the serial and the cross-sectional dimension – i.e. autocorrelated and heteroskedastic at the same time.

The second approach is to estimate a Fixed effects (FE) model. There are at least two concerns about the consequences of using a FE specification. First, unit dummies have been accused of eliminating ‘too much’ cross section variance and second, the inclusion of unit dummies makes it impossible to estimate the effect of time invariant exogenous variables and

severely biases the estimate of partly time invariant variables (Huber and Stephens, 2001). Unit dummies control away the deviation of the variable's mean of one unit from the variable's mean of the base unit. Thus, unit dummies completely absorb differences in the level of independent variables across units. As such, unit dummies can be considered as a vector of the effects of time invariant variables plus some parts of the mean of time varying variables.

A third estimation approach is Random effects (RE) modeling. These models allow the individual specific constant terms to be randomly distributed across cross-sectional units, which is appropriate if we believed that sampled cross-sectional units were drawn from a large population. The payoff is that this reduces the number of parameters to be estimated considerably. In general, random effect models are efficient over fixed effects and should be used if the unit specific effects are orthogonal to the other covariates in the model (Huber and Stephens, 2001). However, in the case where regressors are correlated with the unit effects both random effects (RE) models and pooled OLS results are inconsistent and biased. FE also does not allow the estimation of time-invariant variable effects. "Employing these models trades the unbiased estimation of time-varying variables for the ability to compute estimates of the time-invariant variables. Thus, they may be a second-best solution if researchers are solely interested in the coefficients of the time-invariant variables" (Plümper and Troeger, 2007 p.127).

Considering the problems above, it is tempting to consider specifications without unit dummies. However, omitting fixed effects runs the risk of explaining variance in the dependent variable that existed prior to the period under observation with the variance in the mean of the independent variable in the period under observation (Greene, 1993, Wooldridge, 2002). This may be acceptable if the largest effects on the dependent variable come from independent variables that are permanently time-invariant (e.g., geography).

Taking into consideration the problems and advantages of pooling observations across time and space with and without unit dummies into a pooled regression, I adopt the three stage procedure called FEVD (by Plümper and Troeger, 2007) for the estimation of time-invariant and rarely changing variables in panel data models with unit effects. Their procedure allows estimation of the effects of time-invariant variables and is more efficient than FE in estimating variables that have very little longitudinal variance. This method has been adopted by Belke and Spied (2008), Etzo (2008), Krogstrup and Wälti (2008) among others. Later, I compare the results from this model with the standard panel approaches discussed above.

Plümper and Troeger call their approach the "fixed effects vector decomposition" (FEVD) model because the estimator decomposes the unit FE into an unexplained part and a part explained by the time-invariant or rarely changing variables (henceforth called "time invariant variables"). The FEVD technique involves the following three steps:

(i) Estimate the unit FE by running a FE estimate of the "baseline" model (without time-invariant variables);

(ii) Split the unit effects into an explained and an unexplained part by regressing the unit effects on the time-invariant variables;

(iii) Pooled OLS estimation including all explanatory time-variant variables, the time-invariant variables, and the unexplained part of the FE vector, with corrected SEs for the coefficients of the time-invariant variables.

The authors show, based on Monte Carlo simulations, that the FEVD model has better finite sample properties in estimating models that include either time-invariant variables correlated with unit effects than the random effects model, pooled OLS and the Hausman-Taylor (1981) procedure (which depends on instruments to indentify time-invariant variables).

In sum, this method has advantages over the three more standard approaches to analysis of panel data described above. First of all, only the third stage allows obtaining correct standard errors. Not correcting the degrees of freedom leads to a potentially serious underestimation of SEs and overconfidence in the results in comparison to the pooled model. Second, the third stage also allows researchers to explicitly deal with the dynamics of the time-invariant variables.

2.4 Results

We analyze the impact of two key policy variables, roads and protected areas, on malaria incidence, controlling for demographic, climatic, land use and political economy characteristics.

2.4.1 Descriptive statistics

Results are presented in table 2.2. There are a total of about 500 thousand cases of malaria each year for the entire Brazilian Amazon or an average of 39 cases per 1000 inhabitants (on average 847 cases of malaria per municipality or 0.19 per square kilometer). To investigate the drivers of malaria I focus on policy drivers of deforestation that I hypothesize may also directly affect public health, controlling for the more typical causes of malaria suggested in the literature.

A first group of variables linked to malaria in the literature are demographics: population, population density, migration, and percent of population living in rural areas. On average each municipality has 28 thousand inhabitants, at a density of 11 people per km^2, and 51% living in rural areas. The population of the Amazon includes a small fraction of indigenous people, a large "traditional" population (descended from earlier migrants to the region and sometimes of mixed indigenous, African, and European heritage), and more recent migrants who were born in other parts of Brazil but moved to the Amazon in pursuit of economic opportunity. On average there are 25 thousand migrants in each municipality out of which four thousand come from outside the Amazon. In the case of diseases like malaria,

migrants from outside the disease range area are believed to be more prone to infection (Martens and Hall, 2000).

Socio-economic conditions also influence malaria. These variables measure a municipality's wealth, distribution of resources, and health care access. GDP is quite low in the Brazilian Amazon, averaging R$2.85 per year[2] per capita which shows that a large percentage of the population lives at subsistence levels. The average Theil index (a measure of social inequality) for the region is 0.57, which represents an approximate 80:20 distribution similar to Brazil as a whole. Access to health care is measured through the Family Health Program, or PSF. This health program was expanded by the Brazilian federal government in 1994 and now has primary responsibility for the identification and prevention of malaria. In the Amazon the number of families treated is approximately 1 family per km^2 or 0.35 per km^2 in rural areas.

Climate is also considered to play a major role in determining malaria (Hoshen and Morse, 2004; Pascual and Dobson, 2005; Olson et al, 2009). In the Amazon there are two seasons, rainy (Winter) and dry (Summer). The average rainfall in the Winter is 70mm per month, and in Summer the average is 42mm per month. Over a third of municipalities have total annual rainfall over 2300mm. Temperature is an average of 26^oC in Winter and 28^oC in Summer. Altitude also influences climate, although there is limited variation in the altitude of municipal seats which have a mean of just 120m above sea level.

[2] Exchange rate in 2006 was R$2.20:US$1

A growing literature finds that land use influences malaria incidence (Pattanayak and Levins, 2007; Barbieri et al, 2006; Foley et al, 2005) in the form of annual and accumulated deforestation (Castro et al, 2006), mining, cattle pasture, as well as protected areas and roads. In my sample 41km^2 are deforested per year per municipality and the accumulated deforestation is of 860km^2 (or 14% of the municipality area). Thirteen percent of municipalities have mineral extraction activities, as reported in a survey of municipal governments by IBGE. This includes both legal and illegal (or informal) mining activities. There are no finer data related to location of illegal mining (garimpo). Finally, the average number of heads of cattle per square kilometer is 40 (much higher than the population density) but also highly variable.

Turning to the policy variables, the area of a municipality that is within protected areas varies considerably. While some municipalities have no PAs, others are basically inserted into protected areas. This significantly influences and usually limits their options for economic development and land use. On average 14% of a given municipality is within sustainable use PAs, 3% in strict protection PAs and another 9.5% within indigenous reserves. Each municipality has an average of 21 km of official roads and an average of 391 km of unofficial roads per municipality. This significant difference in means demonstrates the importance of unofficial roads, which facilitate economic activities even where there is little or no government presence.

2.4.2 Correlations

Since my interest is to evaluate the effect of the policy variables on malaria incidence, I now present the correlations between the characteristics of interest. Looking at the first column of the table, we also see that most variables are positively correlated with malaria cases, even if this seems unintuitive. From the table we can see that most variables are also significantly correlated. Therefore, these are important reasons to use a multidimensional approach to analyze malaria drivers.

2.4.3 Regression results

To take into consideration how all these variables interact, I implement the Fixed Effect Vector Decomposition (FEVD) procedure to estimate a complete model of malaria determinants and control for potential confounders. This estimator decomposes fixed effects into a part that is explained by time invariant variables and therefore allows for incorporating my policy variables into a fixed effects framework (for more details see appendix 2.2).

The FEVD results are presented in table 2.4. The model fit is quite good (R^2 is 77%). Demographic variables: as expected, population and percentage of population living in rural areas are positively correlated with malaria incidence. Migrants from outside the Amazon increase malaria, a result in line with arguments that populations outside the natural occurrence area of malaria are more susceptible to the disease. Migrants from within the Amazon actually decrease malaria.

Socioeconomic: GDP does not influence malaria, but a more unequal municipality has more malaria. The main government health service to combat malaria (the Family Health Program) is not significant in explaining the disease incidence. As expected the area of the municipality is positively correlated with the number of cases of malaria.

Climate: The results for the climatic variables indicate the seasonal weather patterns influence malaria. Rainfall in the winter and temperature in the winter are correlated with higher incidence of malaria, while rainfall and temperature in the Summer have the opposite effect. As expected, altitude also has a negative correlation with malaria.

Land use: Annual deforestation presents a quadratic relationship with malaria while accumulated deforestation presents a negative effect, in line with the literature about frontier dynamics (Castro et al, 2006; Sawyer and Sawyer, 1992; Sawyer, 1988). Mining and cattle are not significant, suggesting that it is deforestation itself – rather than the replacement land use – that affects health outcomes.

Policy levers: After controlling for deforestation, strict protection PAs have a negative impact on malaria, indicating that areas with pristine forest and without human populations serve as a barrier to malaria. The same occurs for indigenous reserves, mainly due to the natural resistance of its population. The percent of a municipality in sustainable use PAs does not have a significant impact on malaria. Both non-official as well as official roads are positively correlated with malaria, although the significance of non-official roads is higher.

2.4.4 Robustness check

For comparison, I also estimate 4 other models with the same specification: a pooled OLS regression, a random effects (RE) model, a median regression and a Hausman-Taylor model. As we can see in table 2.4 the results for the random effects model are quite similar to the FEVD results. The other models show some differences with the FEVD, especially regarding the significance of the PA variables. I also estimate all the models above including latitude and longitude, as a way to control spatial autocorrelation in appendix 2.3.

2.5 Discussion

Increasingly the popular media and policy-makers assert that human health depends on ecosystem health, but in fact there is only very limited evidence on the health benefits of ecosystem protection and even less on the net effect of conservation policy levers (e.g., protected areas or roads) on human health (9). This chapter enhances current literature by compiling a rich panel dataset on deforestation, malaria incidence, and their drivers in the Brazilian Amazon.

The data cover 4 years and 378 municipalities in the Brazilian Amazon. Variables include demographic, climatic, land use and socioeconomic characteristics, all of which have appeared separately in models of malaria but not previously all together in a complete model. Thus, I have a more fully specified model, allowing me to evaluate their individual effects on malaria with greater confidence. I also have two key policy variables known to influence deforestation represented in finer detail: protected areas divided into strict protection,

sustainable use and indigenous reserves, as well as official and unofficial roads. This allows us to identify differentiated effects of each category of protected areas and roads on malaria incidence, over and above their impact on deforestation.

Our analysis confirms the central role of human behavior as a modifier of the deforestation and disease linkage (10). Significant demographic factors are population, percentage of population living in rural areas as well as migration. Income inequality within a municipality is also positively correlated with malaria. Weather variables (rainfall and temperature) are consistently significant in these estimations, confirming that climate change could have significant impacts on malaria incidence in the region. Finally, malaria is related to incremental deforestation (quadratic) as well as total accumulated deforestation (negative). Roads and protected areas are well known drivers of deforestation, and thus influence malaria through that pathway. However, my estimation results also show that roads directly increase malaria and strict protection PAs and indigenous reserves directly reduce malaria. Thus, there is considerable scope for complementary in conservation and public health policies.

In sum, I find evidence on the "double-(health)-dividends" of conservation policies. If health is an important by product, a portfolio of conservation strategies including control of unofficial roads (instead of just official roads) and promotion of strict protection and indigenous reserves (instead of sustainable use PAs) merit further consideration as Brazil plans further expansion and management of the vast protected area network in the Amazon.

2.6 Conclusion and policy implications

Increasingly, the popular media and policy-makers assert that human health depends on ecosystem health, but in fact there is only very limited evidence on the health benefits of ecosystem protection and even less on the net effect of conservation policy levers (e.g., payments for ecosystem services) on human health. In this chapter I assess the impact of protected areas and roads on malaria incidence in the Brazilian Amazon. To achieve this I use a unique dataset and fixed effects vector decomposition to efficiently estimate coefficients for both time-varying and time-invariant variables. Considering the non-random nature of testing for malaria (being at least partly driven by the level of malaria risk) I presented results in unweighted form and weighted by the ratio of population of the municipality and the total number of malaria tests.

I reach the same general conclusion of many previous policy analysts and epidemiologists: disease is affected by human actions and therefore policy design and implementation should take this into consideration. My results show that sustainable use protected areas increase malaria, while strict protection PAs have no effect. Roads have no effect in the weighted model, which perhaps indicates counter-balancing effects of easier access to the forest (malaria source) and to health care (treatment). Finally, climatic variables are consistently significant in these estimations, confirming that changes in climatic conditions can have significant impacts on malaria incidence in the region.

In sum, I do not find a clear "win-win" or "double-dividend" of conservation policies also reducing malaria. However, there clearly are linkages between public health and protected areas that merit consideration by Brazil's National Program for Protected Areas, as it plans expansion and management of the vast protected area network in the Amazon.

2.7 References

Andam, K.S.A., Ferraro, P.J., Pfaff, A., Sanches-Azofeifa, G.A., Robalino, J.A. 2008. Measuring the effectiveness of protected area networks in reducing deforestation. Proceedings of the National Academy of Sciences 105(42):16091-16094.

Barreto, P., Pinto, A., Brito, B., Hayashi, S. 2008. Quem é Dono da Amazônia: Uma análise do recadastramento de imóveis rurais. Belém, Imazon: 74p.

Bauch, S.C., Amacher, G.A., Merry, F. 2007. Costs of harvesting, transportation and milling in the Brazilian Amazon: Estimation and policy implications. Forest Policy and Economics 9(8):903-915

Beck, N. 2004. Longitudinal (Panel and Time Series Cross Section) Data. Unpublished manuscript available at: http://www.nyu.edu/gsas/dept/politics/faculty/beck/beck_home.html

Beck, N., Katz, J.N. 1995. What to do (and not to do) with time-series cross-section data. The American Political Science Review 89(3) pp.634-647.

Belke, A., Spies, J. 2008. Enlarging the EMU to the east: what effects on trade? Empirica 35(4)

Chape, S., Harrison, J., Spalding, M., Lysenko, I. 2005. Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets. Philosophical Transactions of the Royal Society B: Biological Sciences 360(1454) pp. 443 – 455.

Chen, S., Khan, S. 2008. Semiparametric Estimation of Nonstationary Censored Panel Data Models with Time Varying Factor Loads. Econometric Theory 24:1149-1173.

Chomitz, K.M. 2007. At loggerheads? Agricultural Expansion, Poverty Reduction and Environment in the Tropical Forests. Washington, D.C. International Bank for Reconstruction and Development.

Chomitz, K. Gray, D. 1996. Roads, land use and deforestation: A spatial model applied to Belize. World Bank Economic Review 10:487-512.

Chomitz, K.M., Thomas, T.S. 2003. Determinants of Land Use in Amazônia: A Fine-Scale Spatial Analysis. American Journal of Agricultural Economics v.85(4), pp. 1016- 1028

Cornell, J. 2000. Assessing the role of parks for protecting forest resources using GIS and spatial modeling. In Quantifying Sustainable Development: the Future of Tropical Economies, ed Hall, CAS (Academic, San Diego), pp.543-560.

Cropper, M., Puri, J., Griffiths, C. 2001. Predicting the location of deforestation: the role of roads and protected areas in North Thailand. Land Economics 77(2):172-186.

Deininger, K., Minten, B. 2002. Determinants of deforestation and the economics of protection: An application to Mexico. American Journal of Agricultural Economics 84:943-960.

Etzo, I. 2008. Determinants of Interregional Migration in Italy: A Panel Data Analysis. Available at SSRN: http://ssrn.com/abstract=1135165

FAO. 2006. Global Forest Resources Assessment 2005: Progress Toward Sustainable Forest Management. United Nations Food and Agricultural Organization Forestry Paper 147.

Filmer, D., Hammer, J.S., Pritchett, L.H. 2000. Weak Links in the Chain: a Diagnosis of Health Policy in Poor Countries. The World Bank Research Observer 15(2):199-224.

Frankenberg, E., Thomas, D. 2001. Women's Health and Pregnancy Outcomes: Do Services Make a Difference? Demography 38(2): 253-265.

Gamper-Rabindran, S., Khan, S., Timmins, C. 2008. The impact of piped water provision on infant mortality in Brazil: a quantile panel data approach. NBER Working Paper 14365.

Greene, W.H. 1993. Econometric Analysis, third ed. Macmillan, New York.

Greenwood, B., Bojang, K., Whitty, C., Targett, G. 2005. Malaria. The Lancet 365(9469):1487-1498.

Hays, S.P. 1987. Beauty, health and permanence: environmental policy in the United States, 1955-1985. Cambridge University Press, New York.

Huber,E., Stephens, J.D. 2001. Political choices in global markets: Development and crisis of advanced welfare states. Chicago, IL: Chicago University Press.

Kaimowitz, D., Angelsen, A. 1998. Economic Models of Tropical Deforestation. Bogor, Indonesia: Cifor.

Krogstrup. S., Wälti, S. 2008. Do fiscal rules cause budgetary outcomes? Public Choice 136: 123-138.

Margulis, S., 2004. Causes of Deforestation of the Brazilian Amazon. World Bank Working Paper No. 22. World Bank: Washington, D.C.

Martens, P., Hall, L. 2000. Malaria on the Move: Human Population Movement and Malaria Transmission. Emerging Infectious Diseases v.6(2): 103-109

Mas, J. 2005. Assessing protected area effectiveness using surrounding (buffer) areas environmentally similar to the target area. Environmental Monitoring and Assessment 105:69-80.

Millennium Ecosystem Assessment. 2005. "Ecosystems and Human Well-being: Biodiversity Synthesis." Washington, DC: World Resources Institute.

Morton, D.C., DeFries, R.S., Shimabukuro, Y.E., Anderson, L.O., Arai, E., del Bon Espirito-Santo, Freitas, R., Morisette, J. 2006. Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. Proceedings of the National Academy of Sciences 103(39): 14637-14641

Olson, S.H, Gangnon, R., Elguero, E., Durieux, L., Guégan, J., Foley, J.A., Patz, J.A. 2009. Links Between Climate, Malaria and Wetlands in the Amazon Basin. Emerging Infectious Diseases v.15(4): 659-662

Parkes, M., Panelli, R., Weinstein, P. 2003. Converging paradigms for environmental health theory and practice. Environmental Health Perspectives 111(5): 669-675.

Pattanayak, S.K., Dickinson, K., Corey, C., Sills, E.O., Murray, B.C., Kramer, R. 2006. Deforestation, Malaria and Poverty: a call for transdisciplinary research to design cross-sectoral policies. Sustainability: Science, Practice and Policy 2(2):1-12.

Pattanayak, S.K., Ross, M.T., Depro, B.M., Bauch, S.C., Timmins, C., Wendland, K.J., Alger, K. 2009. Evaluating the health impacts of climate change and conservation policies using applied CGE. B.E. Journal of Economic Analysis and Policy 9(2): 6.

Pattanayak, S.K., Yasuoka, J. 2008. Deforestation and malaria: Revisiting the human ecology perspective. In C.J.P. Colfer (ed.) Forests, People and Health: A Global Interdisciplinary Overview. Earthscan Publishers. Pp. 197-217.

Pfaff, A. 1999. What Drives Deforestation in the Brazilian Amazon. Journal of Environmental Economics and Management 37:26-43.

Pfaff, A., Robalino, J., Walker, R., Aldrich, S., Caldas, M., Reis, E., Perz, S., Bohrer, C., Arima, E., Laurance, W., Kirby, K. 2007. Road Investments, Spatial Spillovers, and Deforestation in the Brazilian Amazon. Journal of Regional Science 47(1):109-123.

Pitt, M.M., Rosenzweig, M.R., Gibbons, D.M. 1993. The Determinants and Consequences of the Placement of Government Programs in Indonesia. The World Bank Economic Review 7(3)(319-348).

Plümper, T., Troeger, V.E., Manow, P. 2005. Panel data analysis in comparative politics: Linking method do theory. European Journal of Political Research 44 pp. 327-354.

Plümper, T., Troeger, V.E. 2007. Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects. Political Analysis 15 pp. 124-139.

Sachs, J., Malaney, P. 2005. The economic and social burden of malaria. Nature 415(6872):680-685.

Schaeffer, R., Rodrigues, R.L.V., Laurance, W.F., Albernaz, A.K.M., Fearnside, P.M., Vasconcelos, H.L., Ferreira, L.V. 2005. Letters: Underlying Causes of Deforestation. Science 307(18 February): 1046-1047.

Simone, M., Garagorry, F. 2005. The Expansion of Agriculture in the Brazilian Amazon. Environmental Conservation. 32(3):203-212.

Smith, K., Corvalán, C., Kjellstrom, T. 1999. How much global ill health is attributable to environmental factors? Epidemiology 10(5):573-584.

Timmins, C. 2003. "Health and Climate Change in the Developing World: The Impact of Global Warming on Brazilian Morbidity Patterns". Working Paper. Yale University.

Turner, R.K., Daily, G.C. 2008. The ecosystem services framework and natural capital conservation. Environmental and Resource Economics 39(1): 25-35.

Van de Walle, D. 1998. Assessing the Welfare Impacts of Public Spending. World Development 26(3): 365-379.

Vincent, J. 1992. The Tropical Timber Trade and Sustainable Development. Science v.256(5064) pp. 1651-1655

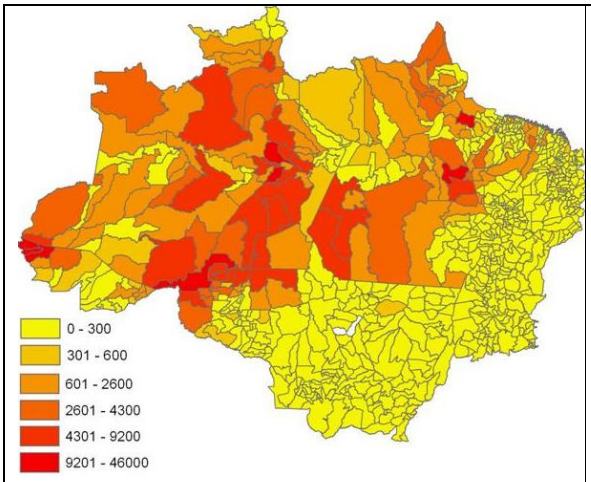
Wang, H., He, X. 2007. Detecting Differential Expressions in GeneChips Microarray Studies: A Quantile Approach. Journal of the American Statistical Association 102(477): 104-112

Wilcox, B., Colwell, R. 2005. Emerging and reemerging infectious diseases: biocomplexity as an interdisciplinary paradigm. EcoHealth 2(4):244-257.

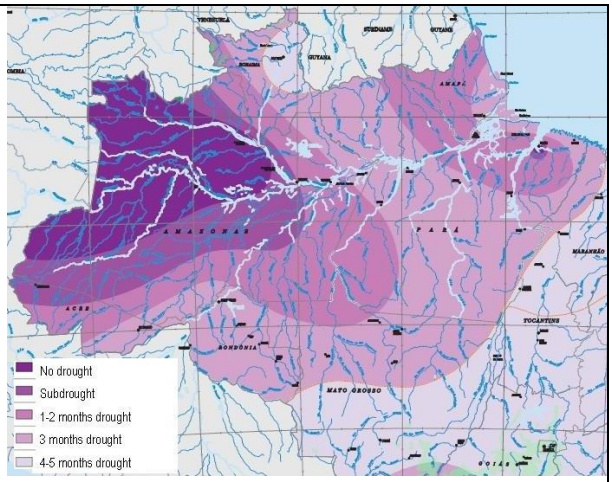
Wilson, M. 2001. Ecology and infectious disease. In J. Aron & J. Patz (Eds.) Ecosystem Change and Public Health. Pp. 285-291. Baltimore, MD: The Johns Hopkins University Press.

Wooldridge, J.M. 2002. Econometric analysis of cross section and panel data. Cambridge, MA: MIT press.

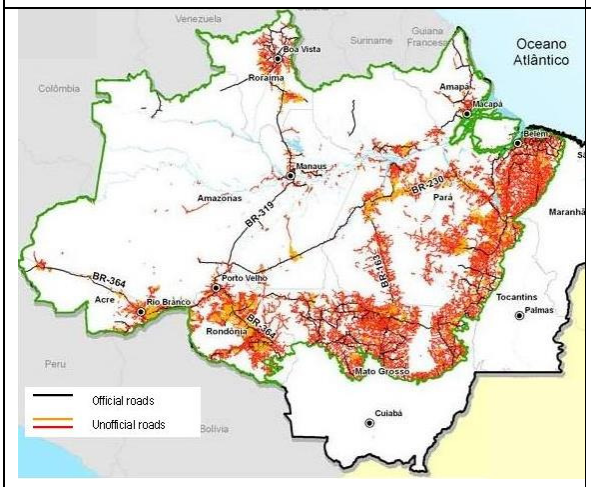
Yasuoka, J., Levins, R. 2007. Impact of Deforestation and Agricultural Development on Anopheline Ecology and Malaria Epidemiology. American Journal of Tropical Medicine 76(3): 450-460.



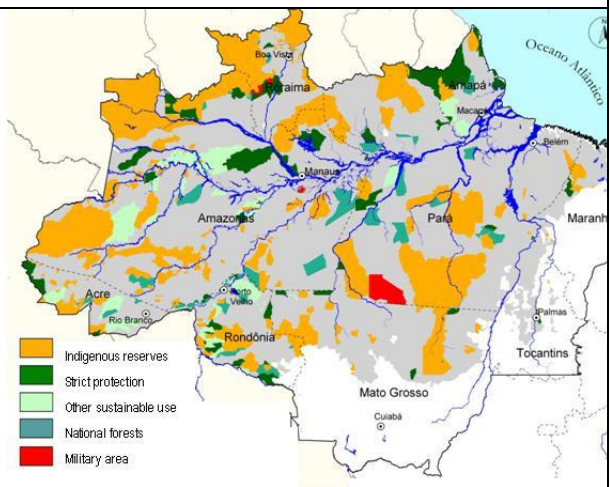
1.A: Malaria (total number of cases in 2006)



1.B: Rainfall (historic averages)



1.C: Official and unofficial road network.



1.D: Protected and indigenous reserves.

Figure 2.1 Malaria, rainfall and policies in the Brazilian Amazon.

Table 2.1 Literature relating deforestation to malaria (N is the sample size, K is the number of covariates).

Author(s)	Deforestation	Other covariates	Data structure	Location	N	K
	Coefficient					
Castro et al, 2006	Estimate is not reported	House characteristics, proximity to forest fringe, education, migration history, protective measure use, economic circumstances	1 settlement over 5 years	Western Brazilian Amazon - Rondônia		
Vittor et al, 2006	(+)***	Proximity to roads and water, human population, land cover	Mosquito landing in 8 sites over a year	Peruvian Amazon	80	8
Vittor et al, 2009	(+)*** for 2ary, (-)** for 1ary	Algae present, inhabitants, distance to 2ary growth	Mosquito counts in visits to water bodies	Peruvian Amazon	2950	12
Barbieri	(+)	Housing quality, sex ratio, age, population	178 localities in 1992 and 1995	Western Brazilian Amazon - Mato Grosso	178	
Pattanayak et al, 2009b	(+)***	Temperature, rainfall, total and rural population, piped water and toilet access, health care, education, GDP, distance to highway	480 microregions over 4 years	Brazil	480	31
Pattanayak et al., 2009a		Ecology and behavioral controls.	Household survey of 500 hh	Indonesia	340	
Pattanayak and Yasuoka, 2008		Ecology and behavioral controls.	cross section of 120 countries	120 countries	120	

Table 2.2 Descriptive statistics for the municipalities in the Brazilian Amazon (N=378). PSF is short for Family Health Program.

Variable type	Definition	Mean	Std Dev	Min	Max
Outcome	Malaria cases	847.54	2951.53	0	45447
Demographic	Population (in 1000s)	27.75	42.38	1.2	381
	Population density	11.23	15.12	0.2	152
	Migrants (in 1000s)	25.25	37.98	1.3	335
	Migrants from outside the Amazon (in 1000s)	3.82	7.27	0	66.24
	% population in rural areas	50.97	20.82	1.7	92
Climatic	Winter rainfall (cm)	6.97	10.08	0	77.7
	Summer rainfall (cm)	4.22	5.87	0	40.7
	Winter temperature (oC)	25.76	1.36	22.75	34.1
	Summer temperature (oC)	27.97	2.47	22.9	34.7
Land use	Area in sustainable use protected area	1153	3727	0	37067
	Area in strict protection protected area	629	3165	0	36477
	Area in indigenous reserves	1408	5687	0	71407
	Km of official roads	21	39.42	0	477
	Km of unofficial roads	391	718	0	7081
	Yearly deforested area	41	108	0	1411
	Accumulated deforestation until 1997	864	960	0	7252
	If has mining	0.13	0.34	0	1
	Heads of cattle (in 10000s)	10.49	14.25	0	159.6
Political economy	GDP per municipality in 2000 (in R$1000)	69.86	160.38	1.67	1434.2
	Theil index in 2000	0.57	0.10	0.29	1.05
	Number of families treated by PSF (in 100s)	18.95	35.23	0	478.9
	Altitude (in 10s of meters)	12.07	11.21	0	92
	Municipality area (in km2)	6220	13739	110.8	159685

Table 2.3 Correlations between variables (*** significant at 1%, ** significant at 5%, * significant at 10%).

	posinf		totpaprot		totpause		tind		roadnonoff		kmrds		migraout		migrants		defors		pops		pctrural	
posinf	1																					
totpaprot	0.14	***	1																			
totpause	0.25	***	0.72	***	1																	
tind	8.44E-02	***	0.49	***	0.40	***	1															
roadnonoff	0.15	***	0.20	***	0.21	***	0.23	***	1													
kmrds	0.35	***	-2.17E-02	*	-7.00E-04		-3.38E-02	***	0.24	***	1											
migraout	0.33	***	2.26E-02	*	3.79E-02	***	9.90E-03		0.29	***	0.50	***	1									
migrants	0.49	***	3.78E-02	***	8.29E-02	***	2.30E-03		7.17E-02	***	0.33	***	0.66	***	1							
defors	0.18	***	0.16	***	0.20	***	0.23	***	0.63	***	7.60E-02	***	0.18	***	4.25E-02	***	1					
pops	0.51	***	3.64E-02	**	7.79E-02	***	-1.00E-04		0.12	***	0.35	***	0.73	***	1.00	***	6.49E-02	***	1			
pctrural	-4.09E-02	**	-4.70E-02	***	1.60E-03		3.86E-02	***	-9.47E-02	***	-0.29	***	-0.36	***	-0.25	***	4.40E-03		-0.27	***	1	
rwint	-5.60E-03		-3.89E-02	***	-6.01E-02	***	-3.51E-02	**	0.15	***	7.72E-02	***	0.10	***	-2.43E-02	*	2.12E-02		-2.39E-02		-0.13	***
rsumm	-9.03E-02	***	-7.45E-02	***	-0.11	***	-7.96E-02	***	0.14	***	0.13	***	9.65E-02	***	-4.82E-02	***	3.05E-02	*	-7.05E-02	***	-0.23	***
twint	-0.21	***	-0.21	***	-0.29	***	-0.19	***	-9.74E-02	***	0.12	***	4.28E-02	***	-7.59E-02	***	-0.16	***	-9.12E-02	***	-0.23	***
tsumm	-0.20	***	-0.19	***	-0.27	***	-0.22	***	-1.80E-02		6.82E-02	***	7.45E-02	***	-0.11	***	-9.22E-02	***	-0.11	***	-0.30	***
deforupto9	0.16	***	6.83E-02	***	0.11	***	0.15	***	0.60	***	0.25	***	0.29	***	0.12	***	0.46	***	0.15	***	4.20E-03	
minext	0.15	***	1.20E-02		-7.40E-03		-5.30E-03		8.82E-02	***	0.16	***	0.13	***	0.11	***	6.97E-02	***	0.11	***	-4.63E-02	***
cattle	4.18E-02	**	4.41E-02	***	4.47E-02	***	0.13	***	0.71	***	0.27	***	0.35	***	3.37E-02	**	0.51	***	5.03E-02	***	-9.53E-02	***
gdp00	0.64	***	3.68E-02	***	6.42E-02	***	-2.70E-03		6.59E-02	***	0.30	***	0.58	***	0.89	***	1.80E-02		0.89	***	-0.19	***
theil00	8.07E-02	***	8.26E-02	***	0.12	***	0.14	***	0.20	***	0.14	***	0.17	***	0.13	***	0.18	***	0.14	***	-0.20	***
psfl	0.54	***	1.20E-03		1.88E-02		-2.38E-02	*	3.29E-02	**	0.33	***	0.56	***	0.81	***	-2.41E-02		0.84	***	-0.27	***
areaima	0.20	***	0.73	***	0.73	***	0.80	***	0.38	***	3.09E-02	**	4.92E-02	***	4.77E-02	***	0.30	***	4.79E-02	***	1.76E-02	
altitude10	-0.11	***	-8.62E-02	***	-0.14	***	-3.97E-02	***	0.17	***	0.11	***	0.11	***	-9.48E-02	***	9.82E-02	***	-9.58E-02	***	-0.27	***

Table 2.3 Continued

	rwint		rsumm		twint		tsumm		deforupto97s		minext		cattle		gdp00		theil00		psfl		areaima	
posinf																						
totpaprot																						
totpause																						
tind																						
roadnonoff																						
kmrds																						
migraout																						
migrants																						
defors																						
pops																						
pctrural																						
rwint	1																					
rsumm	0.49	***	1																			
twint	0.28	***	0.47	***	1																	
tsumm	0.34	***	0.48	***	0.79	***	1															
deforupto9	-8.40E-03		-0.10	***	-0.27	***	-0.20	***	1													
minext	-4.90E-03		-1.35E-02		-3.11E-02	**	-5.44E-02	***	0.10	***	1											
cattle	0.15	***	0.14	***	1.81E-02		0.14	***	0.65	***	8.26E-02	***	1									
gdp00	6.20E-03		-2.80E-03		-4.64E-02	***	-4.82E-02	***	8.46E-02	***	0.12	***	2.34E-02		1							
theil00	4.54E-02	***	5.03E-02	***	3.01E-02	**	9.46E-02	***	0.14	***	2.89E-02	**	0.20	***	0.11	***	1					
psfl	-1.30E-03		-9.50E-03		9.30E-03		-2.65E-02	*	5.12E-02	***	9.20E-02	***	2.33E-02		0.80	***	0.12	***	1			
areaima	-2.62E-02	*	-8.07E-02	***	-0.29	***	-0.28	***	0.29	***	7.00E-03		0.19	***	3.59E-02	***	0.17	***	1.20E-03		1	
altitude10	0.29	***	0.42	***	0.43	***	0.58	***	-1.34E-02		-9.00E-03		0.25	***	-3.01E-02	**	0.15	***	-2.28E-02	*	-7.26E-02	***

Table 2.4 Results of third stage of FEVD and robustness checks of log of malaria cases (*** significant at 1%, ** significant at 5%, * significant at 10%) N=1379.

Type of variables	Variables	FEVD		RE		Pooled OLS		HT		Median	
		estimate	std dev	estimate	std dev	estimate	std dev	estimate	std dev	estimate	
	Intercept	3474.23 ***	1093.38	6193.91 ***	1384.53	12815.73 ***	1897.66	2.08 **	1.06		
Demographic	pops	3.03E-02 ***	1.05E-02	2.68E-02 **	1.12E-02	7.61E-02 ***	2.09E-02			8.62E-02	***
	pctrural	16.93 ***	2.38	13.29 ***	4.96	5.69	5.08	2.06E-02 ***	7.21E-03	7.93	*
	migrants	-3.31E-02 ***	1.13E-02	-2.90E-02 **	1.28E-02	-8.58E-02 ***	2.24E-02	3.66E-07	8.27E-06	-0.10	***
	migraout	5.52E-02 ***	9.34E-03	7.00E-02 ***	2.19E-02	4.65E-02 **	2.09E-02	5.71E-05 **	2.90E-05	3.10E-02	*
Climatic	twint	101.57 **	44.98	135.39 ***	52.42	258.63 ***	74.80			306.11	***
	tsumm	-159.52 ***	33.29	-288.39 ***	38.84	-636.89 ***	47.59			-737.08	***
	rwint	15.54 ***	3.54	16.05 ***	2.71	28.18 ***	5.26			23.74	***
	rsumm	-33.51 ***	8.38	-27.79 ***	5.06	-56.51 ***	8.24			-49.11	***
	altitude10	-68.36 ***	4.86	-47.02 ***	9.04	-32.12 ***	9.76	-8.88E-02 ***	1.90E-02	-30.95	***
Land use	defors	4.29 ***	1.00	1.58 *	0.94	10.44 ***	1.87			11.66	***
	defors2	-5.03E-03 ***	1.08E-03	-2.21E-03 **	9.17E-04	-1.26E-02 ***	2.18E-03			-1.64E-02	***
	deforupto9	-0.31 ***	7.88E-02	-0.16	0.15	-0.41 **	0.17	-1.56E-04	2.03E-04	-0.59	***
	minext	145.32	143.68	185.10	285.77	-173.10	310.79	0.46	0.41	-102.90	
	cattle	8.50E-04	6.20E-04	-1.04E-03	8.06E-04	3.40E-03 **	1.41E-03			5.35E-03	***
	totpause	2.74E-02	2.44E-02	3.03E-02	5.94E-02	7.20E-02	6.16E-02	1.76E-05	1.02E-04	8.00E-02	
	totpaprot	-0.12 ***	3.01E-02	-0.12 **	6.61E-02	-6.47E-02	6.75E-02	-1.33E-04	1.20E-04	-1.95E-02	
	tind	-0.11 ***	1.85E-02	-0.10 **	4.49E-02	-4.14E-02	4.36E-02	-1.27E-04	1.08E-04	-2.58E-02	
	kmrds	2.64 *	1.54	2.64	3.13	4.12	3.28	2.09E-03	5.77E-03	6.22	*
	roadnonoff	0.91 ***	0.12	1.06 ***	0.26	0.72 ***	0.26	9.31E-04 *	5.80E-04	0.90	***
Political	GDP00	2.00E-04	6.60E-04	1.98E-04	1.19E-03	-5.00E-04	1.26E-03	4.90E-07	1.79E-06	-1.27E-03	
economy	theil00	2136.56 ***	429.32	2215.79 **	925.86	2490.19 ***	855.87	1.69	1.45	2355.16	***
	psfl	-2.00E-02	1.66E-02	-0.05116 **	2.19E-02	-1.19E-02	3.50E-02			-3.59E-02	
	areaima	8.68E-02 ***	1.22E-02	7.90E-02 **	3.38E-02	2.16E-02	3.08E-02	1.03E-04 *	5.51E-05	7.39E-02	*
	res2	606.94 ***	30.95								

APPENDICES

Appendix 2.A

Protected Areas in Brazil

The National Nature Conservation Unit System (SNUC) was instituted in Brazil on July 18th 2000 through Law #9985 and regulates protected areas at the municipal, state and federal levels. The SNUC objectives, according to the law, are as follows:

- to contribute to the maintenance of biological diversity, genetic resources in national territory and jurisdictional waters;
- to protect species threatened by extinction at regional and national levels;
- to contribute to the preservation and restoration of natural ecosystem diversity;
- to promote sustainable development through natural resources;
- to promote the use of nature conservation principles and practices in the development process;
- to protect natural and little altered landscapes of notable scenic beauty;
- to protect the geological, geomorphologic, speleological, paleontological and cultural characteristics of nature;
- to protect and restore hydrological and edaphic resources;
- to recover or restore degraded ecosystems;
- to promote means and incentives for scientific research activities, studies and environmental monitoring;
- to value biological diversity economically and socially;

- to favor conditions and promote environmental education and interpretation, recreation in contact with nature and ecological tourism;
- protect natural resources necessary to the subsistence of traditional populations, respecting and valuing their knowledge and their culture and promoting them economically.

The consolidation of the SNUC aims at *in situ* conservation of biological diversity on the long run, centrally to a fundamentally conservationist process. It establishes the necessary condition of complementarity between the different categories of conservation units, organized according to their management objectives and land uses into Integral Protection and Sustainable Use. The basic objective of Integral Protection Areas is to preserve[3] of nature, with only indirect use[4] of its natural resources being permitted, except for cases described in the SNUC law. The basic objective of Sustainable Use Protected Areas is to make nature conservation compatible with the sustainable use of parts of its natural resources.

Integral Protection Areas

This group contains the following categories of protected areas:

---

[3] Preservation aims at keeping a resource in its natural state, without human interference. Opposed to conservation, which aims at keeping the resource through sustainable use.

[4] Indirect use it that which does not involve consumption, collection, damage or destruction of natural resources while direct use is that which involves collection and use, commercial or not, of natural resources.

Ecological Station: has the objective of preserving nature and scientific research. Public visitation is prohibited except with an educational objective and scientific research depends on previous authorization by responsible entity. The only ecosystem alterations allowed in these areas are: measured that aim at restoring modified ecosystems, species management to preserve biological diversity and collection of ecosystem components for scientific ends or scientific research which have a greater impact than the observation or controlled collection of ecosystem components in an area up to 3% of its extention or up to 1500 hectares.

Biological Reserve: has the objective of promoting the integral preservation of the biota and other existing natural attributes within its limits, without direct human interference or environmental modifications, except restoration measures of its altered ecosystems and the management actions necessary to restore and preserve the natural equilibrium, biological diversity and natural ecological processes. Public visitation is prohibited except with an educational objective and scientific research depends on previous authorization by responsible entity.

National Park: has the basic objective of preserving natural ecosystems of large ecological relevance and scenic beauty, allowing for scientific research and development of environmental education and interpretation, recreation in contact with nature and ecological tourism. Public visitation is subject to norms and restrictions established in the

unit's Management Plan, to norms established by entity responsible for its management and those in law. Scientific research depends on previous authorization by entity responsible for the unit's management. Protected units in this category created by the State or Municipality are respectively called State Park and Municipal Natural Park.

Natural Monument: has the basic objective of preserving rare, unique or natural sites of large scenic beauty. This type of conservation unit can be constituted by private lands, as long as the land and natural resource uses by the owners are compatible with the objectives of the conservation unit. Public visitation is subject to conditions and restrictions established in the area's Management Plan, to the norms established by entity responsible for its management and those in law.

Wildlife Refuge: has as objective to protect natural environments where conditions for the existence or reproduction of species or communities of local flora and resident or migratory fauna are ensured. This type of conservation unit can be constituted by private or public lands, as long as the owner accepts the land use restrictions proposed by the entity responsible for the unit's management. Public visitation is subject to conditions and restrictions established in the area's Management Plan, to the norms established by entity responsible for its management and those in law. Scientific research depends on previous authorization by entity responsible for the unit's management.

Sustainable Use Protection Areas

This group contains the following categories of protected areas:

Environmental Protection Area: is in general an extensive area, with a certain degree of human occupation, with abiotic, biotic, aesthetic or cultural attributes especially important for quality of life and wellbeing of human populations, and has the basic objectives of protecting biological diversity, discipline the occupation process and ensure the sustainability of natural resource use. These units are constituted by public or private lands.

Area of Relevant Ecological Interest: Is an area in general of small extension with little or no human occupation, with extraordinary natural characteristics or that shelters rare species of the regional biota and has the objective to maintain the natural ecosystems of regional or local importance and regulate admissible use of these areas, as a way to make it compatible with the objectives of nature conservation.

National Forest: is an area with forest cover of predominantly native species and has the basic objective of sustainable multiple use of the forest resources and scientific research, with emphasis on sustainable management methods of native forests. National Forests are of public domain and private areas within its boundaries should be dispossessed. This category allows the permanence of traditional populations which inhabited the area when

of its creation in conformity with the unit's management plan. Public visitation is permitted, conditional on norms established for the unit's management by the entity responsible for its administration. Research is allowed and incentivized, under previous authorization by the entity responsible. The National Forest has a Consultative Council, presided by the entity responsible for its management and constituted by representatives of public entities, civil society organizations and, if pertinent, resident populations. Units in this category, when created by the State or Municipality will be called respectively State Forest and Municipal Forest.

Extractive Reserve: is an area used by traditional extractive populations whose subsistence is based on extraction of natural resources and, complementarily by subsistence agriculture and raising small animals and has the basic objective of protecting livelihoods and culture of these populations and ensure the sustainable use of natural resources in the unit. The Extractive Reserve is of public domain with use conceded to the extractive traditional populations and private areas within its boundaries should be dispossessed. The extractive reserve will be managed by a Deliberative Council presided by the entity responsible for its management and constituted by representatives of public entities, civil society organizations and resident traditional populations. Public visitation is permitted, as long as compatible with the local interests and according to the area's Management Plan. Scientific Research is permitted and incentivized, pending authorization by the entity responsible for the areas management. The unit's management

plan will be approved by its Deliberative Council. Exploration of mineral resources and professional or amateur hunting is prohibited. The commercial management of its timber resources will only be allowed on sustainable basis in special situation and complementary to the other activities developed in the area.

Fauna Reserve: is a natural area with terrestrial or aquatic native animal populations, resident or migratory, adequate for technical-scientific studies about the economic sustainable management of fauna resources. Public visitation might be permitted, as long as compatible with the unit's management and according to norms established by the entity responsible for its management. Professional or amateur hunting is prohibited. The commercialization of products and byproducts resulting from research are regulated under fauna laws.

Sustainable Development Reserve: is a natural area that shelters traditional populations, whose existence is based on systems of sustainable harvest of natural resources, developed through generations and adapted to the local ecological conditions and that play a fundamental role in nature protection and in the maintenance of biological diversity. Its basic objective is to preserve nature and, at the same time, ensure conditions and means necessary for the reproduction and improvement of livelihoods and life quality and management of natural resources of the traditional populations, as well as value, conserve and improve the knowledge and techniques of environmental

management developed by these populations. The use of areas occupied by traditional populations will be regulated according to specific legislation. The unit will be managed by a Deliberative Council, presided by the entity responsible for its management and constituted by representatives of public entities, civil society organizations and resident traditional populations. Activities carried out in the reserve will follow the conditions:

a) Public visitation is allowed and permitted as long as compatible with local interests and according to the area's management plan.

b) Scientific research aimed at nature conservation, at better relations between resident populations with their environment, and environmental education is permitted and incentivized subject to prior authorization by entity responsible for the area's management.

c) The dynamic equilibrium between the population and conservation.

d) Sustainable management of natural ecosystem components and the substitution of land cover for cultivated species is allowed as long as subject to the zoning, legal limits and the area's management plan.

The area's management plan will determine the areas of integral protection, sustainable use, ecological corridors and mitigation areas and will be approved by the unit's Deliberative Council.

Private Reserve of Natural Patrimony: is a private area, recorded for perpetuity, with the objective of conserving biological diversity. The only activities allowed are: scientific

research, visits with touristic, educational and recreational objectives, natural resource extraction, except timber, that does not put species or ecosystems in risk and justify the creation of the unit.

Appendix 2.B

The FEVD model

Plümper and Troeger (2007) call their approach the "fixed effects vector decomposition" (FEVD) model because the estimator decomposes the unit FE into an unexplained part and a part explained by the time-invariant or rarely changing variables (henceforth called "time invariant variables"). The FEVD technique involves the following three steps:

1) Estimate the unit FE by running a FE estimate of the "baseline" model (without time-invariant variables);
2) Split the unit effects into an explained and an unexplained part by regressing the unit effects on the time-invariant variables;
3) Pooled OLS estimation including all explanatory time-variant variables, the time-invariant variables, and the unexplained part of the FE vector, with corrected SEs for the coefficients of the time-invariant variables.

The authors show, based on Monte Carlo simulations, that the FEVD model has better finite sample properties in estimating models that include either time-invariant variables correlated with unit effects than the random effects model, pooled OLS and the Hausmann-Taylor procedure (which depends on instruments to indentify time-invariant variables). Below I present the derivation of the FEVD procedure (directly from Plümper and Troeger, 2007):

First, I present the data generating process of a FE model with time-invariant variables:

(i) $$y_i = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + u_i + \varepsilon_{it}$$

where the x variables are time-varying, the z variables are time invariant, u_i denotes the N-1 unit-specific effects (FE), ε_{it} is the independent and identically distributed error term, α is the intercept of the base unit, and β and γ are the parameters to be estimated.

In the first stage, the FEVD procedure estimates a standard FE model. The FE transformation can be obtained by first averaging equation (i) over T:

(ii) $$\overline{y_i} = \alpha + \sum_{k=1}^{K} \beta_k \overline{x_{ki}} + \sum_{m=1}^{M} \gamma_m z_{mi} + \bar{e}_i + u_i$$

where: $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, $\bar{e}_i = \frac{1}{T}\sum_{t=1}^{T} e_{it}$ and e stands for the residual of the estimated model. Then equation (ii) is subtracted from equation (i). This transformation removes the individual effects u_i and the time-invariant variables z. I get:

(iii) $$\begin{aligned} y_{it} - \bar{y}_i &= \beta_k \sum_{k=1}^{K} (x_{kit} - \bar{x}_{ki}) + \gamma_m \sum_{m=1}^{M} (z_{mi} - z_{mi}) + (e_{it} - \bar{e}_i) + (u_i - u_i) \\ &\equiv \hat{y}_{it} = \beta_k \sum_{k=1}^{K} \hat{x}_{kit} + \hat{e}_{it} \end{aligned}$$

With $\hat{y}_{it} = y_{it} - \bar{y}_i, \hat{x}_{it} = x_{kit} - \bar{x}_{ki}, \hat{e}_{it} = e_{it} - \bar{e}_i$ denoting the demeaned variables of the within transformation. I run this FE model to obtain estimates of the unit effects $\hat{u}_i$. At

this point, it is important to note that the "estimated unit effects" $\hat{u}_i$ do not equal the unit effects u_i in the data generating process. Rather, these estimated unit effects include all time-invariant variables, the overall constant term, and the mean effects of the time-varying variables x – i.e.

(iv) $$\hat{u}_i = \bar{y}_i - \sum_{k=1}^{K} \beta_k^{FE} \bar{x}_{ki} - \bar{e}_i$$

where β_k^{FE} is the estimate of the demeaned variables in equation (iii). This $\hat{u}_i$ is calculated by subtracting the predicted value and residual from the mean value of y. Therefore, $\hat{u}_i$ includes the unobserved unit-specific effects as well as the observed unit specific effects z, the unit means of the residuals $\bar{e}_i$ and the time varying variables $\bar{x}_{ki}$, whereas u_i only accounts for unobservable unit specific effects. In stage 2, I regress the unit effects $\hat{u}_i$ from stage 1 on the observed time-invariant and rarely changing variables – the z variables to obtain the unexplained part h_i (the residual from regressing the unit-specific effect on the z variables). In other words, I decompose the estimated unit effects into two parts, an explained and an unexplained part that I call h_i:

(v) $$\hat{u}_i = \sum_{m=1}^{M} \gamma_m z_{mi} + h_i$$

The unexplained part h_i is obtained by computing the residuals from equation (v):

$$h_i = \hat{u}_i - \sum_{m=1}^{M} \gamma_m z_{mi}$$

In sum, this crucial stage decomposes the unit effects into an unexplained part and a part explained by the time-invariant variables. I retain the unexplained part h_i for the next stage of estimation.

In stage 3, I rerun the full model without the unit effects but include the unexplained part of h_i of the decomposed unit FE vector obtained in stage 2. This stage is estimated by pooled OLS:

$$y_{it} = \alpha + \sum_{k-1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + \delta h_i + \varepsilon_{it}$$

By design, h_i is no longer correlated with the vector of the z variables. If the time-invariant variables are assumed to be orthogonal to the unobserved unit effects – i.e. if the assumption underlying this estimator is correct – this estimator is consistent. If this assumption is violated, the estimated coefficients for the time-invariant variables are biased, but this bias is just the normal omitted variable bias.

In sum, this method has advantages over the three more standard approaches to analysis of panel data described above. First of all, only the third stage allows obtaining correct standard errors. Not correcting the degrees of freedom leads to a potentially serious underestimation of SEs and overconfidence in the results in comparison to the

pooled model. Second, the third stage also allows researchers to explicitly deal with the dynamics of the time-invariant variables.

To illustrate how different public policies can affect malaria, I estimate the Plümper and Troeger (2007) procedure as described below:

In the first stage, I estimate the fixed effects model in equation 3:

$$\hat{y}_{it} = \beta_k \sum_{k=1}^{K} \hat{x}_{kit} + \hat{e}_{it}$$

where $\hat{y}_{it}$ are changes in malaria incidence (changes in the number of positive malaria cases reported for each municipality), $\hat{x}_{kit}$ are changes in variables I have annual data for and $\hat{e}_{it}$ is the residual term. β are the parameters to be estimates. In my application, $\hat{x}_{kit}$ s are annual changes in Winter (TWINT) and Summer temperature (TSUMM), Summer rainfall (RSUMM), annual deforestation (DEFORSA and DEFORSA2), heads of cattle/km^2 (CATTLEA), percentage of municipality covered by agriculture (AGRIC) and lagged number of families treated by the Family Health Program in the rural region of the municipality/km^2 (PSFRAL)[5]. Considering the heterogeneity in the size of municipalities, I normalize variables by the area of the municipality.

[5] I used the lagged variable for treatment by the Family Health Program for 2 reasons: 1) the focus on preventative care of malaria and 2) to avoid endogeneity of treating more people due to a malaria outbreak.

After estimating the fixed effects model, I calculate the individual effects (equation iv). In the second stage:

$$\hat{u}_i = \sum_{m=1}^{M} \gamma_m z_{mi} + h_i$$

where $\hat{u}_i$ represent the calculated unit effects, z_{mi} are the time invariant variables in my data and h_i are the residuals of this equation (the portion of the unit effects not explained by my model). In my specification z_{mi} represent population density (POPDENS), population growth/km^2 in the 1991-2000 period (POPGROWA), % of population living in rural areas (PCTRURAL), Theil index (measure of economic inequality) in the municipality in 2000 (THEIL00), kilometers of official roads/km^2 (KMRDSA), kilometers of unofficial roads/km^2 in 1997 (ROADNONOFFA), accumulated deforestation (% area deforested) in 2002 (ACUMDEFOR), total number of migrants/ km^2 (MIGRANTSA), number of migrants from outside the Amazon/km^2 (MIGROUTA), % of municipality within sustainable use PA (TOTPAUSEA), % of municipality within strict protection PA (TOTPAPROTA), % of municipality within indigenous reserves (TINDA), if there is mineral extraction in the municipality (legal and/or illegal) (MINEXT), altitude/10 (ALTITUDE10), latitude (LAT98) and longitude (LONG98). Latitude and longitude are included to control for spatial autocorrelation. Considering the heterogeneity in the size of municipalities, I normalize all variables by the area of the municipality.

Finally, in the third stage, I estimate the pooled regression of malaria (POSINFA) as a function of all variables used in the first and second stages plus the residual from the second stage. This residual represent the fixed effects for the municipality.

I estimate both unweighted and weighted version of this model. The dependent variable in both cases is the number of positive tests for malaria per square kilometer of the municipality. However, testing for malaria is not random and is at least partly driven by the level of malaria risk - and indeed, the positive number of tests is related to the total number of tests carried out in a given municipality. To address this, I employ weights equal to the population of the municipality divided by the total number of malaria tests. This gives more weight to malaria cases in municipalities with large populations but few tests.

Appendix 2.C

FEVD results "controlling" for spatial correlation

The inclusion of latitude and longitude into the model allows for some control over the spatial autocorrelation present in the data. Thus, here I present and discuss these results.

Table 2.C.1 Results of first stage of Plümper and Troeger (2007) – fixed effects model of cases of the difference in the log of malaria cases per km^2 unweighted and weighted by population/ total number of malaria tests (*** significant at 1%, ** significant at 5%, * significant at 10%).

		Unweighted regression			Weighted regression		
		R^2 = 2%		F stat = 4.47***	R^2 = 26%		F stat = 71.18***
Description	Variable	Estimate		Std error	Estimate		Std error
	Intercept	-2.64E-02		2.48E-02	1.31	***	4.35E-02
Annual deforestation	ddeforsa	-3.12E-03		4.57E-02	0.83	***	0.12
Annual deforestation squared	ddeforsa2	-2.66E-03		3.78E-03	-0.10	***	1.92E-02
Heads of cattle/km2	dcattlea	-1.09E-02	**	4.41E-03	-8.25E-02	***	1.94E-02
Temperature in Winter	dtwint	-3.00E-02		6.27E-02	-0.68	***	0.12
Temperature in Summer	dtsumm	0.22	***	5.15E-02	8.44E-02		0.11
Rainfall in Summer	drsumm	-1.76E-03		3.13E-03	-3.76E-02	***	2.46E-03
Number of families treated by PSF in rural areas	dpsfral	-0.12		8.98E-02	0.29		0.58

Table 2.C.1 presents the results of the first stage (fixed effects) of the Plümper and Troeger (2007) model for both the weighted and unweighted regressions. There are stark differences between the weighted and unweighted models. Because the weighted model fits the data much better (as evidenced by its R^2 and F statistic), I focus on those results. Annual deforestation shows a significant quadratic relation with malaria incidence. As such, areas with more deforestation have more malaria. Cattle is association with lower incidence of malaria, which might be due to fact that they also serve as potential hosts for the mosquito, leading to less human disease. In the other hand, area in agriculture increases malaria incidence indicating how the mosquito is sensitive not only to the major land change (deforestation) but also to subsequent land uses in line with findings by Yasuoka and Levins (2007). In terms of climatic variables, temperature in the winter is negatively correlated with malaria (meaning that higher temperatures in the wetter months are associated with less disease) and more rainfall is associated with less malaria, a finding that is expected in the Amazon, in line with other studies (i.e. Olson et al, 2009). Finally, the number of families treated by the main health program in charge of malaria prevention and treatment (PSF) is not significantly related to the incidence of the disease.

As these estimates are based on the traditional Fixed Effects model, and taking into consideration the discussion on its limitations I should be careful in assessing these

estimates. For better inference I now discuss the results for the full third stage specification in table 2.C.2:

Table 2.C.2 Results of third stage of Plümper and Troeger – Clustered OLS of log of malaria cases per km^2, unweighted and weighted by population/ total number of malaria tests (*** significant at 1%, ** significant at 5%, * significant at 10%)

Definition	Parameter	Unweighted model R^2 = 68%		Weighted model R^2=58%	
		Estimate	Std error	Estimate	Std error
	Intercept	-6.46 ***	1.96	4.68	5.31
% of area within sustainable use PA	totpausea	1.52 ***	0.33	1.99 **	0.91
% of area within strict protection PA	totpaprota	-8.25E-03	0.79	-0.95	1.79
% of area in indigenous reserves	tinda	-0.91 ***	0.32	0.74	0.69
Km of official roads	kmrdsa	11.02 ***	4.20	1.31	16.30
Km of unofficial roads	roadnonoffa	3.90 ***	0.74	1.84	2.10
Annual deforestation	deforsa	0.29 **	0.13	-5.60E-02	0.27
Annual deforestation squared	deforsa2	-4.66E-02 *	2.63E-02	8.90E-03	4.70E-02
Accumulated deforestation	acumdeforam	2.00E-03	3.56E-03	1.50E-03	8.13E-03
If there is mineral extraction	minext	0.86 ***	0.22	0.69 *	0.36
Heads of cattle/km2	cattlea	-3.64E-05	2.96E-03	-6.01E-04	5.29E-03
Total number of migrants	migrantsa	0.45 *	0.27	-2.29 *	1.32
Total number of migrants from outside the Amazon	migraouta	4.06E-02	3.05E-02	0.13 *	7.62E-02
% of population living in rural areas	pctrural	7.85E-03 **	3.58E-03	-3.83E-03	8.01E-03
Population density	popdens	-0.44	0.27	2.30 *	1.32
Population growth/km2	popgrowa	8.23E-04	1.50E-02	-2.64E-02	5.15E-02
Theil index	theil00	0.68	0.53	-0.15	0.94
Number of families treated by PSF in rural areas	psfral	2.36E-02	0.11	0.49	0.34
Temperature in Winter	twint	4.64E-02	6.58E-02	0.20	0.16
Temperature in Summer	tsumm	-1.02E-02	5.33E-02	-0.26 **	0.10
Rainfall in Summer	rsumm	-5.07E-03	4.73E-03	-2.61E-02 **	1.25E-02
Altitude	altitude10	-3.07E-02 ***	6.40E-03	1.62E-02 **	7.74E-03
Latitude	lat98	0.31 ***	2.84E-02	0.14 *	7.84E-02
Longitude	long98	0.17 ***	1.42E-02	-2.27E-02	3.91E-02
Residual from 2nd stage	res2	0.45 ***	3.75E-02	-0.23 ***	7.99E-02

In this table I again present the weighted and unweighted models of log of malaria as a function of policy variables, biological and socioeconomic covariates. As a reminder, this stage is estimated as a pooled OLS clustered by municipality (since I have one observation for each municipality each year). With respect to the results of these models, the fit of the unweighted regression is better here than in the first stage (R^2=68%) however, both models have a very good fit (weighted model has a R^2=58%). Therefore I will base the discussion on the variables which have significance in both models with some intuition of why the results for both models sometimes differs. First of all, sustainable use protected areas are associated with higher levels of malaria. This is expected as this type of protected area allows for communities to live within its boundaries, which means that humans are directly in contact with the disturbed forest environment. In the other hand, strict protection PAs do not have a significant relation with malaria incidence. This is also explained since populations are not allowed to live within its boundaries. Indigenous reserves are only significant in the unweighted model since this model does not take into account the population in each municipality and the fact that indigenous reserves have less overall tests than other areas (due to natural resistance of the indigenous population to malaria symptoms). Roads are highly significant in the unweighted model but not in the weighted model again due to these factors being a driver of the total number of malaria tests being carried out (which is controlled for using weights). Annual deforestation is also only significant in the unweighted model. Mineral extraction is a major driver of migration into forest areas and

therefore it is positively associated to malaria in both models. Total number of migrants has a different effect in each of the models. In the weighted model, where I control for population in the municipality, migrants are negatively associated with malaria, which might indicate a higher level of education or social capital that these migrants have compared to the local population while migrants from areas outside the Amazon (non-malarial areas) are positively related to malaria incidence, as expected. Finally, population density is also positively associated to malaria incidence in the weighted model, which is expected from an epidemiological point of view, since the disease transmission is facilitated. Several environmental variables are also significant. First of all, temperature in the Summer is negatively related to malaria incidence, which might indicate climate as a limiting factor for the disease vector. Second, rainfall is also negatively associated to the disease, as in other studies on malaria in the Amazon (Olson et al, 2009). Finally, altitude has differing effects in each model. In the unweighted regression the result is negative while the weighted regression results in a positive estimate. Considering the weight of population/ total number of tests and the fact that higher altitude areas are negatively correlated with total number of tests and population, this result might indicate that the weighted regression is putting more weight into this variable that is not necessarily associated to it but might be associated to other characteristics highly correlated with altitude. Finally, latitude is significant, showing that spatial autocorrelation is present and that the further South in the Amazon, the more malaria, which is consistent with the spatial distribution of the deforestation frontier. The

residuals from the second stage in the model are included and represent the individual fixed effects not explained by the time invariant variables in the model, this is always significant.

The results pertaining the effect of sustainable use protected areas might lead us to think that they are inherently bad. However, to show that they indeed have a positive impact in decreasing malaria linked to deforestation directly, I rerun the model without the deforestation variables:

Table 2.C.3 Results of third stage of Plümper and Troeger – Clustered OLS, weighted and unweighted models of log of malaria cases omitting deforestation (*** significant at 1%, ** significant at 5%, * significant at 10%).

Definition	Variable	Unweighted model R^2 = 65%		Weighted model R^2 = 58%	
		Estimate	Std Error	Estimate	Std Error
	Intercept	-5.01 ***	1.84	5.57	5.38
% of area within sustainable use PA	totpausea	1.06 ***	0.29	2.05 **	0.88
% of area within strict protection PA	totpaprota	-4.59 ***	0.75	-0.92	1.71
% of area in indigenous reserves	tinda	0.39	0.34	0.89	0.74
Km of official roads	kmrdsa	-7.24	4.71	6.94	18.67
Km of unofficial roads	roadnonof	4.62 ***	0.80	2.20	2.06
If there is mineral extraction	minext	-0.16	0.23	0.67 *	0.35
Heads of cattle/km2	cattlea	2.35E-03	2.55E-03	8.47E-04	4.66E-03
Total number of migrants	migrantsa	-2.01 ***	0.31	-2.61 *	1.52
Total number of migrants from outside the Amazon	migraouta	0.17 ***	3.14E-02	0.12	7.38E-02
% of population living in rural areas	pctrural	2.57E-03	3.51E-03	-5.05E-03	8.37E-03
Population density	popdens	2.01 ***	0.31	2.63 *	1.52
Population growth/km2	popgrowa	7.70E-02 ***	1.92E-02	-2.60E-02	4.45E-02
Theil index	theil00	3.60 ***	0.60	-0.35	0.88
Number of families treated by PSF in rural areas	psfral	5.73E-02	0.12	0.44	0.37
Temperature in Winter	twint	-4.70E-02	6.79E-02	0.18	0.17
Temperature in Summer	tsumm	6.29E-02	5.33E-02	-0.26 **	0.10
Rainfall in Summer	rsumm	-1.34E-02 *	7.10E-03	-2.54E-02 **	1.20E-02
Altitude	altitude10	-9.04E-03	6.92E-03	1.65E-02 **	7.78E-03
Latitude	lat98	0.37 ***	2.87E-02	0.15 **	7.23E-02
Longitude	long98	0.12 ***	1.25E-02	-2.45E-02	3.79E-02
Residual from 2nd stage	res2	0.50 ***	3.79E-02	-0.24 ***	8.41E-02

From the results in table 2.C.3 we can see that the coefficients on sustainable use protected areas continues to be significant but have decreased in their size. In the other hand, strict protection PAs become significant and negative, as if these areas actually

decrease the incidence of this disease. As such, we can see that omitting deforestation from the model actually shows that the PA variables absorb some of the mitigation effect. Therefore, it is important to consider that while protected areas have negative impacts regarding malaria incidence they are still providing services by decreasing deforestation.

Appendix 2.D

Modeling deforestation

To inform the discussion about the drivers of malaria, I estimated a model of deforestation as a function of policy, demographic, social and climatic variables. Results of the pooled OLS with clustered standard errors are presented in the table below:

Table 2.D.1 Annual deforestation as a function of policy variables and covariates.

Description	Variable	Estimate		Std. Err.
Accumulated deforestation in 1997	deforupto97s	-1.14E-02		7.42E-03
Km official roads	kmrds	0.24		0.17
km unofficial roads	roadnonoff	0.11	***	1.97E-02
Indigenous reserves	tind	3.14E-03		2.72E-03
Sustainable use PA	totpause	3.13E-03		2.68E-03
Strict protection PA	totpaprot	-5.78E-03	**	2.74E-03
Migrants	migrants	-1.63E-04		1.55E-04
Migrants from outside the Amazon	migraout	3.79E-04		8.72E-04
If has mineral extraction	minext	-2.99		10.34
Theil index	theil00	32.84		39.54
% of population living in rural areas	pctrural	0.29	*	0.17
Altitude (in 10s of meters)	altitude10	-0.19		0.24
Area of municipality	areaima	-9.42E-05		1.59E-03
Rainfall in Winter	rwint	-0.16		0.10
Rainfall in Summer	rsumm	-0.50	*	0.26
Temperature in Winter	twint	-4.41	**	1.80
Temperature in Summer	tsumm	2.32	*	1.22
Intercept	_cons	22.26		54.56

As we can see from the table above non-official roads are the most robust driver of deforestation where each extra km of road leads to 0.11 km^2 of deforestation. Strict protection PAs decrease annual deforestation, in line with the fact that only indirect use is allowed within its borders. These results are robust to different specifications of the

model. Finally, having a higher percentage of the population living in rural areas increases deforestation, while climatic variables are also significant in the model.

Appendix 2.E

Clustered Quantile Regression models for malaria

In this appendix[6] I first estimate a clustered quantile regression model (weighted by municipality area[7]) as described below:

Log(posinfa) = F(totpausea, totpaprota, tinda, kmrdsa, roadnonoffa, deforsa, acumdefora, acumdefora2, migrantsa, migraouta, cattlea, agriculture, psfral, popdens, popgrowa, pctrural, theil00, rsumm, tsumm, twint, altitude10)

Where POSINFA is the number of positive malaria cases reported for each municipality; TOTPAUSEA is percent of the municipality within sustainable use protected areas; TOTPAPROTA is percent of the municipality within strict protection protected areas; TINDA is the percent of the municipality within indigenous reserves; KMRDSA is number of kilometers of paved official roads; ROADNONOFFA is the number of kilometers of unofficial roads; DEFORSA is the deforestation for any given year; ACUMDEFORA is the accumulated deforestation up to that year; ACUMDEFORA2 is the square of accumulated deforestation; MIGRANTS is the total number of migrants in the municipality population; MIGRAOUT is the number of migrants from outside the Amazon states; CATTLEA is the number of heads of cattle; agriculture is the percent of

[6] These results were presented at EAERE 2009.
[7] By weighted by municipality area I mean that each observation is divided by the area of the municipality (resulting in per km^2 variables)

municipality area used for growing crops; PSFRAL is the lagged number of families treated by the Family Health Program in the rural region of the municipality; POPDENS is population density; popgrowa is population growth in the 1991-2000 period; PCTRURAL is percent of municipality population living in rural areas; THEIL00 is the Theil Index (measure of economic inequality) in 2000; RSUMM is average monthly rainfall in the Summer; TSUMM is average monthly temperature in the Summerl, TWINT is average monthly temperature in Winter, ALTITUDE10 is the municipality's altitude (in 10s of meters) and POP000 is the municipality's population (in 10s of inhabitants).

I also estimate the clustered OLS regression for comparison, as well as the reduced form model of malaria as a function of policies (without deforestation) to evaluate how the estimated coefficients change. Finally, several robustness checks as specified, such as the estimation of the same model according to the time series cross section estimator proposed by Plümper and Troeger (2007), and inclusion of spatial correlation in both the quantile regression and TSCS models. The inclusion of spatial correlation allows us to correct for effects that are similar across space, while the TSCS model allows us to control for the endogenous placement of policies. Plümper and Troeger's model allows for the incorporation of time invariant (or, rarely changing) variables into a fixed effects model.

Results and analysis

Table 2.E.1 below presents the results from the clustered quantile regression of the full model of malaria as a function of policies (and deforestation), controlling for ecological and population characteristics. For comparison purposes I also present the results from a clustered OLS estimation for the same model.

Table 2.E.1 Results of the clustered quantile regression for the log positive cases of malaria in the 2002-2006 period, weighted by municipality area (*** significant at 1%, ** significant at 5%, * significant at 10%).

	clustered		quantile= 0.25		quantile= 0.5		quantile= 0.75	
Variables	OLS		Coefficients		Coefficients		Coefficients	
totpausea	1.41	**	2.47	**	1.91	***	1.53	**
totpaprota	-6.10E-02		-0.53		0.81		0.32	
tinda	-1.11	**	-0.15		-1.54	***	-2.42	***
kmrds	5.93E-03	***	7.89		7.55		7.84	
roadnonoffa	4.14	***	4.40	***	5.02	***	3.95	***
deforsa	0.16	*	1.92E-02		0.23		0.29	**
acumdefora	2.90E-03		-4.04E-03		2.20E-03		2.24E-02	
acumdefora2	-7.06E-05		1.02E-04		-8.17E-05		-2.59E-04	
migrantsa	0.36		0.79		0.61		0.17	
migraouta	0.13	**	1.22E-02		-1.90E-02		-5.20E-02	
cattle	4.20E-03		8.66E-04		2.40E-03		-5.25E-03	
agriculture	-2.64	**	-1.51		-1.87	*	-3.94	**
psfral	-4.25E-02		-5.85E-02		-8.29E-03		-0.30	**
popdens	-0.36		-0.90		-0.73	*	-0.37	
popgrowa	1.09E-02		9.90E-03		-2.46E-02		2.11E-02	
pctrural	1.05E-02	*	8.34E-03	**	4.67E-03		1.34E-02	**
theil00	-0.21		-0.65		-0.88		0.41	
rsumm	-3.18E-02	***	-3.25E-02	***	-3.09E-02	***	-2.16E-02	*
tsumm	-0.40	***	-0.41	***	-0.60	***	-0.51	***
twint	-0.18		-0.18	**	-5.32E-02		-7.40E-02	
altitude10	-4.80E-02	***	-3.19E-02	***	-5.97E-02	***	-5.17E-02	***

As is known, the median regression is less influenced by extreme observations than the mean, and this shows in the results. We can see that in the case of some variables the OLS and quantile results are quite similar (as with the climatic variables) but sometimes one method finds significance while the other does not. As we expected, sustainable use PAs have a positive effect on malaria (meaning that they increase the number of malaria cases) while strict protection PAs have no significant effect on the disease. Also as expected, indigenous reserves decrease malaria as local populations are resistant to the pathogen. Official roads do not help predict deforestation, and we think this is due to these being older frontier regions, further from where deforestation is occurring. Deforestation is quite significant at the .75 quantile, while in the OLS is it only significant at the 10% level. Percent of land in agriculture reduced deforestation, as does the number of families treated by the Family Health Program (PSF) in rural areas (at the higher quantile) and population density (at the median). The percent of population living in rural areas increases malaria incidence, also as expected. Finally, altitude, temperature and rainfall are all significant and have the expected signs.

From this table we might be tempted to infer that protected areas are bad, in that they "cause" malaria, however, PAs influence malaria primarily through the deforestation channel. Table 2.E.2, helps elucidate this matter, by showing the results of the reduced form estimation of the same model presented in table 2. The expected results, is that, in omitting deforestation the effects of protected areas will be dampened, and this is exactly

what happens: sustainable use PA estimate becomes smaller (still positive) while integral protection PA estimate becomes negative and significant (insignificant before). As such, in the case of sustainable use PAs there is an off-setting direct effect of increasing malaria by increasing human interaction with disturbed forest.

Table 2.E.2 Results of the reduced form model of clustered quantile regression for the sum of positive cases of malaria in the 2002-2006 period, weighted by municipality area (without deforestation) (*** significant at 1%, ** significant at 5%, * significant at 10%).

Variables	Clustered		quantile= 0.25		quantile= 0.5		quantile= 0.75	
	OLS		Coefficients		Coefficients		Coefficients	
totpausea	0.83	**	0.56		0.95	**	0.90	*
totpaprota	-0.86		-1.04		-0.88		-0.56	
tinda	-1.47	***	-0.95	**	-1.36	***	-2.36	***
kmrds	0.00	**	2.32		4.82		5.53	
roadnonoff	3.24	***	1.34		1.91	***	4.66	***
migrantsa	1.14E-03		-0.15		-0.13		3.65E-02	
migraouta	4.68E-02		3.40E-02		1.66E-02		-2.23E-02	
cattle	3.30E-03		2.74E-03	*	1.37E-03		3.07E-03	
agriculture	-2.10	*	-1.54	**	-1.70	*	-1.56	**
psfral	-0.25	***	-0.21	***	-0.44	***	-0.32	***
popdens	3.52E-03		0.13		8.81E-02		-0.18	
popgrowa	1.42E-02		9.01E-04		2.25E-02		1.81E-02	**
pctrural	6.71E-03		2.54E-03		4.86E-03	*	7.70E-03	
theil00	0.30		0.52		0.57		0.36	
rsumm	-1.80E-02	***	-1.65E-02	***	-1.88E-02	***	-1.91E-02	***
tsumm	-0.42	***	-0.42	***	-0.57	***	-0.62	***
twint	-6.48E-02		8.78E-02		0.12	*	5.04E-02	
altitude10	-2.98E-02	***	-1.87E-02	***	-2.37E-02	***	-3.20E-02	***

To check the robustness of my estimates to spatial correlation, we include latitude and longitude of municipality centroids into the analysis (results in table E.3)[8]. From this table we can see that the effects of protected areas remain quite similar to the ones in table 2 (sustainable use PAs have a positive and significant effect while strict protection PAs have no significant effect on malaria). In this model official roads have a significant and positive effect (at lower quantiles). Accumulated deforestation also becomes significant, showing a quadratic relationship with malaria (meaning that municipalities with very high or very low deforestation have less malaria). The main difference in these results compared to table 2.E.1 refers to the climatic conditions, which become less significant.

[8] Other specifications were tested including higher order terms of latitude and longitude, as well as their interactions and produced similar results.

Table 2.E.3 Results of the clustered quantile regression for the log positive cases of malaria in the 2002-2006 period, weighted by municipality area, accounting for spatial correlation (*** significant at 1%, ** significant at 5%, * significant at 10%).

	Clustered		quantile=	0.25	quantile=	0.5	quantile=	0.75
Variables	OLS		Coefficients		Coefficients		Coefficients	
totpausea	1.54	***	2.03		2.39	***	1.83	***
totpaprota	0.07		1.36		-0.54		0.94	
tinda	-1.05	*	-2.10E-02		-0.87	**	-1.24	
kmrds	6.30E-03	***	10.38	**	6.99	**	3.38	
roadnonoff	3.63	***	3.96	***	4.11	***	3.48	***
deforsa	0.23	***	0.10		0.22	**	0.30	***
acumdefor	2.77E-02	**	1.09E-02		1.20E-02		2.53E-02	***
acumdefor	-2.72E-04	**	-1.22E-04		-2.10E-04	*	-3.46E-04	***
migrantsa	0.48		0.67		0.25		0.17	
migraouta	6.55E-02		9.34E-02	**	0.10		-2.85E-02	
cattle	2.36E-04		1.47E-03		-1.62E-04		-1.07E-02	*
agriculture	-2.04	*	-1.48		-2.56	*	-0.36	***
psfral	0.10		-1.71E-02		-8.01E-03		-0.18	
popdens	-0.47		-1.16		-0.97	*	8.31E-04	
popgrowa	2.19E-03		-1.07E-02		5.29E-03		-2.02	
pctrural	8.48E-03		2.62E-03		4.25E-03		-4.85E-02	
theil00	0.27		-0.16		-0.62		-0.37	
rsumm	-1.19E-02	**	-1.42E-02	**	-1.77E-02	**	1.45E-02	
tsumm	-3.56E-02		-0.32	***	-0.38	***	1.29E-02	***
twint	-0.11		6.11E-02		-0.13		0.79	
altitude10	-3.07E-02	***	-2.55E-02	***	-3.65E-02	***	-3.56E-02	***
latitude	0.28	***	0.27	***	0.24	***	0.25	***
longitude	0.14	***	0.15	***	0.16	***	0.14	***

3 Revisiting ICDPs: the Impact of Community Based Enterprises in the Brazilian Amazon

3.1 Abstract

As the debate around Reduced Emissions from Deforestation and Degradation (REDD) as a viable option to curb carbon emissions intensifies, Integrated Conservation and Development Projects (ICDPs) have seen renewed interest as an alternative that could tackle a double objective: reduce deforestation while also addressing poverty issues. In this chapter I evaluate the effectiveness of community based enterprises (CBEs) as parts of ICDPs to address welfare of indigenous communities in the Tapajós National Forest in the Brazilian Amazon. Although ICDPs have been implemented since the 1980s, no rigorous evaluation of their effectiveness has been carried out and thus my contribution is quite policy relevant, as this level of scrutinity is necessary if we are to allocate scarce resources efficiently (Morduch, 1999; Morduch, 2000; Ferraro and Pattanayak, 2006). I use a rich household panel (pre and post intervention) of 216 households in 22 communities to compare outcomes at the community and household level. I start with non-parametric comparisons between communities that received CBEs and those without, then compare communities with differing percentages of household participation. From these comparisons we can see that, as participation in CBEs increases, cash income tends to increase. Cattle tends to decrease as participation increases, which indicates a shift in

household land use decisions in line with the ICDP objective. However, asset wealth indices tend to decline with participation, indicating that ownership of assets has not improved. Finally, I estimate propensity score and panel based fixed-effect estimators (difference in difference) to compare households that were participating in the CBEs 2006 with those that never participated and also all households that have ever participated with those that have not. Matching is necessary as program placement and participation was not random. We see the same trends observed at the community level, but interestingly results are stronger for the larger group of households having ever participated in the CBEs. This indicates that there might be greater benefits to drop out of the program before funding ends and responsibilities for management are pushed to the community. Overall, the analysis indicates that although ICDPs can have positive livelihood impacts through the implementation of CBEs, the sustainability of such actions need further scrutiny.

Keywords: ICDPs, program evaluation, Amazon, DID, Propensity Score, microenterprises

3.2 Introduction

As interest in mitigating climate change through Reduced Emissions from Deforestation and Degradation (REDD) intensifies, it becomes increasingly critical to learn the lessons of previous efforts to conserve tropical forest. One common approach

likely to be recycled in the context of REDD is Integrated Conservation and Development Projects (ICDPs) (Brandon and Wells 2009). This is because in addition to its benefits as a potentially significant, inexpensive, and quick way to reduce carbon emissions, REDD is expected to bring co-benefits for both biodiversity conservation (e.g., supporting the expansion and implementation of protected area networks) and local livelihoods (e.g., compensating traditional populations who conserve forest). There are high expectations for REDD to generate a win-win through large financial transfers and better governance that will benefit the rural poor in developing countries (Peskett and Yanda, 2009). International negotiations over REDD hinge on questions such as: how it can be used as a poverty alleviation mechanism? Or, at least, how can REDD be designed to avoid undermining livelihoods? Proposals range from direct payments to landowners to keep the forest standing to government controlled protected areas. ICDPs are somewhere in the middle of the spectrum between direct payments and government control and have thus received significant attention as a model and base of experience for designing REDD (Angelsen, 2008). Rigorous evaluations of the effectiveness of the ICDPs already in place are required to inform this debate but are not found in the literature. In this chapter, I offer such a rigorous evaluation of a $1.4 million ICDP that implemented community-based enterprises to generate income from sustainable management of the Tapajós National Forest in the Brazilian Amazon. I examine impacts on both the welfare and land use of participants versus non-participants.

ICDPs have been implemented since the mid-1980s (Hughes and Flintan, 2001) but their definition has changed considerably over time. However, most practitioners would agree that ICDPs are biodiversity conservation projects with rural development components (Hughes and Flint, 2001). A broader definition is that by Worah (2000), which describes ICDPs as "an approach that aims to meet social development priorities and conservation goals". Stricter definitions argue that ICDPs are projects implemented in or around protected areas (Barrett and Arcese, 1995). As such, the objective of ICDPs is to enhance livelihoods of communities living within a natural resource rich area while promoting sustainable use and conservation at the same time.

Despite having been implemented for over two decades, conclusions about their effectiveness are, at best, not convincing. As Wells and McShane (2004 p.518) put it: "Whether ICDPs have worked or can work remain complex questions to answer. It is even difficult to assess what has been achieved." Literature reviews on the topic find that success cases are mostly fleeting and fragile (Garnett et al, 2007); that decentralization improves ecological, economic, attitudinal and behavioral measures of success but that market integration and utilization had positive effects only on behavior and economic indicators (Brooks et al, 2006) and that quality of forests is correlated best with funding of enforcement and not at all with the presence of ICDPs (Struhsaker et al, 2005). Several other authors have found that these project struggle to meet both their conservation and

their development goals (McShane and Wells, 2004; Robinson and Redford, 2004; Sayer and Campbell, 2004; Wells et al, 2004, Browder, 2002, Wells and Brandon, 1992).

My opinion is that uncertainty surrounding ICDP effectiveness is rooted in the lack of rigorous evaluation of these projects. Ferraro and Pattanayak (2006) point to the lack of program evaluations of conservation investments, including ICDPs, and highlight the importance of identifying the counterfactual to a project: what would have happened if this intervention had not been implemented? This is the question I answer regarding the community based enterprises financed by the Forest Resource Management Project (ProManejo) initiative in the Tapajós National Forest (TNF) in the Brazilian Amazon. The projects funded in the TNF covered a wide range of initiatives from health to income generation by selling forest products. As is typical in most ICDPs, these projects were targeted to specific communities (selected by ProManejo staff and community leaders) and households within these communities chose whether or not to participate. Therefore I need to control for this non-random selection into the program as I evaluate its impacts on both household welfare and household resource use. This chapter proceeds by discussing the study site and ProManejo in section 2, followed by a discussion of the data in section 3, empirical strategy and econometric issues in section 4, and results in section 5. Section 6 concludes.

3.2.1 Study site

The Tapajós National Forest was established in 1974 and encompasses an area of approximately 545 thousand hectares in four municipalities: Aveiro, Belterra, Rurópolis and Placas, in the west of the state of Pará (see figure 3.1 for map). The northern limit is the urban area of the municipality of Belterra; the eastern border is the highway BR163 that links the cities of Cuiabá and Santarém; to the west is the Tapajós River and to the south the Tinga and Cupari Rivers.

As of 2004, there were approximately 1100 families that live in the five inhabited zones defined by the National Forest's management plan completed in 2004 (Cordeiro, 2004). Table 3.1 describes the land use in these different zones. The riverine and Chibé zones contain 26 communities (in 2006). Most of these families (or their ancestors) already lived in the area before the creation of the National Forest, but there has also been some in-migration. The other zones are a municipality seat (Aveiro) which is more population dense, and some isolated private lots along the BR163 road (distributed by the government in 1972 and 1973). The resident population of the National Forest lives primarily from fishing, hunting, cultivation of manioc, corn, rice and beans, raising animals and collection of non-timber forest products (Cordeiro, 2004). However, the five zones differ in terms of the origin and socio-economic characteristics of residents and in terms of institutional support for development of forest-based sustainable income sources

under ProManejo. This support has been almost exclusively concentrated in the riverine population area.

My study focuses on the riverine area, Chibé, and two similar communities located just outside the northern boundary of the forest (see community location in figure 3.1). Figure 3.2 shows a satellite image of the Tapajós National Forest and surrounding areas. It is evident from this image that there is less deforestation within this protected area than outside. The locations of the communities are indicated by yellow dots. As suggested by the name, the 25 communities in the riverine zone are located along the river, while Chibé is located further inland (close to the large riverine community of Tauari). Since the creation of the National Forest, there have been many conflicts between the communities and the agency that administers the National Forest (IBAMA - Brazilian Institute of Environment and Renewable Natural Resources), as shown in the timeline of events presented in Appendix 3.1.

3.2.1.1 Agriculture in the National Forest

Households in the Tapajós National Forest cultivate mostly manioc but also some other crops such as squash, sweet manioc, corn, rice and beans. Due to the poor quality of most soils, an area is usually cultivated for a period of 2 to 4 years, followed by a fallow period that lasts between 4 and 8 years (Cordeiro, 2004). Households can then reopen land from this secondary growth or open primary forest. Primary forest is preferred due

to the higher quality of organic matter and nutrients, but there is a tradeoff due to access difficulty to these areas.

Since 2001, IBAMA has emitted deforestation authorizations of up to three hectares to traditional resident families that live basically from subsistence slash-and-burn agriculture. Due to increased concern about deforestation, starting in 2003, IBAMA has held meetings with communities prior to emitting those authorizations in order to define collective norms of land use. Officially, IBAMA authorizes up to two hectares of deforestation in secondary growth (capoeira) and up to one hectare in primary forest, while land clearing for pasture is forbidden.

3.2.1.2 ProManejo

The Amazon Sustainable Forest Management Support Project (ProManejo) was a component of the PPG7 (Pilot Program for Tropical Forest Protection), financed by the World Bank, DFID, GTZ, the Brazilian government and the German development bank, KfW. The objective was to finance pilot programs and demonstrations to support the development and adoption of sustainable forest management systems in the Amazon (Dias, 2001). It was implemented by IBAMA and the Environmental Ministry (Biodiversity and Forests division/ National Forests Program). ProManejo had four sub-components: i) strategic studies and analysis; ii) promising initiatives; iii) monitoring and control; and iv) management of the Tapajós National Forest (as a demonstration project

for all sustainable use protected areas in the Amazon) (IBAMA, 2009). The fourth sub-component, the focus of this chapter, aimed to support the conservation of the TNF through natural resource management and community development, with the participation of traditional populations from the region, improving livelihoods in these communities (IBAMA, 2009). Table 3.2 describes the projects supported by Component IV of ProManejo, including their objectives, number of benefitted communities and total funding received between 1999 and 2006.

Among the projects in the TNF supported by ProManejo were community based enterprises (CBEs) designed to generate cash income from non-timber forest products and services: eco-tourism; natural oils, wood artisan products, and "ecological leather" made from natural rubber. (The ecotourism CBE received money through the "promising initiatives" component of ProManejo, while the others are described in the shaded rows of Table 3.2.) The CBE's objectives were to complement community-based forest management and to increase and diversify household income. Some of the CBEs were supported by local NGOs who assisted community associations, while in other cases community associations had full responsibility. In all cases, ProManejo provided start-up capital as well as technical training and support in commercialization. Each CBE was given an initial term of 5 years to become sustainable on its own. These terms were

extended[9] and lasted the full 7 years ProManejo invested in the region. In total, ProManejo invested about $1.4 million directly in the communities of the TNF.

3.3 Data

The data for this chapter come from administrative records of ProManejo and survey data on a panel of households interviewed in 1997 and 2006.

In 1997, the USDA Forest Service funded a study to explore the potential for ecotourism in the region, including the first survey of 322 households in 18 communities located in the "riverine" zone of the National Forest. Coincidently, two years after this questionnaire was carried out, ProManejo started its activities in the National Forest. As such, the first round of surveys provides a good baseline to analyze the impacts of the CBEs.

The second wave of the panel survey was implemented in 2006, as the funding and activities of ProManejo in the Tapajós were ending. In this wave, 355 households located in the same communities visited in 1997 (now divided into 22 communities) were interviewed, resulting in a panel of 216 households interviewed in both the first and second rounds[10]. Figure 3.3 provides a timeline of ProManejo and the surveys.

---

[9] These were no-cost extensions necessary for the conclusion of activities and spending present in the approved proposals, but no additional funds were invested.

Of the 106 households lost from the panel, 56% correspond to households moving out of the communities (to cities or communities outside the survey area), 26% were households not available to be interviewed (mostly because they were traveling), and 7% no longer existed due to death of heads of household and no remaining children. In the remaining 11% of cases, the households could not be clearly matched, or the reason for attrition could not be clearly established. In the full sample of 355 households interviewed in 2006, 35% participated in the CBEs, while in the panel 33% participate in the CBEs.

Table 3.3 provides descriptive statistics for variables that characterize the households, shape their decisions whether to participate in the CBEs, and define livelihood and environmental outcomes of interest. Significant changes have occurred in the National Forest in the nine years between 1997 and 2006. For example, the number of children (less than 10 years old) has decreased, while the proportion of women in the household has increased, although the size of the household has remained the same. Education of the household heads has increased, as has the number of assets a household owns. The number of children living outside of the National Forest has increased, but the number of remittances to family has decreased. The hectares of cropland per household have decreased, while the number of poultry and cattle has remained steady. The number of milking cows has decreased though as has the number of hogs. The latter is generally

[10] Same households were defined as: i) same heads of household; ii) surviving head of household; iii) same physical location if child was living there; or iv) household of oldest child.

attributed to IBAMA's campaign to decrease the number of hogs due to concerns regarding their roaming (destroying other people's plantations and yards).

In terms of forest income, the percentage of households selling non-timber forest products (NTFPs) also decreased (from 64% to 22%), as did the number of products collected from the forest. On the other hand, rubber tapping increased though, probably as a result of better access to markets. The presence of roads to some communities – resulting in more commercial establishments in those communities - might also explain why families were making fewer trips to Santarém in 2006 compared to 1997.

Cash income from a variety of other sources increased between 1997 and 2006. This includes retirement pensions and other government transfers (virtually nonexistent in 1997), as well as wage labor. Approximately a third of households participated in one of the CBEs evaluated in this chapter. Mean household cash income in 2006 was about R$2624 per year[11].

As we can see from table 3.3 there are several changes that occurred in the National Forest in the 1997-2006 period, which I need to address if I am to identify the impact of the CBEs. In the next section I outline the empirical and econometric strategy

[11] In 1997 R$ (corrected for inflation).

used to address endogeneity and confounding issues in program participation and outcomes.

3.4 Empirical strategy

The key feature that allows program evaluation is that some units get the program and others do not (Ravallion, 2001). In observational studies, the investigator does not have experimental control over which units get the program (Rosenbaum, 2002). This is the case with the Tapajós, and thus I do not have the option of randomizing participation and controlling for differences between participants and non-participants (or controls) experimentally. Thus, my starting point for the evaluation of ProManejo is dealing with differences between the participants and controls.

Program evaluation analyzes outcomes of a specific program relative to one or more counterfactuals (i.e. what would have happened without the intervention). In the case of ProManejo, the four income-generating CBEs were each implemented in some communities but not in others. Within communities that received the program, some households participated and others did not. Therefore, data are collected on an outcome indicator Y_i for unit i in a sample of size n. In the case of ProManejo, Y_i can be the income, assets, health status, or education by household. For further explanation of the methods I will consider a household as the unit that received the program or not (this is readily translated into communities receiving programs or not by substituting the

subscript i for j in the explanations to follow). Some of the households receive the program (D_i=1) and others do not (D_i=0). The value of Y_i is denoted Y_{1i} (for treated) if D_i=1 or Y_{01} (for counterfactual) if D_i=0. There is also information available on control variables (or covariates), X_i. (From this point on I will drop the subscript i to avoid clutter).

If the program were implemented as a random experiment, each observational unit would have the same probability of participating in the program and therefore the group of participants (D=1) and the group of controls (D=0) would be identical *ex-ante* (Ravallion, 2001). This would eliminate bias in comparing the groups and all that would be needed to non-parametrically estimate the Average Treatment Effect would be: ATE = $E(G) = Y^T - Y^C$. This would facilitate the inference of program impacts, but such program design is not common in practice, due to political sensitivities and social ethics. It is also important to note that with randomized assignment we only know mean outcomes for the counterfactual, so we cannot infer the joint distribution of outcomes (Ravallion, 2001).

Considering that the programs of interest in the TNF were not randomly placed, I need to consider alternative methods. According to Heckman and Navarro (2004), when participation is voluntary (as is the case of ProManejo), agents decide to participate or not by utility maximization. Thus, let utility be denoted V:

(i) Max $V = f_v(Z,U_v)$

where Z are factors (observed by the analyst) determining choices, U_v are unobserved (by the analyst) factors determining choice. Then, potential outcomes are written in terms of observed variables (X) and unobserved (by the analyst) outcome-specific variables:

(ii) $Y_1 = f_1(X,U_1)$

(iii) $Y_o = f_o(X,U_0)$

And I assume that U_0, U_1, U_v are continuous random variables with finite means. The individual-level treatment effect is then defined as: $\Delta = Y_1 - Y_o$. The traditional means conditional on covariates used in program evaluation are:

(iv) Average Treatment Effect: ATE: $E(Y_1-Y_o|X)$

(v) Treatment on the Treated: TT: $E(Y_1-Y_o|X,D=1)$

(vi) Marginal Treatment Effect: MTE: $E(Y_1-Y_o|X,Z,V=0)$

Following I briefly describe the two program evaluation techniques I use in this chapter: matching and panel based difference-in-difference.

3.4.1 **Matching** (Based on Heckman and Navarro, 2004)

The method of matching discussed here was developed by Rosenbaum and Rubin (1983) and predecessors and, as usually formulated, makes no distinction between X and

Z. Let's define the conditioning set as W=(X,Z). The strong form of matching advocated by these authors assumes that:

(vii) $(Y_1, Y_0) \perp\!\!\!\perp D|W$ and

(viii) $0 < Pr(D=1|W) = P(W) < 1$

where $\perp\!\!\!\perp$ denoted independence given the conditioning variables after |. P(W), the propensity score, is the probability that the observational unit will receive treatment conditional on the available covariates (Heckman and Navarro, 2004). If I do not use the propensity score, as in covariate matching, a balancing problem arises due to the necessity to address deviations from one or another covariate. On the other hand, Propensity Score Matching (PSM) is created to address the balancing problem you can have when matching on multiple dimensions.

Condition (viii) implies that the mean treatment parameters can be defined for all values of W (no common support problem). Rosenbaum and Rubin show that under conditions (vii) and (viii):

(ix) $(Y_1, Y_0) \perp\!\!\!\perp D|P(W)$

Which reduces the dimensionality of the matching problem. Assuming P is known and conditioning on P eliminates ATE and TT bias because:

$E(Y_1|D=0,P(W)) = E(Y_1|D=1,P(W)) = E(Y_1|P(W))$

$E(Y_0|D=1,P(W)) = E(Y_0|D=0,P(W)) = E(Y_0|P(W))$

Observe that because ATE=TT for all X,Z under conditions (xii) and (xiii), the effect for the average person participating in the program is the same as the effect for the marginal person, conditional on W. This is an unattractive implication of these assumptions, in which all dependence between U_v and (U_1,U_0) is eliminated by conditioning on W:

$U_v \perp\!\!\!\perp (U_1,U_0)|W$

This is the origin of the term "selection on observables" introduced by Heckman and Robb (1985, 1986).

If I do not want to impose condition (vii) I can replace it by the weaker alternative:

(x) $E(Y_1|W,D=1) = E(Y_1|W)$ and $E(Y_0|W,D=0) = E(Y_0|W)$

However, then I have to assume that I can replace W with P(W), since this does not follow from the weaker assumption (x) above.

Also, assumption (viii) implies that, if the researcher has too much information about the decision of who decides to participate, so that P(W)=1 or =0 the method breaks down because similar people cannot be found in the D=1 and D=0 groups. Thus, matching assumes away the possibility of perfect prediction, which can be potentially problematic. This potential problem can be solved if I introduce the distinction between Z and X, if there are some variables Z not in X.

In terms of how to find a good match for a given household, the three most common method are nearest neighbor, kernel and radius. Here is a brief explanation of each matching option:

- Nearest neighbor: Each treated observation is matched to the control with the most similar propensity score value. This is the most commonly used method in literature. Also, it is necessary to specify the number of matched I want to consider for each treated household. In the analysis presented in this chapter I match each treated household to two control households.
- Caliper: Each treated observation is matched to all control observations within a maximum propensity score distance (caliper).
- Kernel: uses the weighted average of all cases in the control group to estimate counterfactual outcomes. The weight is calculated by the propensity score distance between a treatment case and all control cases and the closest control cases are given the greatest weight (using a Normal function).

  The results presented in this chapter use nearest neighbor and kernel matching.

In general, the advantages of the matching method with a known conditioning set that produces condition (x) does not require separability of outcome or choice equations, exogeneity of conditioning variables, exclusion restrictions, or adoption of specific functional forms of outcome equations. These features are commonly used in control

functions and conventional applications of IVs, although more recent semiparametric estimation relaxes many of these assumptions.

3.4.2 **Panel based fixed-effect estimators (DID)** (Based on Ravallion, 2001 and Albouy, 2005)

The double difference (or difference in difference) method is a popular approach to addressing concerns about the exogeneity assumption in single-difference cross-sectional comparisons. This method compares participants and comparison groups in terms of outcome changes over time relative to the outcomes observed for a pre-intervention baseline. Basically, the method allows for endogeneity in the initial placement of the program, but treats the changes in placement as exogenous to the changes in outcomes.

DID compares groups of participants and non-participants before and after the program. Typically there is a baseline survey before the intervention, covering both participants and non-participants (in the Tapajós case, I do have a baseline survey, in which some households participated in the programs and others did not, but almost all communities surveyed have received some kind of intervention – even if just health related programs). As in the Tapajós, often it is not known who will participate and what the program design will be. After program placement one or more follow up surveys are carried out. Finally, one calculates the mean difference between the "after" and "before"

values of the outcome indicator for each of the treatment and comparison groups. The difference between these two mean differences is the estimate of ATE.

In more formal terms (including a second subscript t to indicate: 0=baseline, 1="after"):

$(Y_{1t}|D=1)=Y_{0t} + G_t + e_t$ where Y_{0t} is the counterfactual outcome measure if the program had not existed, G_t is the gain attributable to the program and e_t is a zero-mean innovation error term uncorrelated with program participation, to allow for measument error in Y_t. An indicator of the counterfactual is available from a comparison group and is given by Y_{0t}'. This might be a noisy indicator due to selection bias or measurement errors (for example, from a change in survey design between the baseline and the follow up surveys).

The two key assumptions of a DID estimator are: i) on average, the changes over time in Y_{0t}' reveal the changes in Y_{0t}, i.e. $E(\Delta Y_{0t}') = E(\Delta Y_{0t})$, which holds if selection bias and other sources of error are separable and time invariant, and so the are swept away by taking differences over time, and ii) period 1 outcomes are not contaminated by the expectation of the program's future placement, i.e. baseline $G_t=0$). Under these assumptions, when taking the expectation over all participants, the DID estimator of mean impact is the single-difference difference impact estimate for the second period minus the single difference in the baseline.

$$ATE = E[(Y_{11}-Y_{01}')- (Y_{10}-Y_{00}')|D=1] = E(G_1|D=1)$$

Equivalently, DID is the outcome gain observed over time for the treatment group minus that for the comparison group. This can be readily generalized to multiple time periods and DID can then be estimated by the common effect regression of Y_t on the (individual and date-specific) participation dummy variable, D_t, with individual and time fixed effects. It is important to remember that when the error term is serially correlated one must take account of this in calculating the standard errors of the DID estimator.

Notice that when mean outcomes for the comparison group are time-invariant ($E[Y_{01}'-Y_{00}'|D=1] = ()$), equation (xxvi) collapses to a reflexive comparison in which one only monitors outcomes for treatment units. Having unchanging mean outcomes for the counterfactual is probably an unfeasible assumption, unless I have enough observations over time and use methods of testing for structural breaks in the time series of outcomes for participants.

Considering data quality, DID designs are particularly vulnerable to measurement errors in poor data. When the changes over time are measured with greater error than the levels, a trade-off emerges between the ability of a DID design to estimate impacts robustly to time-invariant selection bias and the attenuation bias and imprecision that arise from identifying impacts off poorly measured changes over time. Even with good

data, the DID assumption of time-invariant and additive selection bias is implausible for some programs. DID will give a biased participation estimate if the subsequent outcome changes are a function of initial conditions that also influenced the assignment of the sample between the two groups. For example, some ProManejo programs were targeted to communities with greater community organization and this targeting criteria could very well influence the subsequent growth rates. See figure 3.4 for an illustration of this point.

Figure 3.4 shows mean outcome plotted over time, before and after the intervention. The DID estimator is unbiased as long as the selection bias is time invariant, as in figure 3.4.A. In the case where the bias varies with time (as in figure 3.4.B), DID estimates are biased. Controlling for initial heterogeneity is therefore crucial for the credibility of DID estimates. Combining DID with PSM can address this issue to some extent. Using PSM to select the comparison group with DD can reduce (though probably not eliminate) the boas found in other evaluation methods, including matching, control functions and IVs.

Although panel data are not necessary for calculating the DID estimator (all I need is the value of the outcome variables pre and post intervention), household-level panel data open up further options for counterfactual analysis of the joint distribution of outcomes over time for the purpose of understanding the impacts on dynamics. An

example can be found in Ravallion et al (2005) in which impacts of changes were measured not only in the form of a general average but in averages for different groups.

To do a program evaluation of the ProManejo microenterprises I prefer a difference-in-difference estimator, using a PSM to select the comparison group using the 1997 data. As such, *ex post*, I select families that were similar in 1997 and compare the differences in their outcomes in 2006. This takes advantage of the panel data that I have, as well as allow for differences in families to be "corrected" before submitting the population to the treatment. In the following section I present results from matching and difference-in-difference estimators.

3.5 Defining treatment, controls and outcomes

Since program placement was not random, I evaluate the effect of CBEs at four different treatment measures as defined below:

i) Communities with versus communities without CBEs

ii) % of households participating in the CBEs by community

iii) Households participating versus households that never participated in the CBEs.

iv) Households that ever participated versus households that never participated in the CBEs.

Looking at communities with versus communities without CBEs allows for a broader impact of the projects, such as leakage across households within a community. The same idea applies when I compare outcomes in communities with different percentage of households participating. Finally, since the household was the decision unit on whether to participate in the projects or not, I evaluate whether participating households have different outcomes than those that have never participated. This discussion is permeated by an implicit definition of controls: who should I compare treated with? First of all, some might argue that controls should have been picked further from the treated communities. My argument is that this National Forest has long been a pilot for many policies and potential projects to be scaled up (such as logging concessions) and has also received many other non-income generating projects (such as mobile health clinics, school construction, strengthening of community organization programs, etc) that render comparisons with other areas quite difficult. It is also located fairly close to a major regional urban center, which allows communities to sell their production fairly easy (which is not the case for most of the Amazon communities).

When evaluating outcomes at the household level, it is also necessary to define controls clearly. This is due to the fact that some households declared having participated in the past but were not participating when the survey was conducted in 2006. Since program benefits could potentially have been captured earlier, I drop this group from the controls and use only households that never participated in the CBEs as controls.

In terms of outcome variables, I evaluate participation in CBEs with respect to direct outcomes on household welfare such as cash income and wealth indices and also land use, since ProManejo's main objective was to divert households from slash-and-burn agriculture and cattle. Below I present each outcome and its justification:

- Cash income or the sum of cash revenues: due to data limitations it is not possible to construct full income measures. However, since labor and land are the major inputs for households and own consumption is supporting labor supply, cash income is an acceptable measure of household value added (Yang, 1997).
- Wealth indices: Based on assets, as proposed by Cohen and Barnes (1996). Principal component analysis was chosen to construct wealth indices because it reduces the dimensionality of variables for comparison while accounting for the maximum portion of the variance present in the original data (McKenzie, 2005). The choice of assets and variables to be included were based on Vyas and Kumaranayake (2006), Moser and Felton (2007), and Howe et al (2008). After testing various combinations, two groups of assets were chosen for further analysis:
- Liquid assets: gas stove, well, water pump, radio, clock, TV, sowing machine, canoe, bike, chainsaw, shotgun, motor boat, set net, cast net, and livestock. These assets are used in the household, in productive activities or for transportation, and as such have a market within the communities, can be traded and have some liquidity to them.

- Total assets: all liquid assets above, livestock, land (total and types), and house quality. This represents the total wealth of the household.

  For the principal component analysis, the 1997 and 2006 observations were stacked and then the principal component was estimated for the pooled observations. Therefore it necessary to consider the assets for which I have the data in both 1997 and 2006. Items that were first asked about in 2006 (such as motorbikes and cell phones) could not be included in the analysis. These assets were included in the questionnaire in 1997 because a very low percentage of households had them.
- Area in cropland (measured in tarefa = ¼ of a hectare). As a main cause of deforestation and a main income source, slash-and-burn agricultural area is a natural outcome choice of policy relevance.
- Heads of cattle: cattle keeps deforested areas from regenerating into forest. As such, it is an important outcome in terms of environment, but also for livelihoods as cattle as seen as investment and savings.

The next section presents the results of the program evaluation considering the treatments and outcomes discussed above.

3.6 Results

This section is divided according to the three treatment definitions described before:

3.6.1 Communities with versus without CBEs

In total I have data on 22 communities and CBEs were implemented in 12. In this definition of treatment I evaluate the differences in mean household outcomes in communities with CBEs and communities without CBEs. To validate this evaluation, table 3.4 presents descriptive statistics for the outcome variables for treatment and control communities in 1997. As we can see there are no significant differences between the two groups of communities before the implementation of ProManejo.

Next, table 3.5 presents the same outcome variables for treatment and control communities in 2006. In this table we can see that there are some significant differences in the outcome variables. The difference in the total asset index is negative indicating that communities with CBEs are worse off. However, the index for 2006 is smaller that than of 1997, indicating that there might be substitution from the assets I asked about in both 1997 and 2006 to others. Finally, communities with CBE's have less cattle, indicating that the income generating programs might have succeeded in shifting households away from this productive activity.

3.6.2 Percentage of households participating in CBEs

Since the CBEs varied in terms of numbers of households participating in each of them and communities vary in size I can use this difference to try to identify effects of having more households participating in the CBEs. Figure 3.5 shows the difference in 1997 and 2006 income for treatment and control communities and a positive trend as communities have a higher percentage of households participating in the CBEs. The positive outlier in the treatment group refers to a community that started its own project (not financed by ProManejo) in which they plant a fibrous plant and sell directly to the processor in the city.

In terms of wealth, Figure 3.6 shows the difference in liquid assets between 1997 and 2006 for communities without CBEs and communities with different percent of households participating in CBEs. As we can see there is a down sloping trend in the results, indicating that households in communities with a higher participation in CBEs tend to have fewer liquid assets. The trend in total assets (Figure 3.7) is also similar. This may indicate that households are shifting away from the assets considered or that they are actually worse off than communities without CBEs.

Land use is the other outcome of interest. Figure 3.8 shows the change in heads of cattle per household between 1997 and 2006. Although there is a declining trend there is also much heterogeneity in this outcome. For example, the same community that shows

an increase in cash income due to their own project is also the one that presents the greatest decline in cattle. Overall there is a decline in heads of cattle as percentage of households in CBEs increases. Finally, figure 3.9 shows the difference in agricultural area in all communities. It is clear that all communities have decreased their planted area, however, there is a trend that shows that communities with more households participating in CBEs have decreased their agricultural area less than others.

Overall, these trends in the data at the community level seem to suggest that communities with higher percentage of households in CBEs tend to have more cash income, although this pattern is not repeated for wealth indices or cattle.

3.6.3 Households in CBEs versus non participants

I now present the comparison at the decision making unit level: the households. As CBEs were implemented in the communities, households chose whether to participate or not. To evaluate the outcome for these households I use propensity score matching and difference in difference (DID) estimators. To define treatment I use households currently participating in CBEs versus households that never participated in CBEs and, separately, households that have ever participated in the CBEs and those that have not. First, tables 3.6 and 3.7 show the descriptive statistics for the outcome variables and the variables used to construct the propensity score for both 1997 and 2006 by treatment variable (currently participating and ever having participated).

The propensity score function was estimated with the covariates presented in the table as well as the outcome variables for 1997 (pre intervention). See appendix 3.3 for propensity score estimation and balancing results. The resulting balance between treatment and control observations is good, with no significant differences occurring below the 10% significance level. Figure 3.10 shows the common support graph of treatment and control observations after matching (2 of the treated observations are off support and therefore dropped).

Next, tables 3.8 and 3.9 show the outcome results for the matching and DID estimators. As we can see from table 3.8, 2006 participants in the CBEs had higher cash income that non-participants, less liquid assets and more heads of cattle (although the latter result does not hold for the DID estimator). It is also interesting to note that none of these results are significant in the Kernel matching estimator. Compared to the previous results at the community level, the effect of cash income and assets is similar, although the decrease in heads of cattle we saw earlier is not seen in the household analysis.

Outcomes can be influenced by ever having participated in the CBEs rather than current participation, and therefore I estimate separate results using current and past participation as treatment. A separate propensity score function was estimated which also has good balance. The common support graph for this specification is shown in figure 3.11 (Appendix 3.3 also shows the matching results). In table 3.9 I show the results for

ever having participated in the CBEs versus never having participated. Here we see a positive (and significant at the 5% significance level) cash income effect, a positive effect on total assets (although not in the DID estimator) and negative effects in heads of cattle and land in agriculture. The latter effects are indicative that having once participated in the CBEs was enough to change the land use of these households and that ProManejo might have achieved their objective whether or not households continue in the CBEs. This might indicate that ProManejo succeeded in shifting households from subsistence agriculture and integrate them into a cash economy, although this has not improved their overall welfare.

Finally, table 3.10 presents the results of a logit model to explain who drops out of the CBEs conditional on having ever participated. As we can see, the best predictor of who stays is whether they collected NTFPs in 1997. These are households that already carried out forest based activities before CBEs were implemented and therefore would have benefited more (and had to change their behavior less). Other predictors of who stays in projects are: soil quality (which might indicate that less effort has to be put into agricultural production to produce the same output), the interaction variable of receiving pension and number of children (indicating that pension and more children serve as income sources, decreasing the opportunity cost to participating in projects), and the interaction of children and children out of TNF (indicating that household with more overall children tend to stay in CBEs). Variables that explain who drops out are having

more total assets in 1997 (wealthier households who might have higher opportunity cost of time or alternative income sources) and planting a greater variety of crops in 1997 (this could indicate innovative or entrepreneurial households that drop out as benefits to participating decrease). Overall this model indicates that households might have chosen to stay in CBEs due to lack of alternatives or while the initial capital inflow into the CBEs increased the benefits of participating.

3.7 Conclusions

In this chapter I evaluate the effectiveness of CBEs as parts of ICDPs to address welfare of indigenous communities in the Tapajós National Forest in Brazil. As the debate on alternative REDD models intensifies, ICDPs have seen a resurgence in interest as an option that could reduce carbon emissions while addressing poverty issues. Although these projects have been implemented since the 1980s, no rigorous evaluation of their effectiveness has been carried out. This level of scrutiny is necessary if we are to allocate scarce resources efficiently (Morduch, 1999; Morduch, 2000; Ferraro and Pattanayak, 2006). The evaluated CBEs are part of a $1.4 million dollar ICDP. I use a rich pre and post intervention household panel dataset to compare outcomes at the community and individual household level. I compare communities with CBEs and communities non-parametrically while also using matching techniques to compare households (taking into consideration that the program placement and household participation were not random).

From the results we can see that the CBEs had some effects at both the community and the individual household level in terms of increasing cash income. This was not translated into greater wealth, as shown by the mostly negative results of asset indices. Finally, there is some evidence that households participating in CBEs have less cattle, indicating that ProManejo has had some success in shifting land use.

It is interesting to see that effects are more pronounced in the group of households that ever participated in projects rather than the households participating in 2006. When analyzing who dropped out of the CBEs, we see that households that have lower opportunity costs and less alternative income sources tended to stay. Although I cannot test the reason for this given data constraints, my opinion is that households might receive higher benefits from participation when there is more external effort coming into the CBEs and that when responsibilities are transferred to the communities (as the end of funding approaches) it makes sense for some households to leave the project.

In terms of policy implications, this chapter shows that the CBEs of ICDPs in the TNF had positive outcomes. However, considering the lack of robust results and some evidence that suggests that households leaving the CBEs are better off, the sustainability of the CBEs and future household participation is questionable. The matter of sustainability of activities is likely to be partially addressed by the constant inflow of transfers occurring in the new REDD model (in which proponents would receive

payments for avoided deforestation every year), but my results demonstrate the shifting nature of household decisions towards a short run utility maximizing objective.

3.8 References

Albouy, D. 2005. Program Evaluation and the Difference in Difference Estimator. Manuscript. Available at: http://www.er.uqam.ca/nobel/k31160/diff.pdf

Angelsen, A. 2008. REDD models and baselines. International Forestry Review 10(3) pp. 465-475

Barrett, C., Arcese, P. 1995. Are Integrated Conservation-Development Projects (ICDPs) Sustainable? On the Conservation of Large Mammals in Sub-Saharan Africa. World Development 23(7) pp. 1073-1084.

Brooks, J. S., Franzen, M. A., Holmes, C. M., Grote, M. N., Mulder, M. B. (2006) Testing hypotheses for the success of different conservation strategies. Conservation Biology 20:1528-1538.

Browder, J.O. (2002) Conservation and Development Projects in the Brazilian Amazon: Lessons from the Community Initiative Program in Rondonia. Environmental Management 29(6), pp.750-762.

Buckley, G. 1997. Microfinance in Africa: Is it either the Problem or the Solution? World Development 25(7): 1081-1093

Cohen, M., Barnes, C. 1996. Assets and the Impact of Microenterprise Finance Programs. AIMS project. Washington DC: Management Systems International.

Cohenca, D. 2005. A expansão da fronteira agrícola e sua relação com o desmatamento detectado em imagens Landsat TM e ETM+ na região norte da BR-163, Pará entre os anos de 1999 a 2004. Lavras, MG. 59 f. Monograph (Graduate Degree *Lato Sensu*) - FAEPE, UFLA. Available at: http://www.ibama.gov.br/flona_tapajos/download.php?id_download=136

Cohenca, D. 2007. Evolução anual de desmatamento na Floresta Nacional de Tapajós: 1997-2005. Natureza & Conservação 5(1) pp. 35-44

Cordeiro, A. (Coord.). Floresta Nacional do Tapajós - Plano de Manejo. Belterra: IBAMA, 2004. 373 v. I- Informações Gerais. Available at http://www.ibama.gov.br/flona_tapajos/ .

Dias, A. da S. 2001. Social and silvicultural considerations for the diversified forestry management in a riverine community in the "Floresta Nacional do Tapajós", Brazilian Amazon. M.S. thesis CATIE, Turrialba.

Ferraro P.J., Pattanayak S.K. 2006. Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments. PLoS Biol 4(4). Available online at: http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0040105

Garnett, S.T., Sayer, J., du Toit, J. (2007) Improving the effectiveness of interventions to balance conservation and development: a conceptual framework. Ecology and Society 12(1): 2. [online] URL: http://www.ecologyandsociety.org/vol12/iss1/art2/

Heckman, J., Navarro, S. 2004. Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models. Review of Economics and Statistics 86(1) pp. 30-57

Howe, L.D., Hargreaves, J.R., Huttly, S.R.A. 2008. Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. Emerging Themes in Epidemiology 5(3) Available at: http://www.ete-online.com/content/5/1/3

Hughes, R., Flintan, F. 2001. Integrating Conservation and Development Experience: A Review and Bibliography of the ICDP Literature. London: International Institute for Environment and Development.

Hulme, D. 2000. Impact Assessment Methodologies for Microfinance: Theory, Experience and Better Practice. World Development 28(1): 79-98.

IBAMA, 1998. Plano de Manejo da FLONA do Tapajos Informações Gerais Volume I, 256, 304-306

McShane, T.O. and Wells, M.P. (eds). 2004. Getting Biodiversity Projects to Work: Towards More Effective Conservation and Development. Columbia University Press, New York.

McKenzie, D.J. 2005. Measuring inequality with asset indicators. Journal of Population Economics, 18:229-260

Morduch, J. 2000. The Microfinance Schism. World Development 28(4): 617-629

Morduch, J. 1999. The Microfinance Promise. Journal of Economic Literature 37: 1569-1614.

Morduch, J. 1998. Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh. Unpublished Manuscript. http://darp.lse.ac.uk

Moser, C., Felton, A. 2007. The Construction of an Asset Index Measuring Asset Accumulation in Ecuador. The Brookings Institution CPRC Working Paper 87.

Mosley, P. 1997. The use of control groups in impact assessment for microfinance. International Labour Office - Social Finance Unit. Working Paper #19

Mosley, P., Hulme, D. 1998. Microenterprise Finance: Is There a Conflict Between Growth and Poverty Alleviation? World Development 26(5): 783-790

Peskett, L., Yanda, P. 2009. The REDD+ outlook: how different interests shape the future. ODI Background Note. Available at: http://www.odi.org.uk/resources/download/4543.pdf

Pitt, M.M., Khandker, S.R. 2002. Credit Programmes for the Poor and Seasonality in Rural Bangladesh. The Journal of Development Studies 39(2): 1-24.

Pitt, M.M., Khandker, S.R. 1998. The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter? Journal of Political Economy 106(5): 958-996.

Pitt, M.M. 1999. Reply to Jonathan Morduch's "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh" Unpublished manuscript.

Pretes, M. 2002. Microequity and Microfinance. World Development 30(8): 1341-1353

Ravallion, M. 2001. Evaluating Anti-Poverty Programs. In Evenson, R.E., Schultz, T.P. Handbook of Agricultural Economics, Volume 4. Elsevier Science, Amsterdam.

Ravallion, M., Galasso, E., Lazo, T., Philipp, E. 2005. What Can Ex-Participants Reveal About a Program's Impact? Journal of Human Resources 40(Winter) pp.208-230

Robinson, J. G., Redford, K. H. (2004) Jack of all trades, master of none: inherent contradictions among ICD approaches. Pages 10-34 in T. O. McShane and M. P. Wells, editors. Getting biodiversity projects to work: towards better conservation and development. Columbia University Press, New York, New York, USA.

Rosenbaum, P.R. 2002. Observational Studies. Springer Series in Statistics. Springer Science+Business Media, New York.

Rosenbaum, P.R., Rubin, D. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 70(1) pp.41-55

Sayer, J., Campbell, B. (2004) The science of sustainable development. Cambridge University Press, Cambridge, UK.

Schreiner, M. 2002. Evaluation of Microenterprise Programs in the United States. Journal of Microfinance 4(2): 68-91.

Schreiner, M. 2003. A Cost-Effectiveness Analysis of the Grameen Bank of Bangladesh. Development Policy Review 21(3): 357-382.

Servon, L., Bates, T. 1998. Microenterprise as an exit route from poverty: recommendations for programs and policy makers. Journal of Urban Affairs, 20(4): 419-441

Soares, E. "Desafios, resultados, ameaças e oportunidades de uma unidade de conservação na Amazônia: A Floresta Nacional do Tapajós." IBAMA 2004.

Struhsaker, T.T., Struhsaker, P.J., Siex, K.S., 2005. Conserving Africa's rain forests: Problems in protected areas and possible solutions. Biol Conserv 123: 45–54.

Vyas, S., Kumaranayake, L. 2006. Constructing socio-economic status indices: how to use principal components analysis. Health Policy and Planning 21(6): 459–468.

Wells, M.P., Brandon, K. 1992. People and Parks: Linking Protected Area Management with Local Communities. The World Bank, U.S. Agency for International Development and World Wildlife Fund, Washington DC.

Wells, M.P. McShane, T.O. 2004. Integrating Protected Area Management with Local Needs and Aspirations. Ambio 33(8) pp.513-519.

Wells, M. P., McShane, T. O., Dublin, H. T., O'Connor, S., Redford, K. H. (2004) The future of integrated conservation projects: building on what works. Pages 397-422 in T. O. McShane and M. P. Wells, editors. Getting biodiversity projects to work: towards better conservation and development. Columbia University Press, New York, New York, USA.

Worah, S. 2000 International History of ICDPs. In:UNDP (2000) Proceedings of Integrated Conservation and Development Projects. Lessons Learned Workshop, June 12-13, 2000. Hanoi:UNDP/World Bank/WWF.

Yang, D.T. 1997. Education in Production: Measuring Labor Quality and Management. American Journal of Agricultural Economics 79(3), pp. 764-772.

Table 3.1 Land use in the five population areas within the TNF (Source: Cohenca, 2007).

Name	Total area (ha)	Primary forest (%)	Secondary forest (%)	Water and wetlands (%)	Deforestation (%)
Riverine	64,271	83.5	9.9	0.7	5.9
Chibé	1,832	54.1	19.4	0.1	26.5
Aveiro	25,219	84.7	8.7	0.3	6.3
Lots on BR163	6,853	55.0	21.7	0.0	23.3
Plateau	15,159	53.4	21.4	0.1	25.1
Total	113,334	77.6	11.9	0.5	10.1

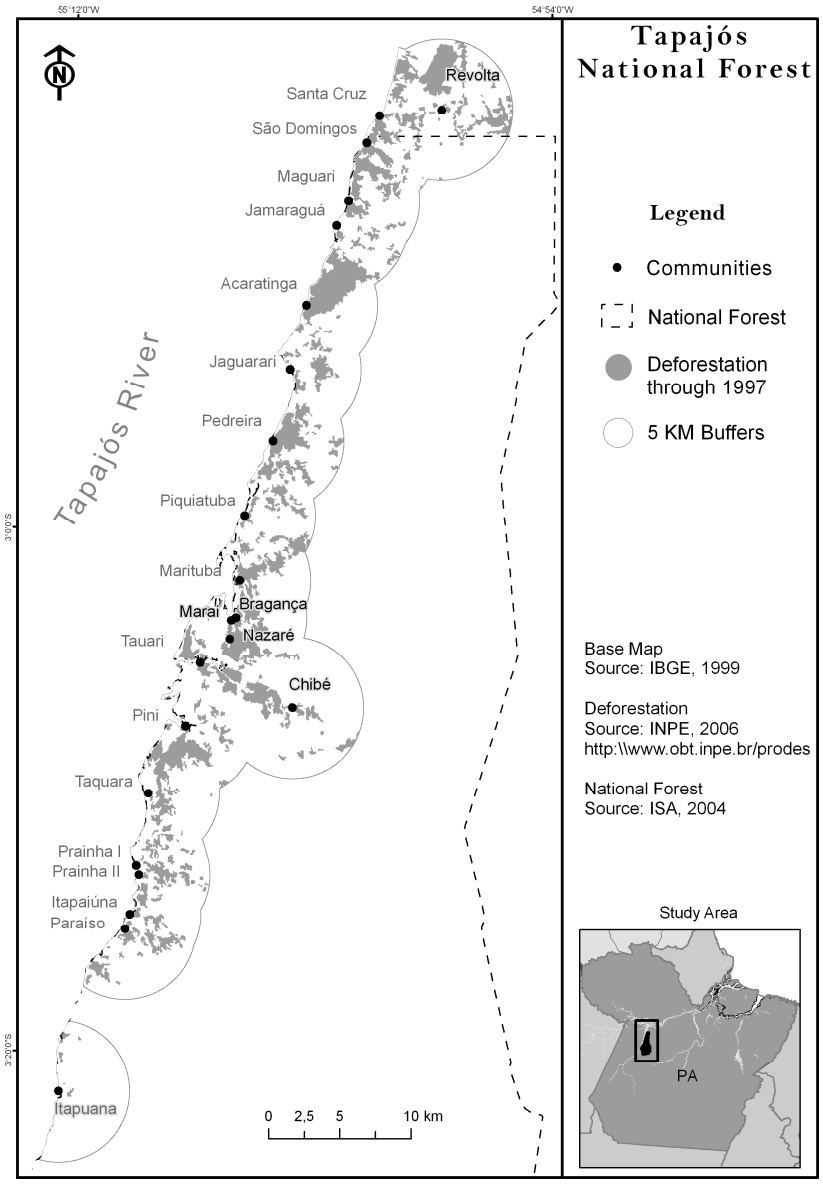


Figure 3.1 Map of communities in the Tapajós National Forest.

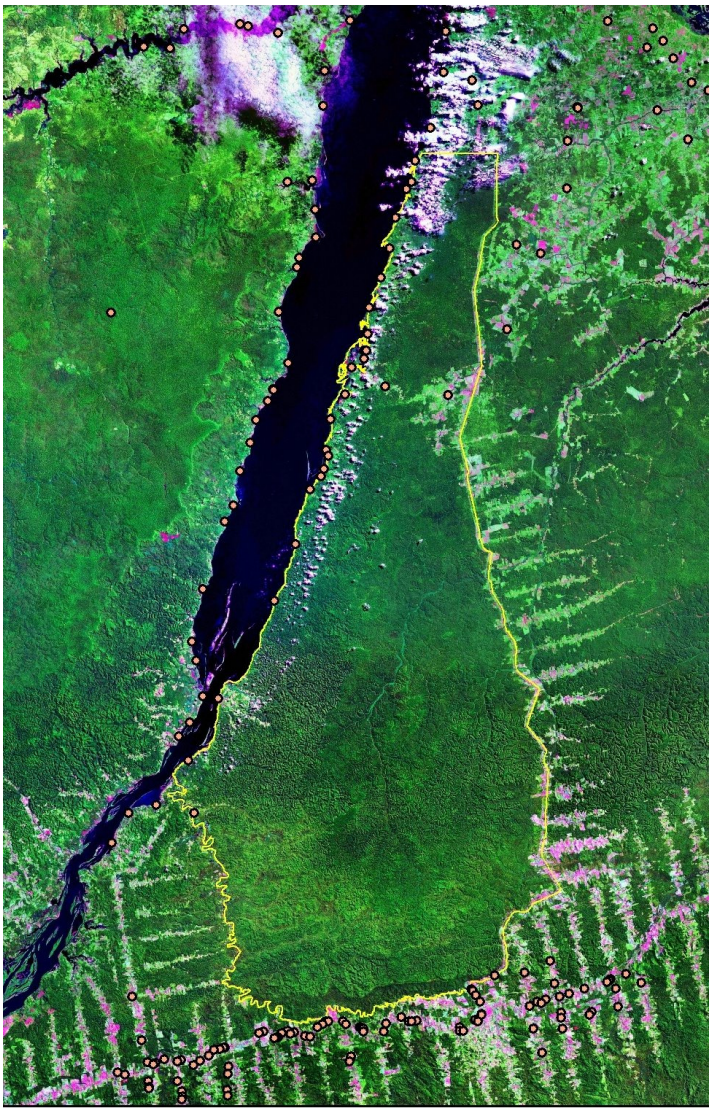


Figure 3.2 Satellite image of the Tapajós National Forest (limits in yellow), communities (dots) and deforested areas (pink and white).

Table 3.2 Projects supported by Component IV of ProManejo in the Tapajós National Forest (highlighted projects refer to CBEs evaluated here).

Name of project	Proposing agency	Objective	Value (US$)
Support for community management and organization	STR S	Institutional enhancement support, popular education, community organization and development	29.985
Promotion of basic health actions	CEAPS	Promote basic health actions in the communities in the Flona	181.136
Support for productive management and natural resources community projects	CEAPS	Promote technical assistance	235.578
Support for andiroba and copaiba oil processing	ASMIPRUT	Infrastructure improvement and processing of vegetable oils	26.555
Valuation of agricultural plots through agroforestry and community development	ASMIPRUT	promote community insertion through production diversification using agroforestry in ag plots management to improve quality of life	105.928
Pilot Community Forest management project	ASMIPRUT	Develop and establish conditions through community forest management to improve infrastructure, increase in information, income generation and involvement of residents	168.153
Support for andiroba and copaiba oil processing	ASMIPRUT	Improve the production, commercialization processes, quality and marketing of vegetable oils and also the quality of life of families involved	117.455

Table 3.2 Continued

Name of project	Proposing agency	Objective	Value (US$)
Improvements in processing and quality of vegetable leather	ASMIPRUT	Improve the practices in extraction and handling of the latex, improve the prod process of vegetable leather by improving quality of products, consolidate already identifies market, improve existing infrastructure, divulge products, promote social organiz	60.071
Honey bees: a sweet alternative	ASMIPRUT	Estimulate residents in the Flona and surroundings to use ntfps through the management of indigenous honey bees, supplying the regional market with a product with greater quality, ecologically correct, with high nutritional and economic value.	100.215
Chicken and agroforestry community projects	ASMIPRUT	Implement an integrated system of chicken raising and agroforestry systems diversifying income and improving subsistance	36.879
Support for institutional strengthening	ASMIPRUT	Enhance the administrative, financial and operational capacity; support the community management projects, divulge the activites developed	18.135
Support for institutional strengthening	AITA	Enhance the administrative, financial and operational capacity; support the community management projects, divulge the activites developed	20.080

Table 3.2 Continued

Name of project	Proposing agency	Objective	Value (US$)
Support for andiroba and copaiba oil processing	AITA	Infrastructure improvement and processing of vegetable oils	9.506
Community forest management nad production of furniture and crafts	AITA	Improve quality of life of these communities through the enhancement of AITA community development and creation of opportunities of training, jobs and income generation for the families involved in the project	286.259
Chicken and agroforestry community projects	AITA	Implement an integrated system of chicken raising and agroforestry systems diversifying income and improving subsistance	12.293
Community forest management and Oficinas Caboclas	IPAM	develop a community forest management model based on the use of forest resources to diversify the hh income and improve quality of life of people directly and indirectly involved	284.736
Community forest management and marketing training	ASMOP	Improve quality of life of families involved in the project, through generation of income and jobs with the implementation of a marketry workshop and implementation of forest management areas for input supply, consolidating forest management in this community	63.271

Table 3.2 Continued

Name of project	Proposing agency	Objective	Value (US$)
Complementation projects to community forest managemnt and rustic furniture training	ACOMTEP	Increase the capacity of management of the group in the production, processing, commercialization and forest management activities; improve production structure and quality of crafts developed; support forest certification as a means to access new markets	63.630
Pilot Community Forest management project - Projeto Ambe	COOMFLONA	Develop sustainable forest management through the multiple use of forest resources and reduced impact logging in the areas destined to timber forest management in the Flona, managed by the local communities to attain social, environmental and economic ben	892.612
TOTAL			2.997.212

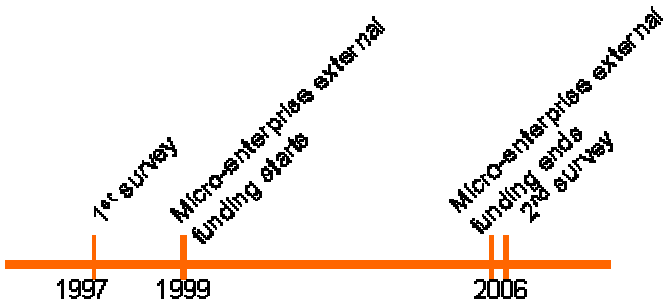


Figure 3.3 Timeline of interventions and surveys in the Tapajós National Forest.

Table 3.3 Descriptive statistics and T tests (*** significant at 1%, ** significant at 5%, * significant at 10%).

Definition	Year	Variable	Obs	Mean	Std. Dev.	Min	Max	T test	
Age of male head of households	1997	agep	188	44.49	14.98	20	97		
Age of female head of households	1997	agem	193	41.44	15.34	16	89		
Number of people in household	1997	family	208	5.98	2.95	1	15	1.34	
	2006	family06	216	5.62	2.58	1	15		
Proportion of women in household	1997	women	208	0.30	0.16	0	1	-6.53	***
	2006	women06	216	0.41	0.19	0	100		
Number of children (<10 years old)	1997	kids	208	2.24	1.82	0	7	3.52	***
	2006	kids06	216	1.62	1.78	0	7		
Max years of education among heads of hh	1997	educ	158	3.86	2.03	0	8	-2.47	**
	2006	educ06	216	4.45	2.43	0	14		
Years in the Amazon	1997	yearam	208	44.67	16.69	8	97		
Whether has children living outside of FLONA	1997	childout	214	0.75	0.43	0	1	-11.71	***
	2006	childnfl	152	2.90	2.64	0	11		
Number of remittances per year	1997	remitfry	121	5.84	12.48	0	52	5.31	***
	2006	remitfry06	212	1.06	3.05	0	12		
Count of assets	1997	posses	208	5.56	2.67	1	14	-5.04	***
	2006	posses06	214	6.88	2.69	0	14		
Hectares of cropland	1997	cropland	208	1.90	1.82	0	13.25	7.03	***
	2006	cropland06	215	0.95	0.81	0	5.5		
Hectares of land burnt	1997	burn96	208	1.44	1.65	0	13.25		
If has terra preta (soil quality indicator)	1997	terap	207	0.38	0.50	0	2		
Number of crops planted	1997	nocrops	208	2.39	1.61	0	5		

Table 3.3: Continued

Definition	Year	Variable	Obs	Mean	Std. Dev.	Min	Max	T test	
Number of ducks and chickens	1997	poultry	208	13.70	12.82	0	60	0.92	
	2006	poultry06	202	12.24	18.81	0	201		
Number of pigs	1997	pigs	208	1.15	5.63	0	70	2.33	***
	2006	qhog	214	0.24	1.15	0	13		
Number of cows	1997	qcattle	208	2.23	8.10	0	75	-0.24	
	2006	qcattle06	216	2.41	7.90	0	70		
Number of beef cows	1997	cowsbeef	208	1.96	7.58	0	73	0.23	
	2006	cowsbeef06	216	1.79	7.12	-5	67		
Number of trips to forest to collect products	1997	fptimes	158	33.70	156.53	0	999	2.50	**
	2006	fptimes06	211	6.70	7.75	0	60		
Whether sold non-timber forest products	1997	soldntfp	208	0.64	1.25	0	9	4.65	***
	2006	soldntfp06	216	0.22	0.42	0	1		
Number of non-timber forest products collected	1997	collntfp	208	4.44	3.14	0	13	2.14	**
	2006	collntfp2	216	3.87	2.25	0	10		
Whether taps rubber	1997	sering	208	0.17	0.38	0	1	-2.54	**
	2006	sering06	210	0.28	0.45	0	1		
Number of trips to Santarém per month	1997	viajarstm	206	13.50	8.95	0.5	52	6.67	***
	2006	viajarstm06	215	7.01	10.89	0.5	52		
If travels to Santarém at least once per month	1997	viajm	216	0.74	0.44	0	1	10.30	***
	2006	viajm06	216	0.30	0.46	0	1		

Table 3.3: Continued

Definition	Year	Variable	Obs	Mean	Std. Dev.	Min	Max	T test	
Number of people receiving pension	1997	pension	208	0.28	0.54	0	2	-3.97	***
	2006	pensionn	216	0.54	0.75	0	3		
Other government transfers received	2006	govinc	216	1.24	1.65	0	8		
Whether received wages	1997	wtab	208	0.34	0.47	0	1	-2.28	**
	2006	wtab06	216	0.44	0.50	0	1		
Participation in ProManejo microenterprises	2006	emoc	216	0.33	0.47	0	1		
Cash income (R$)	2006	income	216	2588	2679	-232	14753		

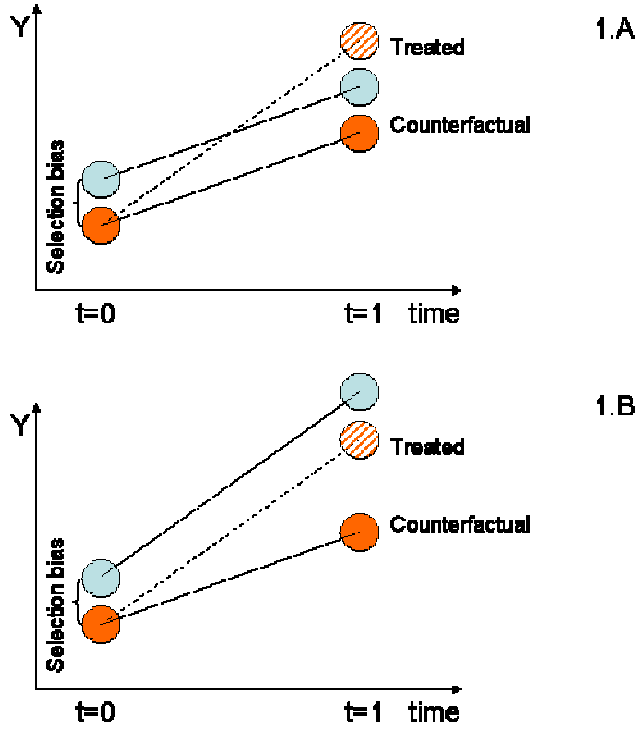


Figure 3.4 Bias in double difference estimates for a targeted problem (from Ravallion, 2001).

Table 3.4 Outcome for treated and control communities in 1997 (None of the variables is significantly different at the 10% significance level).

Definition	Variable	Treatment		Control		Difference	Ttest
		Mean	Std Dev	Mean	Std Dev		
Cash income	income97	829.42	308.82	1007	793	-177.58	
Liquid assets	fliq971	0.27	0.28	0.26	0.60	0.01	
Physical assets	fphys97	-0.08	0.25	-0.32	0.38	0.23	
Heads of cattle	cattle	2.59	3.16	1.56	2.67	1.04	
Area in agriculture	cropland	2.01	0.87	2.56	1.70	-0.56	

Table 3.5 Outcomes of treated and control communities in 2006 (*** significant at 1%, ** significant at 5%, * significant at 10%).

Description	Variable	Method	Treatment		Control		Difference	Ttest	
			Mean	Std Dev	Mean	Std Dev			
Cash income	income06	2006 only	4195.00	1166.80	3353.80	1595.30	841.20	-1.31	
	dincome	DID	3365.60	1100.70	2346.80	1772.20	1018.80	-1.47	
Liquid assets	fliq061	2006 only	-0.22	0.24	0.16	0.63	-0.38	2.05	*
	dpcaliq	DID	-0.48	0.23	-0.04	0.75	-0.43	1.95	*
Physical assets	fphys06	2006 only	0.39	0.37	0.71	0.70	-0.33	1.44	
	dpcaphys	DID	0.50	0.35	1.12	0.70	-0.62	2.67	**
Heads of cattle	cattle06	2006 only	1.68	1.93	2.91	4.65	-1.23	0.88	
	difcattl	DID	-0.77	2.70	1.36	2.18	-2.13	1.98	*
Area in agriculture	cropland	2006 only	1.03	0.31	0.87	0.28	0.16	-1.25	
	dcroplan	DID	-0.99	0.71	-1.69	1.57	0.71	-1.42	

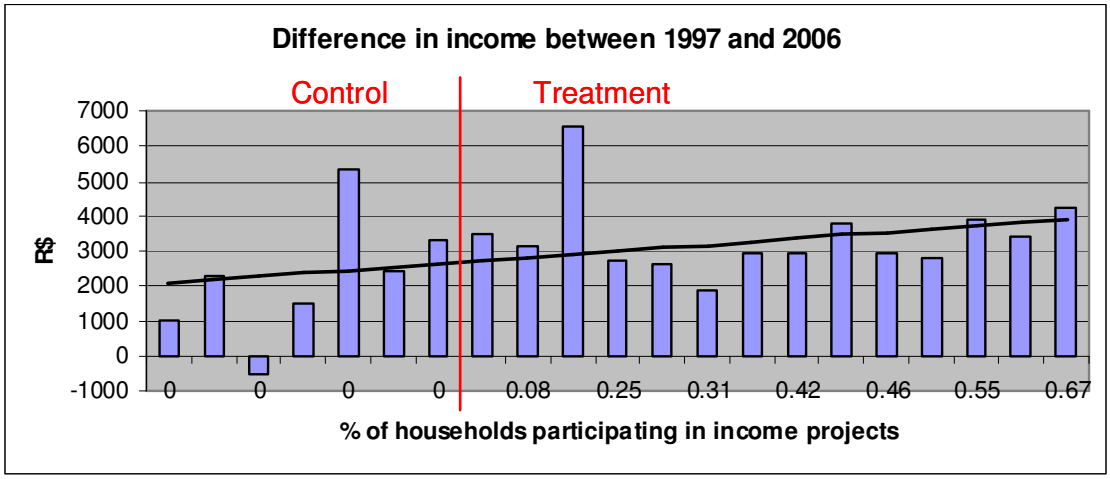


Figure 3.5 Difference in cash income between 1997 and 2006 for control and treatment communities.

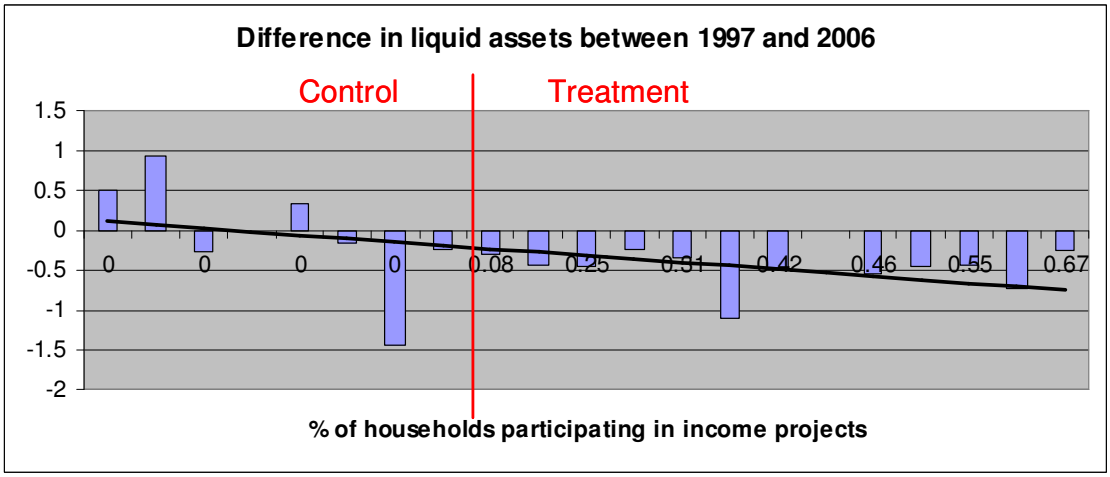


Figure 3.6 Difference in liquid assets between 1997 and 2006 for control and treatment communities.

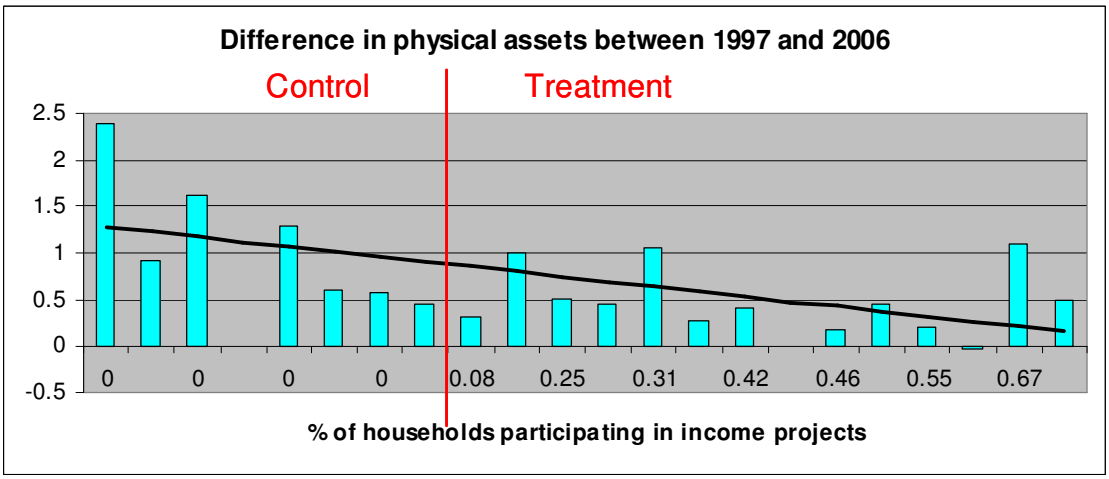


Figure 3.7 Difference in total assets between 1997 and 2006 for control and treatment communities.

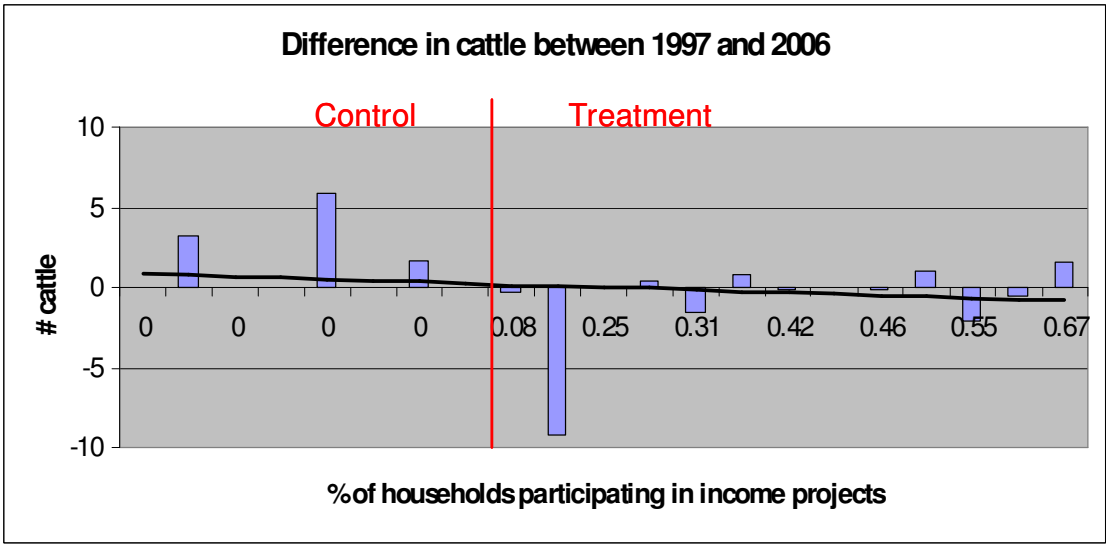


Figure 3.8 Difference in heads of cattle between 1997 and 2006 for control and treatment communities.

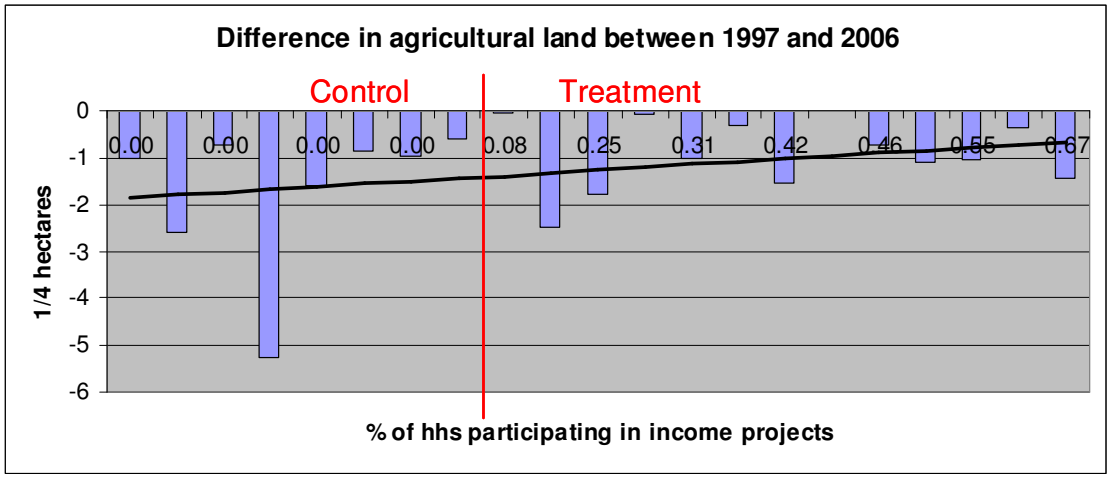


Figure 3.9: Difference in land in agriculture between 1997 and 2006 for control and treatment communities.

Table 3.6 Descriptive statistics for 1997 and 2006 by current participants in the CBEs.

Type	Definition	Year	Variable	Currently participating			
				No		Yes	
				Mean	Std. Dev.	Mean	Std. Dev.
Treatment	Currently participating in CBEs	2006	emoc2	0	0	1	0
	Ever participated in CBEs	2006	emocever	0.36	0.48	1	0
Outcomes	Cash income (R$)	1997	income06	3777.85	4114.37	4608.14	6305.61
		2006	income97	794.20	1197.56	930.28	966.80
	Physical assets	1997	fphys97	-0.12	0.94	0.02	1.02
		2006	fphys06	0.35	1.03	0.45	1.09
	Liquid assets	1997	fliq971	0.26	1.30	0.25	0.81
		2006	fliq061	-0.11	1.06	-0.28	0.67
	Number of cows	1997	qcattle	2.38	8.91	1.50	4.20
		2006	qcattle06	2.75	8.87	1.29	2.69
	Hectares of cropland	1997	cropland	1.83	1.85	2.04	1.65
		2006	cropland06	0.95	0.89	0.92	0.44
Propensity score explanatory variables	Age of male head of households	1997	agep	44.30	14.79	45.55	16.42
	Age of female head of households	2006	agem	41.91	15.61	39.13	14.07
	Number of people in household	1997	family	5.83	2.97	6.30	2.54
		2006	family06	5.55	2.62	5.84	2.46
	Proportion of women in household	1997	women	0.31	0.17	0.27	0.13
		2006	women06	42.30	19.48	37.38	18.23
	Number of children (<10 yo)	2006	kids	2.12	1.81	2.57	1.72
		1997	kids06	1.62	1.78	1.62	1.79

Table 3.6 continued

Type	Definition	Year	Variable	Currently participating			
				No		Yes	
				Mean	Std. Dev.	Mean	Std. Dev.
Propensity score explanatory variables	Max years of education among heads of hh	2006	educ	3.89	2.06	3.65	2.02
		1997	educ06	4.51	2.52	4.26	2.08
	Years in the Amazon	2006	yearam	44.29	16.99	46.38	16.32
	Whether has children living outside of TNF	1997	childout	0.72	0.45	0.86	0.35
		1997	childnfl	3.02	2.64	2.60	2.64
	Number of remittances per year	1997	remitfry	6.23	12.93	2.46	4.76
		1997	remitfry06	1.02	2.98	1.19	3.32
	Hectares of land burnt	2006	burn96	1.39	1.71	1.64	1.43
	If has terra preta (soil quality indicator)	1997	terap	0.36	0.48	0.45	0.54
	Number of crops planted	2006	nocrops	2.36	1.62	2.43	1.61
	Number of ducks and chickens	1997	poultry	12.71	11.56	16.81	16.17
		2006	poultry06	12.77	20.52	10.47	11.65
	Number of pigs	1997	pigs	1.17	6.04	1.11	3.91
		2006	qhog	0.27	1.29	0.14	0.45
	Number of trips to forest to collect products	1997	fptimes	34.31	156.36	32.39	161.19
		2006	fptimes06	6.15	6.45	8.67	10.91
	Whether sold NTFPs	1997	soldntfp	0.70	1.31	0.45	0.97
		2006	soldntfp06	0.18	0.39	0.38	0.49
	Number of NTFPs collected	1997	collntfp	4.25	3.19	4.94	3.02
		2006	collntfp2	3.86	2.28	3.96	2.12
	Whether taps rubber	1997	sering	0.17	0.38	0.17	0.38
		2006	sering06	0.25	0.43	0.39	0.49
	Number of trips to Santarém per month	1997	viajarstm	13.36	8.48	14.00	10.48
		1997	viajarstm06	7.28	11.49	6.09	8.68
	Number of people receiving pension	2006	pension	0.29	0.54	0.30	0.55
		1997	pensionn	1.39	0.49	1.41	0.62
	Other government transfers received	2006	govinc	2.47	1.78	1.83	0.79
	Whether received wages	1997	wtab	0.32	0.47	0.38	0.49
		2006	wtab06	0.41	0.49	0.56	0.50

Table 3.7 Descriptive statistics for 1997 and 2006 by having ever participated in CBEs.

Type	Definition	Year	Variable	Currently participating				Ever participated			
				No		Yes		No		Yes	
				Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Treatment	Currently participating in CBEs	2006	emoc2	0	0	1	0	0	0	0.45	0.50
	Ever participated in CBEs	2006	emocever	0.36	0.48	1	0	0	0	1	0
Outcomes	Cash income (R$)	1997	income06	3777.85	4114.37	4608.14	6305.61	3657.39	4406.36	4271.33	4985.17
		2006	income97	794.20	1197.56	930.28	966.80	790.74	1342.47	859.39	926.96
	Physical assets	1997	fphys97	-0.12	0.94	0.02	1.02	-0.15	0.99	-0.03	0.93
		2006	fphys06	0.35	1.03	0.45	1.09	0.42	1.13	0.33	0.95
	Liquid assets	1997	fliq971	0.26	1.30	0.25	0.81	0.30	1.49	0.22	0.84
		2006	fliq061	-0.11	1.06	-0.28	0.67	-0.01	1.21	-0.30	0.65
	Number of cows	1997	qcattle	2.38	8.91	1.50	4.20	2.64	9.94	1.74	5.74
		2006	qcattle06	2.75	8.87	1.29	2.69	3.16	10.07	1.70	4.94
	Hectares of cropland	1997	cropland	1.83	1.85	2.04	1.65	1.88	2.14	1.88	1.43
		2006	cropland06	0.95	0.89	0.92	0.44	0.98	0.95	0.92	0.64
Propensity score explanatory variables	Age of male head of households	1997	agep	44.30	14.79	45.55	16.42	44.93	14.70	44.29	15.60
	Age of female head of households	2006	agem	41.91	15.61	39.13	14.07	42.89	16.57	39.70	13.77
	Number of people in household	1997	family	5.83	2.97	6.30	2.54	5.85	2.93	6.01	2.85
		2006	family06	5.55	2.62	5.84	2.46	5.54	2.56	5.69	2.61
	Proportion of women in household	1997	women	0.31	0.17	0.27	0.13	0.30	0.17	0.30	0.16
		2006	women06	42.30	19.48	37.38	18.23	41.10	19.64	41.23	18.99
	Number of children (<10 yo)	2006	kids	2.12	1.81	2.57	1.72	2.12	1.82	2.32	1.78
		1997	kids06	1.62	1.78	1.62	1.79	1.67	1.86	1.57	1.70
	Max years of education among hea	2006	educ	3.89	2.06	3.65	2.02	3.64	2.11	4.04	1.97
		1997	educ06	4.51	2.52	4.26	2.08	4.42	2.60	4.48	2.25
	Years in the Amazon	2006	yearam	44.29	16.99	46.38	16.32	45.19	17.34	44.33	16.38

Table 3.7 Continued

Type	Definition	Year	Variable	Currently participating				Ever participated			
				No		Yes		No		Yes	
				Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Propensity score explanatory variables	Whether has children living outsid	1997	childout	0.72	0.45	0.86	0.35	0.72	0.45	0.78	0.41
		1997	childnfl	3.02	2.64	2.60	2.64	2.99	2.42	2.84	2.83
	Number of remittances per year	1997	remitfry	6.23	12.93	2.46	4.76	6.55	13.86	4.02	8.37
		1997	remitfry06	1.02	2.98	1.19	3.32	1.10	3.15	1.02	2.96
	Hectares of land burnt	2006	burn96	1.39	1.71	1.64	1.43	1.30	1.91	1.59	1.34
	If has terra preta (soil quality indica	1997	terap	0.36	0.48	0.45	0.54	0.35	0.48	0.41	0.51
	Number of crops planted	2006	nocrops	2.36	1.62	2.43	1.61	2.22	1.69	2.52	1.53
	Number of ducks and chickens	1997	poultry	12.71	11.56	16.81	16.17	12.33	10.86	14.91	14.43
		2006	poultry06	12.77	20.52	10.47	11.65	14.35	24.25	10.06	10.39
	Number of pigs	1997	pigs	1.17	6.04	1.11	3.91	1.25	7.30	1.06	3.28
		2006	qhog	0.27	1.29	0.14	0.45	0.32	1.45	0.15	0.76
	Number of trips to forest to collect	1997	fptimes	34.31	156.36	32.39	161.19	37.14	166.71	31.20	149.70
		2006	fptimes06	6.15	6.45	8.67	10.91	6.06	5.92	7.35	9.13
	Whether sold NTFPs	1997	soldntfp	0.70	1.31	0.45	0.97	0.62	1.17	0.66	1.32
		2006	soldntfp06	0.18	0.39	0.38	0.49	0.20	0.40	0.25	0.44
	Number of NTFPs collected	1997	collntfp	4.25	3.19	4.94	3.02	3.94	3.32	4.86	2.93
		2006	collntfp2	3.86	2.28	3.96	2.12	3.97	2.40	3.79	2.09
	Whether taps rubber	1997	sering	0.17	0.38	0.17	0.38	0.18	0.39	0.16	0.37
		2006	sering06	0.25	0.43	0.39	0.49	0.23	0.42	0.33	0.47
	Number of trips to Santarém per mo	1997	viajarstm	13.36	8.48	14.00	10.48	13.25	8.48	13.75	9.41
		1997	viajarstm06	7.28	11.49	6.09	8.68	7.17	11.57	6.85	10.26
	Number of people receiving pensio	2006	pension	0.29	0.54	0.30	0.55	0.28	0.53	0.30	0.55
		1997	pensionn	1.39	0.49	1.41	0.62	1.39	0.49	1.40	0.54
	Other government transfers receive	2006	govinc	2.47	1.78	1.83	0.79	2.26	1.40	2.35	1.77
	Whether received wages	1997	wtab	0.32	0.47	0.38	0.49	0.38	0.49	0.29	0.45
		2006	wtab06	0.41	0.49	0.56	0.50	0.41	0.49	0.48	0.50

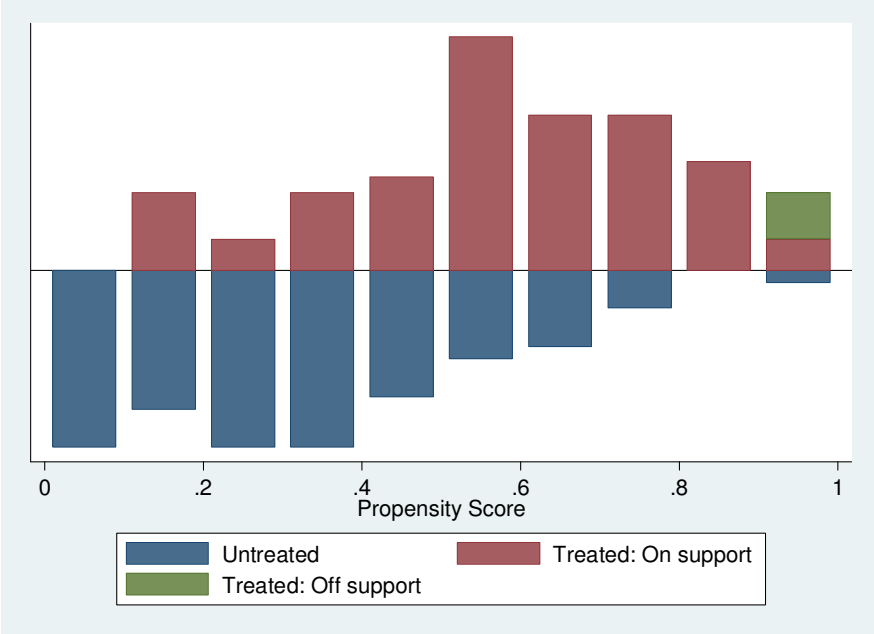


Figure 3.10 Common support graph for current participants (using households that never participated as controls).

Table 3.8 Matching results for 2006 participants versus never having participated in CBEs.

Outcome	Method	Nearest Neighbor			Kernel	
		ATT		Std error	ATT	Std error
Cash income	2006	1604.87	*	1025.07	1054.21	975.40
	DID	1548.01	*	1114.18	1063.58	1095.63
Liquid assets	2006	-0.124		0.188	-0.095	0.162
	DID	-0.294	*	0.185	-0.143	0.170
Physical assets	2006	0.098		0.213	0.048	0.190
	DID	-0.078		0.240	-0.041	0.185
Heads of cattle	2006	1.177	*	0.847	0.106	0.841
	DID	-0.097		0.842	-0.525	0.729
Land in agriculture	2006	0.065		0.189	-0.043	0.162
	DID	-0.131		0.338	-0.117	0.276

Table 3.9 Matching results for having ever participated in CBEs versus not.

Outcome	Method	Nearest Neighbor ATT		Nearest Neighbor Std error	Kernel ATT		Kernel Std error
Cash income	2006	1632.66	**	861.43	697.57		742.19
	DID	1889.70	***	918.74	576.00		805.52
Liquid assets	2006	0.048		0.138	-0.181		0.151
	DID	-0.107		0.169	-0.129		0.176
Physical assets	2006	0.343	**	0.192	0.062		0.169
	DID	0.102		0.192	-0.104		0.155
Heads of cattle	2006	-2.253	*	1.702	-1.794	*	1.348
	DID	-0.800		1.861	-1.706	*	1.088
Land in agriculture	2006	-0.246	*	0.193	-0.204	*	0.142
	DID	-0.179		0.451	-0.200		0.291

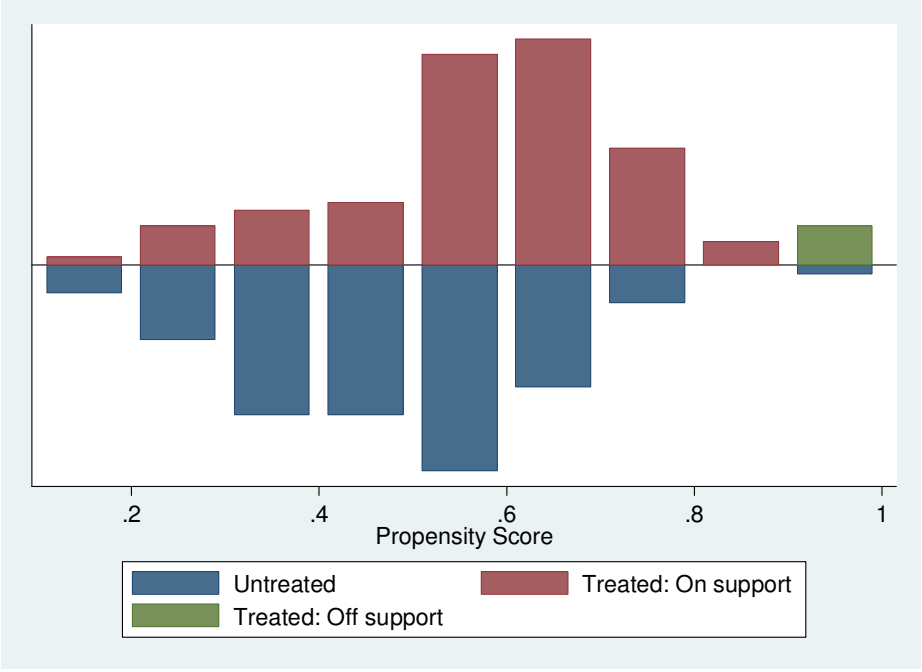


Figure 3.11 Common support graph for having ever participated in CBEs.

Table 3.10 Logit of CBE dropouts (Likelihood ratio test is significant at 2% significance level and pseudo R^2 is 25%, sample size of 145 households).

Description	Variable	Estimate		Std. Err.
household size	family	-0.17		0.55
household size squared	family2	1.53E-02		3.25E-02
# children in hh	kids	-6.28E-02		0.76
# children in hh squared	kids2	0.20	*	0.11
# children outside TNF	childout	0.57		1.12
Age of hh head	agep	8.25E-02		9.03E-02
Age of hh head squared	agep2	-1.07E-03		8.59E-04
# of trips to city per year	viajarstm	8.62E-02		5.64E-02
# people in hh receiving pension	pension	1.14		0.79
If collected NTFPs	collntfp	-0.35	***	0.13
Cash income in 1997	income97	-4.44E-04		3.72E-04
Heads of cattle in 1997	cattle	-3.15E-02		5.87E-02
Physical capital in 1997	fphys97	0.73	**	0.38
Soil quality dummy	terap	-1.17	**	0.57
Number of crops	nocrops	0.46	**	0.21
Pension*Kids	penkids	-0.64	*	0.35
Kids*Kidsout	kidsout	-0.76	*	0.46
Viajarstm*kids	viajkids	-2.28E-02		1.55E-02
	_cons	-1.19		2.29

APPENDICES

Appendix 3.A

Timeline of relevant events for the communities and IBAMA in the Tapajós National Forest

Although nowadays the Environmental Agency (IBAMA) prides itself in working with the communities, it was not always this way. Below I present a timeline with the events that have marked the history of the TNF and its communities:

1974: Creation of the TNF, the first National Forest in the Amazon.

1978: the first population survey is carried out by IBDF (entity responsible for management of the national forest) of households on the margins of the Tapajós River to present the dispossessment policy.

1980: Communities open trail in the forest to mark their land use limits in protest against the dispossessment policy.

1983: IBDF recognizes the exclusion of community use areas temporarily

1988: Government sends IBAMA managers to negotiate with communities to solve conflicts.

1989: Creation of the ITTO (International Tropical Timber Organization) project without community involvement, the first Brazilian attempt at forest concessions.

1990: Forest inventory of the ITTO project upsets residents and provokes discussions and assemblies.

1992: The Santarém city hall forms the TNF Study Group with the participation of representatives of the communities, rural workers syndicate, executive and legislative powers, IBAMA, Incra, Emater, Embrapa, FCAP, PSA;

1993: Conclusion of a socioeconomic study of the communities in the TNF by IBAMA;

1995: Presentation of the law that would exclude all communities and population centers from the protected area

Creation of the TNF Work Group composed by IBAMA, Incra, Environmental Secretary of Santarém, Rural workers syndicate, PSA, GDA, CPT, AITA, ASMIPRUT, CCB – Community Council of Belterra, CEAPAC, and four representatives of the 16 communities;

Through the support of PPG7, legal support service was hired;

1996: Plebiscite and voting on proposals to exclude communities from National Forest areas. Of the 1291 residents, 763 (59.1% voted) and 36.73% voted to remain within the limits of the National Forest but with legal use of their resources, 61.73 decided to leave the Flona.

Through PPG7 the Director Plan of the TNF, forest zoning, and community mapping were elaborated;

1997: Approval of the Use Plan for the communities in the TNF;

PPG7 and GTZ financed a study of markets for non-timber forest products asked for by the communities. As one of the first results the communities of São Domingos, followed by Nazaré and Pedreira, started the production and commercialization of copaiba and andiroba oils.

1998: Conduct Adjustment Term between the public ministry and IBAMA is signed to allow traditional populations to use the National Forest;

1999: ProManejo projects started with the objective of diverting income sources away from slash-and-burn agriculture.

2000: New system of protected areas, SNUC, is approved. It allows the presence of traditional populations to live within Sustainable Use Protected Areas;

2001: New project to ensure the communities the right to use the land, as encompassed by SNUC;

Creation of the National Forest Managing Council;

2002: Communities create the Land Tenure Commission;

2004: Communities create the Federation of Organizations and Traditional Communities of the TNF.

Appendix 3.B

Sample attrition in TNF database between 1997 and 2006

To investigate if the sample attrition in the database of households in the Tapajós National Forest (TNF) I calculate correlations between attrition and several household variables and also a multivariate probit. I define attrition as all households that were interviewed in 1997 but not in 2006. Results are presented below:

Table 3.B.1 Correlation between leaving the sample and 1997 household characteristics

Definition	Variable	if left sample	
Has "terra preta"	terap	-0.12	**
Number of crops	nocrops	-0.17	***
Household size	family	-0.20	***
Number of children	kids	-0.06	
If has children living outside of TNF	childout	0.33	***
Age of father	agep	-0.03	
Age of father squared	agep2	-0.01	
Trips to city (per month)	viajar	0.04	
If receives pension	pension	0.01	
If collected NTFPs	collntfp	-0.16	***
If sold NTFPs	soldntfp	-0.12	**
Walking distance to school (in min)	msch	0.09	*
Area planted in agriculture	cropland	-0.07	
Cash income	income97	-0.04	
Heads of cattle	cattle97	-0.06	

As we can see in the table above, leaving the sample is positively correlated with having children living outside the TNF and living further from the school. Attrition is negatively correlated with having good soils (“terra preta”), planting more crops, having

a larger household, and collecting and selling non-timber forest products (NTFPs). To investigate the issue further, I estimated a multivariate probit with the same variables presented above:

Table 3.B.2 Multivariate probit of sample attrition (model significant at 1%).

Definition	Variable	Estimate		Std error
Intercept	Intercept	1.11		0.75
Has "terra preta"	terap	-0.20		0.18
Number of crops	nocrops	-6.58E-02		5.99E-02
Household size	family	-0.15	**	5.71E-02
Number of children	kids	0.12		8.47E-02
If has children living outside of TNF	childout	0.26	***	6.51E-02
Age of father	agep	-3.51E-02		3.04E-02
Age of father squared	agep2	2.38E-04		3.03E-04
Trips to city (per month)	viajar	-2.03E-02		8.16E-02
If receives pension	pension	5.06E-03		0.19
If collected NTFPs	collntfp	-3.05E-02		3.32E-02
If sold NTFPs	soldntfp	-9.15E-02		8.96E-02
Walking distance to school (in min)	msch	6.11E-03		3.84E-03
Area planted in agriculture	cropland	3.17E-02		5.51E-02
Cash income	income97	4.70E-05		7.80E-05
Heads of cattle	cattle97	-2.31E-02		1.71E-02

The model is significant despite having a questionable fit (very low maximum likelihood) and that is not surprising given that only two variables are significant in the model: household size (which makes a household less likely to drop from the sample) and whether the household has children living outside the TNF. The significance of these variables is expected and easy to explain: having a larger household made it easier for enumerators to find it and also might make it more difficult for the household to move

elsewhere. It there are children living outside the TNF the household has a link to another location, which makes it easier to move elsewhere.

Appendix 3.C

Propensity score estimation

In this appendix I present and discuss the propensity score estimation used for the household matching. The discussion is divided in two: first I discuss the results using 2006 participation as the treatment variable (and excluding households that had participated at an earlier point in time) and then I present results using "ever participated" in the CBEs as the treatment variable.

- **Currently participating (as of 2006):**

The table below presents the results of the propensity score logit, where "1" equals households currently participating in the CBEs and "0" represents households that have never participated in the CBEs:

Table 3.C.1 Results of the propensity score logit for current participation.

Description	Variable	Estimate		Std. Err.
Good soil quality dummy (terra preta)	terap	0.97	**	0.43
Number of crops planted	nocrops	8.93E-02		0.16
Household size	family	-0.17		0.13
Number of children	kids	-0.82	**	0.39
Number of children living outside TNF	childout	-5.61E-02		0.94
Age of male head of household	agep	-0.14	*	0.08
Age of male head of household squared	agep2	1.46E-03	*	7.57E-04
Number of trips to city per month	viajarstm	-0.10	**	4.78E-02
If receives pension	pension	-1.43	*	0.74
If collected NTFPs	collntfp	0.19	**	0.08
If sold NTFPs	soldntfp	-0.20		0.17
Interaction (pension*kids)	penkids	0.91	**	0.41
Interaction (kids*childout)	kidsout	0.65	*	0.36
Interaction (viajarstm*kids)	viajkids	3.28E-02	**	1.55E-02
Minutes walking from school	msch	-2.10E-02	*	1.25E-02
Cash income	income97	9.41E-05		2.24E-04
Area planted in crops	cropland	7.14E-02		0.13
Number of heads of cattle	cattle	-1.28E-02		4.40E-02
Principal component total assets	fphys97	0.15		0.30
Principal component liquid assets	fliq971	-0.15		0.28
Intersect	_cons	3.66	**	1.76

The model is significant at the 1% significance level and the pseudo R^2 is 21%. As we can see, a larger number of children in a household decreases the probability that a household will participate in the CBEs. Receiving retirement pension also decreases these odds. Age has a quadratic relationship with participation, with younger and older households having a higher probability of participating. Having good soil increases participation, as does experience collecting NTFPs. These characteristics might indicate that the household has a higher productivity or experience with the products promoted by the CBEs were more likely to participate. Finally some interaction terms are significant: pension and children (indicating some tradeoff between having children and other sources of income), children and children living outside of the TNF (which indicates that

the total number of children might have a larger influence on participation than where they live) and number of trips to the city and children. The next table shows the balance between treatment and control households before and after the matching:”

Table 3.C.2 Treament and control characteristics before and after matching (based on matching with 2 nearest neighbors).

		Mean			%reduct		
Variable	Sample	Treated	Control	%bias	|bias|	t-test	
terap	Unmatched	0.46	0.31	29.8		1.8	*
	Matched	0.44	0.50	-12.9	56.7	-0.69	
nocrops	Unmatched	2.35	2.45	-6.1		-0.36	
	Matched	2.35	2.55	-12.2	-101.3	-0.74	
family	Unmatched	6.37	6.00	13.4		0.8	
	Matched	6.21	5.92	10.5	21.4	0.58	
kids	Unmatched	2.66	2.13	30.5		1.82	*
	Matched	2.52	2.27	13.7	54.9	0.8	
childout	Unmatched	0.83	0.71	28.3		1.68	*
	Matched	0.82	0.81	3.9	86.4	0.23	
agep	Unmatched	43.45	45.20	-11.5		-0.69	
	Matched	43.26	43.24	0.1	99.1	0.01	
agep2	Unmatched	2128.40	2261.60	-8.8		-0.53	
	Matched	2116.00	2080.10	2.4	73.1	0.14	
viajarstm	Unmatched	13.79	13.85	-0.7		-0.04	
	Matched	13.03	12.32	7.2	-985.4	0.45	
pension	Unmatched	0.26	0.26	-0.2		-0.01	
	Matched	0.23	0.27	-8.9	-4932.3	-0.45	
collntfp	Unmatched	5.22	4.13	34.9		2.07	**
	Matched	5.18	6.06	-28.4	18.6	-1.44	
soldntfp	Unmatched	0.65	0.65	-0.3		-0.02	
	Matched	0.65	0.53	8.5	-2835.5	0.51	
msch	Unmatched	11.22	16.90	-29		-1.71	*
	Matched	11.08	10.70	1.9	93.3	0.16	
income97	Unmatched	953.71	889.35	5		0.3	
	Matched	914.85	857.99	4.4	11.7	0.26	
cropland	Unmatched	1.84	1.95	-6.3		-0.37	
	Matched	1.83	1.63	10.7	-70.9	0.78	
cattle	Unmatched	1.35	2.84	-18.7		-1.08	
	Matched	1.40	0.12	15.9	14.9	2.46	**
fphys97	Unmatched	-7.04E-02	-0.12	5.2		0.31	
	Matched	-0.11	-0.31	20.6	-292.5	1.31	
fliq971	Unmatched	0.17	0.33	-12.6		-0.73	
	Matched	0.16	1.12E-02	11.5	9	1.27	

As we can see, the matching process diminished the differences between the covariates pre-intervention. Prior to matching, treated households had better soil, more children, more children living outside the household, higer collection of NTFPs. After

matching all covariates are balanced except for number of heads of cattle owned in 1997 (significantly different at the 5% significance level).

Finally, table 3C3 below shows the same results for kernel matching, where we see that matching also decreased bias.

Table 3.C.3 Treament and control characteristics before and after matching (based on kernel matching).

		Mean			%reduct		
Variable	Sample	Treated	Control	%bias	|bias|	t-test	
terap	Unmatched	0.46	0.31	29.8		1.8	*
	Matched	0.45	0.54	-17.8	40.3	-0.92	
nocrops	Unmatched	2.35	2.45	-6.1		-0.36	
	Matched	2.38	2.39	-0.5	91.8	-0.03	
family	Unmatched	6.37	6.00	13.4		0.8	
	Matched	6.12	5.81	11.1	17.2	0.6	
kids	Unmatched	2.66	2.13	30.5		1.82	*
	Matched	2.38	2.21	9.9	67.6	0.58	
childout	Unmatched	0.83	0.71	28.3		1.68	*
	Matched	0.81	0.84	-8	71.8	-0.47	
agep	Unmatched	43.45	45.20	-11.5		-0.69	
	Matched	44.03	43.57	3.1	73.4	0.16	
agep2	Unmatched	2128.40	2261.60	-8.8		-0.53	
	Matched	2189.10	2115.90	4.8	45	0.26	
viajarstm	Unmatched	13.79	13.85	-0.7		-0.04	
	Matched	12.90	13.12	-2.2	-237.9	-0.13	
pension	Unmatched	0.26	0.26	-0.2		-0.01	
	Matched	0.21	0.19	2.9	-1542.4	0.17	
collntfp	Unmatched	5.22	4.13	34.9		2.07	**
	Matched	5.09	5.64	-17.7	49.3	-0.91	
soldntfp	Unmatched	0.65	0.65	-0.3		-0.02	
	Matched	0.69	0.64	3.5	-1107.9	0.19	
msch	Unmatched	11.22	16.90	-29		-1.71	*
	Matched	11.44	12.45	-5.1	82.3	-0.34	
income97	Unmatched	953.71	889.35	5		0.3	
	Matched	881.09	927.23	-3.6	28.3	-0.2	
cropland	Unmatched	1.84	1.95	-6.3		-0.37	
	Matched	1.84	1.75	5.1	18.8	0.32	
cattle	Unmatched	1.35	2.84	-18.7		-1.08	
	Matched	1.47	0.69	9.7	48	0.85	
fphys97	Unmatched	-0.07	-0.12	5.2		0.31	
	Matched	-0.10	-0.16	6.1	-17.2	0.32	
fliq971	Unmatched	0.17	0.33	-12.6		-0.73	
	Matched	0.19	0.11	6.1	51.9	0.48	

- **Ever participated:**

Here I define the treatment as households that have ever participated in the CBEs (regardless of whether they were participating in 2006). Table C4 presents the results of the propensity score logit:

Table 3.C.4 Propensity score logit for having ever participated in the CBEs.

Description	Variable	Estimate		Std. Err.
Good soil quality dummy (terra preta)	terap	0.35		0.33
Number of crops planted	nocrops	7.19E-02		0.11
Household size	family	-0.11		0.10
Number of children	kids	-0.56	**	0.29
Number of children living outside TNF	childout	-0.41		0.67
Age of male head of household	agep	-4.71E-02		6.24E-02
Age of male head of household squared	agep2	5.67E-04		6.10E-04
Number of trips to city per month	viajarstm	-6.68E-02	*	3.55E-02
If receives pension	pension	-0.80		0.50
If collected NTFPs	collntfp	0.15	**	6.42E-02
If sold NTFPs	soldntfp	-0.12		0.14
Interaction (pension*kids)	penkids	0.55	**	0.28
Interaction (kids*childout)	kidsout	0.42	*	0.25
Interaction (viajarstm*kids)	viajkids	2.09E-02	*	1.15E-02
Intercept	_cons	1.95		1.42

The model has a pseudo R^2 of 8% and is significant at the 10% significance level. Here we can see that having more children and making more trips to the city are associated with a lower probability of ever participating in the CBEs. In the other hand, prior experience collecting NTFPs, and interaction terms increase the probability of participation. Next, I perform the matching using the propensity score. Table 4 presents the mean comparison between treatment and control households in 1997 (both prior and after matching):

Table 3.C.5 Treatment and control characteristics for households who ever participated in CBEs before and after matching (based on matching with two nearest neighbors).

		Mean			%reduct		
Variable	Sample	Treated	Control	%bias	|bias|	t-test	
terap	Unmatched	0.41	0.3253	17.2		1.15	
	Matched	0.4	0.48421	-17.1	0.6	-1.14	
nocrops	Unmatched	2.54	2.4217	7.4		0.5	
	Matched	2.5579	2.5684	-0.7	91.1	-0.04	
family	Unmatched	6.18	6.0482	4.5		0.3	
	Matched	6	5.4947	17.4	-283.3	1.21	
kids	Unmatched	2.43	2.1446	15.7		1.06	
	Matched	2.2526	2.0842	9.3	41	0.68	
childout	Unmatched	0.78	0.72289	13.2		0.89	
	Matched	0.76842	0.83158	-14.6	-10.6	-1.09	
agep	Unmatched	44.1	45.048	-6.3		-0.42	
	Matched	44.2	42.463	11.5	-83.2	0.79	
agep2	Unmatched	2176.2	2245.3	-4.7		-0.31	
	Matched	2189.9	2024.8	11.1	-138.7	0.77	
viajarstm	Unmatched	13.84	13.783	0.6		0.04	
	Matched	13.389	13.463	-0.8	-29.6	-0.06	
pension	Unmatched	0.27	0.26506	0.9		0.06	
	Matched	0.23158	0.18947	7.7	-752.4	0.6	
collntfp	Unmatched	4.99	4.0602	29.6		2.01	**
	Matched	4.8947	4.3789	16.4	44.5	1.15	
soldntfp	Unmatched	0.67	0.63855	2.4		0.16	
	Matched	0.68421	0.89474	-16.4	-569.5	-1.13	
penkids	Unmatched	0.56	0.20482	30.1		1.96	**
	Matched	0.30526	0.14737	13.4	55.5	1.46	
kidsout	Unmatched	2.04	1.5301	26.5		1.78	*
	Matched	1.8421	1.9158	-3.8	85.5	-0.28	

From table 3.C.5 we can see that there were a few differences between the treated and control households prior to matching (in terms of collection of NTFPs, interaction of number of children and receiving pension and number of children and number of children living outside the TNF). These differences are corrected for after the matching, as none of

the means for 1997 characteristics is significantly different between control and treatment groups after matching. Finally, table 3.C.6 shows the same statistics pre and post kernel matching:

Table 3.C.6 Treatment and control characteristics for households who ever participated in CBEs before and after matching (based on matching with two nearest neighbors).

		Mean			%reduct		
Variable	Sample	Treated	Control	%bias	|bias|	t-test	
terap	Unmatched	0.41	0.33	17.2		1.15	
	Matched	0.40	0.40	-0.9	94.5	-0.06	
nocrops	Unmatched	2.54	2.42	7.4		0.5	
	Matched	2.56	2.38	11.1	-50.9	0.78	
family	Unmatched	6.18	6.05	4.5		0.3	
	Matched	6.00	6.02	-0.6	87.5	-0.04	
kids	Unmatched	2.43	2.14	15.7		1.06	
	Matched	2.25	2.26	-0.4	97.7	-0.03	
childout	Unmatched	0.78	0.72	13.2		0.89	
	Matched	0.77	0.78	-2.7	79.1	-0.2	
agep	Unmatched	44.10	45.05	-6.3		-0.42	
	Matched	44.20	46.00	-11.9	-89.4	-0.79	
agep2	Unmatched	2176.20	2245.30	-4.7		-0.31	
	Matched	2189.90	2363.10	-11.7	-150.4	-0.76	
viajarstm	Unmatched	13.84	13.78	0.6		0.04	
	Matched	13.39	13.22	1.8	-196.3	0.14	
pension	Unmatched	0.27	0.27	0.9		0.06	
	Matched	0.23	0.25	-3.1	-246.5	-0.23	
collntfp	Unmatched	4.99	4.06	29.6		2.01	**
	Matched	4.89	4.87	0.9	96.9	0.06	
soldntfp	Unmatched	0.67	0.64	2.4		0.16	
	Matched	0.68	0.75	-5.4	-122.2	-0.36	
msch	Unmatched	10.70	17.07	-32		-2.18	**
	Matched	10.62	18.41	-39	-22.2	-2.68	***
income97	Unmatched	947.96	892.00	4.1		0.28	
	Matched	924.02	802.45	8.8	-117.2	0.66	
cropland	Unmatched	1.96	1.96	0.2		0.01	
	Matched	1.98	1.96	1.3	-484.9	0.09	
cattle	Unmatched	1.66	2.74	-13.1		-0.9	
	Matched	1.72	1.81	-1.1	91.8	-0.09	

4 Microenterprises after the funding ends: Two Steps Forward and One Step Back

4.1 Abstract

Community-based natural resource management has become central to international conservation efforts in recent decades, as part of a pro-poor environmental agenda (Adams et al. 2004; Berkes 2007). The goal is to simultaneously improve local welfare and provide incentives for conservation through economic development based on sustainable use of natural resources. A key assumption is that the best way to accomplish this is by working with groups of households in communities, associations, or other types of networks. Meanwhile, as community-based conservation and development projects continue to attract funding, there are persistent concerns about the lack of long-term impacts on local welfare and environmental conservation (McShane and Wells 2007; Garnett et al. 2007). In fact, this lack of sustainability beyond the period of external funding is an important problem for international aid in general (Kareiva et al. 2008). Given the emphasis placed on cooperation among households in the implementation phase, I hypothesize that social networks are also critical to the process of sustaining (or abandoning) activities promoted or discouraged by the projects. I examine this hypothesis in the context of a pilot program to engage resident communities in sustainable management of a national forest in the Brazilian Amazon.

Social networks are important sources of information, especially in remote rural areas where information asymmetry characterizes most economic exchange (Conley and Udry,

forthcoming and Loader, 1997). Therefore households create networks comprised of other households they trust to exchange information. While the role of social networks in technology adoption has been the subject of much recent research, this chapter is one of the first assessments of social networks as drivers of "disadoption" or exit from programs. In this chapter, I consider a household's social network to be comprised of its immediate neighbors, fellow members of the community association, and two stated groups of contacts (who the household talks to about projects, and who the household would choose to work with). The individual network matrices for each of these definitions are added together and row normalized. The result is a directed social network (by construction), which allows identification of endogenous and exogenous network effects (Bramoulle et al, 2009).

In this chapter, I examine the influence of social networks on household decisions of whether to continue participating in forest-based community micro-enterprises after external funding ends, and whether to reduce land area in slash-and-burn cultivation. These are two key outcomes of concern to the ProManejo pilot program, which sought to engage local housholds in the sustainable management of the TNF in the Brazilian Amazon. I estimate the probability of reducing agricultural area between 2006 and 2008 as a function of household characteristics (including participation in the micro-enterprises) and network characteristics (including agricultural area and its determinants). Estimation results are reasonably consistent across different empirical approaches to

identifying social network effects and the relation between microenterprise participation and trends in agricultural land area. Social networks - but not microenterprise participation per se - influence the decision whether to decrease agricultural land area. With regards to the microenterprises implemented, results are heterogeneous and vary depending by type of project. While some achieved success in diverting labor from agriculture, others have a higher retention of participants. This result has both policy and academic importance. In the policy arena it shows that actions of households in the same network matters both in terms of decreasing agriculture and project participation and therefore outcomes of community-based interventions should take this into account. For example, these results indicate that, similar to conventional wisdom, it is important to get weak ties (who are part of many networks) to participate. The results also show that conditional government transfers may help divert people from agriculture but health of network (not just household) is a constraining factor. Finally, results show that outmigration of children decreases the probability of decreasing land in agriculture and does not impact microenterprise participation.

Keywords: social networks, microenterprise, land use, Amazon

4.2 Introduction

Community-based natural resource management has become central to international conservation efforts in recent decades, as part of a pro-poor environmental

agenda (Adams et al. 2004; Berkes 2007). The goal is to simultaneously improve local welfare and provide incentives for conservation through economic development based on sustainable use of natural resources. A key assumption is that the best way to accomplish this is by working with groups of households in communities, associations, or other types of networks. In the case examined in this chapter, donors invested in community enterprises that process and market non-timber forest products. This is an example of "microequity," which Pretes (2002) defines as startup grants and equity finance with the goal of supporting the implementation of microenterprises even among the very poor.

While community-based conservation and development projects continue to attract funding, there are persistent concerns about the lack of long-term impacts on local welfare and environmental conservation (McShane and Wells, 2007; Garnett et al, 2007). In fact, this lack of sustainability beyond the period of external funding is an important problem for international aid in general (Kareiva et al. 2008). Given the emphasis placed on cooperation among households in the implementation phase, I hypothesize that social networks are also critical to the process of sustaining (or abandoning) activities promoted or discouraged by the projects. I examine this in the context of a pilot project to engage resident communities in microenterprises to promote sustainable management of a national forest in the Brazilian Amazon.

Social networks are increasingly recognized as a pervasive and important influence on the economic behavior of rural households in developing countries, similar to any other agents who make decisions in the face of incomplete information. Information asymmetry leads to inefficiencies that agents try to overcome by creating informal ties, links and contracts (Loader, 1997). These are usually trust based and therefore depend on the knowledge that each agent has about others to mitigate risk related to trusting (Dasgupta, 2003, Fafchamps, 2003, Fafchamps and Minten, 2002). Accordingly, agents create networks comprised of other agents they trust to exchange information. These networks are necessary for the survival of an agent and the better connected the agent, the greater the amount and reliability of information available to her (Conley and Udry, forthcoming).

In this chapter, I analyze the role of networks among households in the TNF in the Brazilian Amazon. Specifically, I analyze the impact of networks on household decisions about agricultural land area and microenterprise participation. Traditionally, households in this region earn their livelihood through slash-and-burn agriculture. However, for the communities inside the National Forest, this has been a source of conflict with forest managers, whose mandate is to protect the forest from clearing and fire. Under the Pilot Program for the Brazilian Rainforest, over $1.4 million was invested in the search for more sustainable alternatives for these communities, including microenterprises aimed at shifting land use away from slash-and-burn and projects to improve community

organization and household welfare. In this chapter I evaluate what happened after this external funding ended in 2006: did local households remain engaged in the microenterprises? Did they shift back into slash-and-burn agriculture? In particular, I evaluate the role of social networks in these decisions. My hypothesis is that the decisions made by a household's social network influences the household's own decision.

The next section presents a literature review on papers estimating social networks effects in developing country applications. After that, section 3 discusses the motivation for an empirical model of project participation and agricultural land use influenced by social networks. The data used to estimate this model are described along with the study site in section 4. Empirical results are provided in section 5, and finally conclusions and policy implications in section 6.

4.3 Literature review

The economic literature on social networks is recent but rapidly growing. I reviewed empirical studies of social networks in rural areas of developing countries, which have been published in peer-reviewed journals or as Ph.D. dissertations since 1992 (See Appendix 4.1). This starting point corresponds to one of the earliest and highly influential publications in this field, by Case (1992). She considers the impact of neighbors' characteristics and attitudes on a farmer's decision to adopt a new agricultural technology. She concludes that neighbors' characteristics and expected profits play an

important role in determining a farmer's decisions and that ignoring these social effects would lead to overestimates of the effects of own-farm characteristics on the adoption decision.

All of the studies reviewed here also take another seminal article by Manksi (1993) as their starting point. Manski identifies three separate mechanisms through which an individual's actions may be correlated with the actions of her peers, and discusses whether and how these different mechanisms can be identified. These mechanisms are:

i) endogenous effect (individual behavior varies with the behavior of the group);

ii) exogenous effects (individual behavior varies with the exogenous characteristics of the group); and

iii) correlated effects (individuals in the same groups tend to behave similarly because they have similar individual characteristics or face similar institutional environments).

Endogenous and exogenous effects are the most policy relevant since they affect, for example, how policy interventions are implemented, and how technology is adopted. Literature follows this argument and usually papers try to estimate endogenous and exogenous effects while sometimes controlling for correlated effects. However, separating these effects is considered the key empirical challenge by most authors in this field. Next, I summarize some key dimensions of these 35 studies that I reviewed, starting with the outcomes that they model. Nearly half (46%) of the studies examine

adoption of new technologies promoted for environmental or other social reasons while another 26% studied risk sharing and shock smoothing.

In terms of networks, most of the studies focus on the role of networks in exchanging information, but some (e.g. Luke and Munshi, 2007; Dercon and Krishnan, 2003; Fafchamps and Lund, 2003; Fafchamps, 2003) also acknowledge that networks can operate through mechanisms such as economies of scale in transportation, access to credit, and prices. Fafchamps and Gubert (2007) provide an overview of different pathways through which social networks can affect decisions. The studies also vary in terms of how they define networks. The earliest studies by Case (1992), Foster and Rosenzweig (1995), and Besley and Case (1997) define networks as all members of the same community or village. Other measures include caste, religion, ethnicity, schools, producer associations and other formal group membership. Twenty-five percent of the studies use self-reported networks, identified using answers to questions such as "who do you talk to," "who do you send money to," "who would you go to for help." These questions are more likely to identify "directional networks" since households do not have to cite each other. This allows for more estimation flexibility as will be discussed later. All of the studies reviewed assume the same strength of ties among all households in the same network, and none of the studies have panel data on networks

Finally, the studies employ a variety of empirical modeling strategies. 54% assume that networks are exogenous, and the others use linear in means models with instrumental variables. The Brock and Durlauf approach I use later on is used by Dickinson and Pattanayak (forthcoming) as well as other non-developing country applications that are not included in the literature review. Another model I estimate is the linear-in-means model that is also used by Moffitt (2001) and Bajari et al. (2006).

Appendix 4.A summarizes the literature reviewed on empirical analyzes of how social networks influence economic decisions in developing countries. It shows the problem discussed in the article, the theoretical and empirical strategies employed, network definition used and whether the networks were considered exogenous or endogenous.

This chapter extends the existing literature on social networks in several ways. First, rather than adoption of new technology, the motivating question for this chapter is what occurs after external funding for micro-enterprise projects ends. Thus, the question is why or why not do households give up on a new "technology." The collapse of externally funded projects is a very common problem in rural development (Kareiva et al. 2008). I examine two key outcomes of concern to the project and National Forest managers: participation in the microenterprise project and changes in agricultural land area (which is directly linked to deforestation). Second, having asked directly about

producer's networks, I combine both stated networks and more conventional networks (defined spatially and by membership in formal associations) to construct a general measure of social networks in the communities.

Economic theory and empirical work show that relative expected utility from different activities are the key driving force in labor use; therefore, it is critical to understand the factors that determine relative preference between agriculture and other activities. The point of this chapter is to evaluate whether social networks are among the factors affecting expected utility from labor allocation. Basically, a household with a better network of contacts should have better information about labor allocation therefore their decisions should reflect the importance of their network. As such, I argue that information from social networks shapes expectations of returns to different activities, but I do not attempt to disentangle this information effect from other potential effects.

The key hypothesis to be tested is whether networks among household influence the decision regarding household labor allocation. Therefore, this research will shed light both on the key determinants of household participation in externally funded micro-enterprises after the funding ends and their land use choice.

4.4 Motivation and empirical model

The motivation for the empirical model focuses on household decisions about allocation of labor, because it is their most important fully fungible resource. A household's social network supplies information about the benefits of uses of time (such as microenterprises), including both productivity and profitability (or potential and challenges of commercialization). The mechanism by which a household's social network influences its expectations and therefore its time allocation is detailed in an appendix. In practical terms, the empirical model relates project participation and agricultural land use to (a) household and community characteristics, (b) project participation and agricultural land use of the household's network, and (c) each other.

In Appendix 4.B, I draw on Munshi (2004) to develop a simple theoretical framework to motivate how a household's social network influences its labor allocation between agriculture and non-agriculture. This framework represents the typical household living on the forest margins in the Brazilian Amazon. Such households are very familiar with traditional agriculture production, but the potential income that they could earn from other activities is less well known. Social networks can diminish uncertainties by providing a household with information on the expected returns to labor from those other activities.

Expectations are formed based on the household's external information and decisions of their network about whether or not to participate in microenterprises, possibly with some time lag. Appendix 4.B expands the model to specify how households use information from their social networks to update their expectations of the returns to labor allocated to the microenterprise. To the extent that households are similar (or differences that affect Y can be observed), then households in the same social network will rely on information about each other's participation. If returns are heterogeneous across households, then social network's decisions matter less. Thus, even this very simple model generates testable hypotheses about the influence of social networks.

Turning to the empirical model, household decisions about microenterprise participation would be influenced by household characteristics and by expectations of the output of microenterprises. In parallel, decisions about agriculture are influenced by the same household characteristics and the agricultural output. According to this model, the relationship between microenterprise participation and labor allocated to agriculture is straightforward: there is a direct trade-off such that participating in the project must mean spending less time on agriculture. In reality, there are other uses of time. The relationship between microenterprise participation and the land area dedicated to agriculture is further muddied by the fact that labor and land are not strictly Leontief inputs into agricultural production, but rather there is some scope for substitution. To allow for the different possible relationships between agricultural land and microenterprise participation, the

empirical models include lagged land in agriculture in 2006 and participation in the three largest microenterprises in 2006 as measures of accumulated human capital and potential determinant of both participation in microenterprises and decreasing agriculture[12]. To reflect the fact that these are really simultaneous decisions, I also estimate a bivariate probit with correlated errors.

Because both microenterprise participation and agricultural land use are determined by a common labor allocation problem, they are influenced by social networks through their influence on expectations of output (*Y*). As will be further discussed below, in the analysis I include the endogenous effect (expectations of *Y* are influenced simultaneously by other households decisions about *Y*) and two kinds of exogenous effects (the network's previous decisions about *Y* and related *Y*s that may also influence decisions by offering opportunities to learn, encouraging either conformance or diversification). I use a two year lag which is consistent with the fact that funding for the microenterprises ended two years prior to the current period, and thus was a particularly telling time in terms of household expectations of returns from the microenterprises themselves (distinct from program activities such as workshops and consultant visits). For agriculture, the two year lag is also logical because the main agricultural product (manioc) takes two years to mature, so that households are now reaping harvests resulting from land use decisions made two years prior). Finally, networks may influence decisions

[12] I also tested models with the other contemporaneous (rather than lagged) outcome as a determinant (with similar results), and models of change in agricultural land area as a function of a general dummy for participation in any of the micro-enterprises in 2006 (not significant).

through other pathways, e.g. to the extent that there is sharing or risk management through the same networks.

I divide household characteristics into physical capital, labor, human capital and accessibility. Since agriculture is mostly labor intensive I use transportation variables as a measure for physical capital (the number of trips to the city per month) as well as a soil quality dummy. Labor is represented by size of household, proportion of adults (over 10 years old), proportion of children under 5 years of age, and age of heads of household. I also include days not working due to health problems or taking care of another household member as a measure of labor quality. Additional external and largely exogenous income sources such pension paid by the government to anyone over 65 years of age, government transfers for keeping children under the age of 15 in school and remittances (measured by number of children living outside the TNF) are also expected to influence households' decisions. Human capital is measured by experience in agriculture and non-agricultural income opportunities (represented by amount of land in agriculture, heads of cattle, and microenterprise participation in 2006). I also consider age of heads of household and proportion of life spent in the community as measures of experience. Finally, access is represented by two community characteristics: a dummy for whether the community can be accessed by road (vs. boat only) and distance from the regional market center.

4.5 Data and context

Data for this study come from household surveys in the Tapajós National Forest (TNF). This protected area, created in 1974, is considered an example for other protected areas in the Brazilian Amazon. For example, it hosted the first timber concession contract in Brazil in 1999 (Bacha and Rodriguez, 2005), and it was recently cited in the New York Times as an example of sustainable forest development (Friedman, 2009). Thus, this analysis describes household decisions about project participation and agricultural land under a "best case scenario" of significant external funding and support for incorporating community interests into protected area management. The TNF is clearly not representative of the typical protected area in the Brazilian Amazon, but it is a very influential and well-know case study from which many have sought to draw policy implications (Bacha and Rodriguez, 2005).

The TNF has served as a pilot and demonstration site for the Pilot Program for the Amazon Rainforest, launched in 1999 with funding from the G7 to implement and support pilot programs and demonstrations to promote sustainable forest management systems in the Amazon (Dias, 2001). It was executed by IBAMA (the Brazilian Institute of Environment and Renewable Natural Resources – Forest Division) and the Environmental Ministry (Biodiversity and Forests division/ National Forests Program). Within the broad program of ProManejo, there were 4 components: i) strategic studies and analysis; ii) promising initiatives; iii) monitoring and control; and iv) Management of

the Tapajós National Forest (IBAMA, 2009). In the Tapajós, one key activity was to develop microenterprises to process and sell forest products to provide the 1200 families residing in the National Forest with an alternative livelihood to traditional slash-and-burn agriculture. These microenterprises were explicitly community projects, requiring cooperation among households in community associations, but also implicitly relying on information sharing and mutual support among households for moving beyond traditional activities. PPG7 funding ended in 2006, and social networks could have played an important role in determining whether and which activities are maintained.

A second reason for selecting this study site was the opportunity to use previous rounds of household survey data. In 2008, I surveyed 355 households in 22 communities inside the National Forest (see map). Surveys had been carried out in the same communities (with the same households) in 1997 (2 years before the financing of microenterprises began) and 2006 (when the external funding ended) forming a panel of 230 households. Since very limited information on networks was collected in the previous rounds I cannot use panel information on network membership, similar to all previous literature on networks in rural areas of developing countries (eg. Conley and Udry, forthcoming, who also have a panel of outcomes but not networks). However, the panel can be used in three ways: first, the panel can be used to explore potential confounders that might be driving both network participation and outcomes of interest; second, I can construct lagged measures of network behavior to better reflect the time

frame for decision-making about agricultural areas, and third, I can examine differences in outcomes to sweep out time-invariant unobservables (and correlated effects) and more precisely estimate the effects of networks, microenterprises, and other household and community characteristics.

The 22 communities included in the survey varied in size from 10 to 63 households. In 2008, a team of six enumerators working under my supervision interviewed as many as possible of the households who had been interviewed in 1997 and 2006. Thus, the sampling was based on the previous samples. In 1997, a random sample of households stratified by community was interviewed (excluding two communities due to time constraints and participation in pre-test). In 2006, the sample consisted of the same and descendent households and randomly drawn replacements. The final sample of 167 households refer to the panel for which I have network information (excluding households with self reported links that are not part ofmysample). Table 4.1 provides descriptive statistics for the households.

Households in the Tapajós National Forest obtain most of their income from slash-and-burn agriculture. In the context of low population density, traditional slash-and-burn agriculture is part of a sustainable land use system in which households shift from one area to the next over very long cycles. Nonetheless, inside a protected area, the loss of mature forest and the use of fire are the main forms of environmental degradation,

followed by shorter fallow periods associated with current practices and the introduction of cattle, which do not allow forest regeneration. Under rules in place in 2008, households are officially limited to clearing 2 hectares (8 tarefas)[13] of agricultural land per household per year (plus an extra 2 ha for each child over 18 years of age). Discussions with key informants and community meetings indicated that most local people do not consider this constraint binding, both in the sense that most households are cultivating less land and that they do not expect to be penalized if they exceed the limit. In addition to this regulatory approach, there is a strong policy mandate to develop alternative means of income generation more compatible with forest conservation as a means to attract households out of agriculture. In fact, the survey demonstrates that land used per household for agriculture in the TNF has been decreasing. While the average in 1997 was over 6 tarefas of agricultural land per household, in 2006 this had decreased to 3.85 tarefas and in 2008 to 3.3 tarefas per household. Between 2006 and 2008 51% of households decreased the land they had in agriculture, and among these, the area reduction was of 44% on average (same pattern holds in terms of land area per capita). One key outcome of interest in this chapter is the decision whether to decrease land in agriculture between 2006 and 2008.

The ProManejo Tapajós component sought to divert labor from agriculture towards environmentally sustainable activities such as medicinal oils, seeds, ecological leather (made of latex), and ecotourism in the years 1999 to 2006. As shown in Table 4.1,

[13] Tarefa is the local land measure, which we keep throughout the text. 1 tarefa = 0.25 hectares

36% of households participated in microenterprises in 2006 (of which 17% of households were participating in the ecotourism, 19% in the furniture, 4% in oil and 3% in ecotourism microenterprises). By 2008, only 23% of households participated in at least one microenterprise. This is the second key outcome considered in this chapter: whether or not a household is still participating in the microenterprise projects in 2008, two years after external funding ended.

In terms of physical capital, 35% have good soils - "terra preta de índio," very rich in organic matter and a major determinant in agricultural productivity in the region (Glaser et al, 2001), households are located on average 14 minutes walking from the community center (with a maximum of 2.5 hours) and average 2.2 trips to the city per month. In terms of labor, on average a household is comprised of 6 people, and the average age of the heads of household is 46 years. The age distribution within the household is as follows: 11% are less than 5 years old and 73% are over 10 years of age. In this study we consider children over the age of 10 as a working adult since they tend to work with their parents from this age on. Days not working (DNW) is the sum of days that either head of household could not go to work due to sickness or taking care of someone else sick in the household (therefore, the maximum is 60, if both heads of household could not work). On average a household misses 11 days of work a month.

Human capital in these communities is measured mainly by experience, so agricultural area and prior participation in microenterprises (discussed above) are considered in this category, as well as the proportion of their life that the heads of household have spent in the community (on average 79%). Finally, a household can also receive exogenous income in the form of government pension (30% of the sample, receiving on average R$185 per month), government transfers (67% of the sample, receiving on average R$65 per month[14]) and children living outside of National Forest (on average 1.5 per household with a variation from 0 to 9). Finally, some community variables are included in the table, such as distance in hours by boat from the community to the main urban center in the region (Santarém) (on average 5.8 hours) and whether the community has road access (83% of sample).

The definition of social network used in this chapter is similar to that of Carter and Maluccio (2002), Bandiera and Rasul (2006) and Nepal et al (2007). These authors combine different definitions of networks to build a single social network for each household. Having a unique measure of social network allows for each household to receive and rely on information from their preferred network of contacts (e.g., neighbors, friends, and confidants), in addition to the information generally available to all members of a certain group. Therefore, I build four NxN interaction (or distance) matrices for:

[14] Most of this transfer comes in form of "bolsa família", a program that gives approximately R$35 for each child kept studying. However, not all households with children in school receive this payment (although there is no evidence of targeting at the household level).

- Association membership: a_{ij} =1 if i and j are members of the same community association, 0 otherwise;
- Geographic neighbors: a_{ij} =1 if i and j live next to each other, 0 otherwise;
- Project: Who each household talks to about microenterprises: a_{ij} =1 if i said he talks to j about microenterprises, 0 otherwise;
- Choice: Who each household would choose if they were to work on a "project" together: a_{ij} =1 if i said he talks to j about projects, 0 otherwise.

The last two definitions of networks do not require reciprocity (i.e. *i* might cite *j* as a source of information about microenterprises but *j* might not cite *i*), i.e., they are directed social networks (Bramoulle et al, 2009). The four matrices were then summed and row normalized, so that each household has a weighted average of importance of information links. The resulting normalized matrix is used as the network matrix for all analyses in this chapter. The four different networks were elicited in three ways in the survey:

(a) open-ended question about who is in network (i.e. "Who are your direct neighbors?" and "Who would choose to work with?"),

(b) open-ended question followed by query about random sub-sample of households (i.e. "Who do you work with in the microenterprises?"), and

(c) closed-ended question about whether a member of a pre-defined network (i.e. "Are you a member of the community association?).

Not all households indicated as part of a household's network are necessarily part ofmyrandom sample of households. For neighbors, the average number of responses per household was 2.25[15] of which 0.83 (37% of reported links) are inmysample. For the choice network (who would you choose to work with) the average number of responses was 5.25 out of which 2.14 (42% of reported links) are in my sample. The project network (who do you talk to about microenterprises) had an average of 3.95 responses per household of which 2.58 (65% of reported links) are also in my sample[16]. Finally, on average there are 5.97 within sample households as members of the same association.

Following Quisumbing et al (2008) I investigate which characteristics influence network formation and compare those across different types of networks, some being "more exogenous" than others. Table 4.2 presents the results of OLS regressions which explain which household characteristics determine how many links a household has for each type of network. Overall, there are few significant determinants of size of network. For example, the first network presented is the neighbors' matrix. We can see that having more siblings living in the community and living closer to the community center is associated with having more neighbors. This may be due to siblings living in clusters and therefore acknowleding more neighbors. Similarly, people living further from the community center are more dispersed and therefore have fewer neighbors. Wealthier

[15] This number is lower than the expected 4 neighbors because i) most households have the river at the front or back of their house, which would decrease the maximum number of neighbors to 3, and ii) if the next household is far away, a respondents may not list as a neighbor.

[16] This percentage is higher since each household was also asked whether they talked about project to a random set of other household in my sample.

households are more likely to be members of the community association. Having more agriculture and fewer siblings in the community is associated with talking to fewer people about microenterprises. It is intuitive that households more engaged in agriculture would talk to fewer people about microenterprises, but the connection with fewer siblings is not obvious. Finally, the older the household head, the fewer people the respondent would choose to work with. Overall these results seem to indicate that there are few observable characteristics that determine participation in these social networks. Based on this finding, we treat the network as exogenous in the empirical analysis presented in the next section.

4.6 Empirics

There are several different ways to approach estimation of social network effects on household decisions like microenterprise participation and change in crop area. Most broadly, these can be divided into two categories: spatial econometrics (see review in Anselin, 2002) and network effects (see Bramoulle et al. 2009). In this chapter I follow the network effects literature, but to provide context, I first briefly describe these two different branches of the literature. Both are based on Spatial AutoRegressive (SAR) models, and accordingly they employ row-normalized interaction weight matrices (e.g., to obtain the weighted average of a household network's decisions). However, they differ in that they seek to identify different types of network effects, make different assumptions, and use different terminology.

Following Manski (1993), the literature that focuses on network effects starts with the following model:

(iii) $$y = \delta Wy + \gamma Wx + \beta X + e$$

This includes two social networks effects: δ represents the endogenous effect and γ the exogenous effect, while controlling for correlated effects by modifying the error specification. It is important to note that x is a subset of X which includes variables that are likely to be influential as exogenous network effects. In contrast, the spatial econometrics literature usually begins with the specification:

(iv) $$y = \delta Wy + \beta x + \eta$$

where the coefficient on Wy (δ) represents spatial lags (network effects) and the error includes a spatial component to account for correlated effects. Comparing these models we can see that the latter ignores exogenous effects (that an agent's outcome can be affected by her network's covariates). If exogenous factors also drive y the estimate of δ will be confounded; otherwise the exogenous effects would be relegated to the error term. In spatial econometrics, W is always considered exogenous[17], which means that equation

[17] The spatial econometric approach evolved from thinking about W as exogenous physical space, however, people usually have some degree of choice about where they live, which is considered by another branch of the literature on sorting from Tiebout (1956) and Schelling (1969 and 1970) to Fujita (1999), Bayer and Timmins (2005), and Ioannides and Zabel (2003).

(ii) can be solved for *y* on the left hand side and estimated as a non-linear maximum likelihood model (which requires inversion of the *Wy* matrix).

There is a large and evolving literature on how to estimate network effects models, focused on how to obtain separate estimates for the three different effects identified by Manski (endogenous, exogenous, and correlated). The main difficulty is the reflection problem "that arises when a researcher observing the distribution of behaviour in a population tries to infer whether the average behaviour in some group influences the behaviour of the individuals that comprise the group" (Manski, 1993 p.532). Two leading approaches are the binary choice model (Brock and Durlauf, 2001a, 2001b, 2007) and the linear-in-means or, as I will refer to, the Moffit model (Manski, 1993, Moffitt, 2001).

The Brock and Durlauf (2001a) identification strategy relies on non-linearities embedded in the binary choice model to separate endogenous from exogenous effects under the assumption of no correlated effects. Brock and Durlauf (2007) use the same framework but extend the model to allow for correlated effects. They argue that many outcomes of interest are binary in nature and that "binary choice models of social interactions possess a number of interesting theoretical properties, such as multiple equilibria and phase transition (the potential for qualitative changes in the properties of the model to change with small changes in model parameters) that are not present in

linear-in means models" (Brock and Durlauf, 2007: 53). Their model is implemented by simply including the outcome variable network average as a regressor in a non-linear (probit) regression. In the application considered here, the two binary dependent variables are a) whether the household decreased land in agriculture (1 if agricultural area decreases from 2006 to 2008, 0 otherwise) and b) whether the household is participating in the microenterprise program (1 if not participating in any microenterprise in 2008, 0 otherwise). This approach assumes that individual error terms are independently and identically normally distributed and have a likelihood function of the form:

(v) $P(y_i = 1 \mid x_i, Wx, Wy) = 1 - F(-\beta x_i - \delta Wx - \gamma Wy)$

where *i* indexes individuals and *F(.)* is the cumulative density function. Regressors include household (*x*) and network characteristics (*Wx*) and the average outcome variable for the individual's social network (*Wy*). This approach is simple to implement, but has the drawback of relying on the functional form specification to identify the social interaction effects.

Turning to the second estimation strategy, the Moffitt (or linear-in-means) model is only identified with certain definitions of the interaction matrix (i.e. if W is linearly independent or W is linearly dependent and no individuals are isolated[18]) or with the use of instrumental variables that meet exclusion restrictions (i.e., variables that determine

[18] This means that each household's network cannot be of the same size.

your neighbor's behavior but otherwise do not directly affect your own behavior, Moffitt, 2001, Bajari et al, 2006, and Bramoulle et al, 2009). Exclusion restrictions in the literature include: remittances (DeWeert and Dercon, 2006), land and livestock (Dercon and Krishnan, 2003), expenditures and religion (Quisumbing et al, 2004).

Because I consider binary outcomes, I can employ both the Brock and Durlauf and the instrumental variables estimation strategies to identify endogenous and exogenous effects. I estimate the Moffitt model as a 2 stage linear probability model, which has consistent results with the probit when further from the distribution tails (Blundell and Powell, 2003).

Overall, the model can be represented by equation (vi):

(vi) $$y = \delta Wy + \gamma Wx + \beta X + e$$

where y is the binary outcome variable (either microenterprise participation or whether decreased agriculture), X are own household characteristics, Wx is the weighted average (of some) of the network's household characteristics, Wy is the weighted average of the network's outcome. Following Manski's definition, δ refers to the endogenous social effect and γ to the exogenous (or contextual) social effect. W is the NxN row normalized interaction matrix. In the Brock and Durlauf specification the non-linear nature of the model allows for the identification of y. In the Moffitt model this is achieved by first estimating Wy as a function of Wxs and exclusion restrictions and then using the

predicted values in the estimation of the second stage. Therefore, I use the soil quality, whether receive government pension or transfer, number of children living outside of the TNF, distance to community center, and household size as instruments that affect the household's own decision but not its' networks decision. With these instruments (IVs), I estimate a 2SLS model:

(vii) 1st: $Wy^* = b_1 Wx + b_2 W^* IVs + \varepsilon$ and then

(viii) 2nd $y = \delta Wy^* + \gamma Wx + \beta X + e$

This model specification is similar to that of Moffitt (2001) and Bajari et al. (2006).

In the following tables I show results for three different sets of analyses. The first two estimate the probability that a household is still participating in the microenterprises in 2008 and the probability that a household decreases the land area that it has in cultivation between 2006 and 2008. Both models include own household characteristics representing *X* as discussed a, a subset of these *X* as exogenous effects (the network's land in agriculture in 2006, microenterprise participation in 2006, and days not working in 2008), and endogenous effects (network's microenterprise participation in 2008 or network's decision whether to decrease land area in crops between 2006 and 2008). The third model more fully captures the relationship between microenterprise participation and agricultural land area, estimating a bivariate probit model that allows these to be simultaneously determined.

There is a common set of household and community characteristics hypothesized to influence the household's own labor allocation, and hence both whether a household participates in the microenterprises and whether it decreases its area under shifting cultivation. These represent the categories discussed in section 2, with their variables described below:

Physical capital:

- soil quality dummy: expected to affect the preference for agriculture and/ or the need for less land and labor to produce the same amount of agricultural output;
- distance from community center: being further from the community center might hinder the participation in microenterprises but at the same time incentivize agriculture;
- number of trips to the city per month: going to the demand center more often can help in selling household production and accordingly focus on scale of one activity rather than diversification between different activities.

Labor:

- Average age of household heads: younger heads of household might not have children to take care of and might have more time available. Older heads of household might not be able to work on the fields anymore and therefore prefer microenterprises;
- Household size: number of people in household might influence household preferences regarding allocation of labor;

- Proportion of working adults in households (older than 10 but younger than 70 years of age): direct measure of labor supply available to the houehold;
- Proportion of small children in household (less than 5 years old): the care of small children absorbs labor;
- Days not working: included as a measure of labor quality.

Human capital: measures of experience in productive activities

- Land in agriculture in 2006;
- Proportion of life spent in community;
- Dummy for whether participates in ecotourism microenterprise in 2006;
- Dummy for whether participates in oil microenterprise in 2006;
- Dummy for whether participates in ecological leather microenterprise in 2006;

Exogenous income: measures of exogenous cash transfers that can affect cash constraints and access to financial capital.

- Dummy whether receives pension;
- Amount received in government transfers;
- Number of children living outside of community.

Community variables:

- Dummy for whether community has all weather road access;
- Distance (in hours) by boat to community from main urban center in the region (Santarém);

The exogenous effects considered were chosen as a subset of the household characteristics presented above that i) capture history of decisions about Ys (and the possibility of learning from these decisions) and ii) capture other contemporaneous mechanisms by which social networks might influence outcomes. These are:

- DNW: the household's response to: "How many days have the heads of household not been able to work in the last month due to sickness or taking care of another household member?". I argue this variable affects a household's own labor allocation decision and may affect their network's;
- Network's area in agriculture;
- Network's participation in microenterprises.

The linear in means model requires exclusion restrictions to identify the endogenous effect. We use the following identifying instruments:

- Soil quality dummy. If your network partner has better soils he would be more likely to do agriculture, but that should not influence your decision;
- Pension: whether your network receives retirement pension should not affect your labor allocation decision;
- Government transfers: whether your network receives government transfers should not affect your labor allocation decision;

- Children living outside the TNF: whether your network has more or less children living outside the community should not affect your labor allocation decision;
- Distance to community center: time walking to community center from household;
- Household size: the number of household members in the households within your network should not determine your own labor allocation.

Table 4.3 presents estimation results for the model of participation in the microenterprise program. This is a key measure of the long-term sustainability of the project, after the end of external funding. The results of the Brock and Durlauf and Moffit approaches are similar.[19] In this specification, the coefficient estimate for the network's participation in a microenterprise (Wyproj2s) is the endogenous effects. Exogenous (or contextual) social effects are captured by the coefficients on the network's lagged participation in microenterprises (wsemoc06), lagged land in agriculture (wslagr), and days not working (wsdnw). As such, we see that the endogenous effect is positive and significant in both estimating approaches, indicating that if your network has higher participation levels you are more likely to participate. Proportion of life spent in community is also positively related to microenterprise participation, meaning that a household which spent a higher proportion of their life in the community is more likely to participate than a household that spent less of their life in the community, *ceteris paribus*.

[19] Estimation results for the first stage of the 2SLS are not presented, but the joint test for significance of the exclusion restrictions is significant at the 1% level.

Finally, participants of the ecotourism microenterprise in 2006 were more likely to continue in the program. This can be seen as a measure of success of that microenterprise as compared to the other enterprises. One hypothesis is that this microenterprise does not require machinery to produce goods and does not depend on external markets (since tourists arrive at the communities).

Table 4.4 presents estimation results for the model of decreasing land in agriculture. This is a key measure of the environmental success of the microenterprise program, because the external funders and the National Forest have a fundamental interest in decreasing deforestation. It is also a key indicator of household livelihoods, as crops are traditionally the primary source of income in these and many communities living on the forest edge in the Brazilian Amazon. The microenterprises were implemented with the objective of shifting households into activities based on the forest management and hence more compatible with the mission of the protected area. Again, the joint test for the significance of the exclusion restrictions in the first stage of the Moffit model is significant at the 1% significance level. In the second stage (presented in the table), there are some differences between the Moffitt and the Brock and Durlauf approaches, most notably the significance of the endogenous effect estimate (significant in Brock and Durlauf but not in Moffit results).

Both approaches show a significant exogenous effect: the weighted average of days not working in each household's network. If a household's network is "more sick", there is a less chance that a given household will decrease its land in agriculture. This is plausible as having people in your network that are incapacitated (due to their own illness or occupied with taking care of another household member) might encourage you to maintain a large area in production, both in case you need to help your network and in recognition that your network has limited ability to help you.

In terms of own characteristics, a household's own area in agriculture in 2006 is a very significant determinant of whether to decrease land in agriculture between 2006 and 2008: the more land in agriculture a household had in 2006 the more likely it was to reduce production in 2008. Government transfers also serve as an income source and therefore, the higher the transfer the more likely households were to decrease land in agriculture. On the other hand, children living outside of the community can mean either remittances or greater expenses if they depend on help from parents in community (e.g., children attending school in the city). The latter appears to be the case in the Tapajós, as having more children living outside of the community decreases the probability of reducing land in agriculture. Finally, participants in the oil microenterprises in 2006 were more likely to decrease their area in agriculture, which suggests that this project may have been more successful in diverting labor away from this activity (although not significant in the Brock and Durlauf specification).

Finally, table 4.5 shows the results of a bivariate probit model in which households co-determine whether to participate in microenterprises and whether to decrease land in agriculture. Here the errors of the two equations are correlated, reflecting the joint determination of decreasing agriculture and participating in microenterprises. Despite interesting results, the correlation estimate rho is only significant at the 27% significance level. However, the AIC criterion for the bivariate model is lower than for the separate Brock and Durlauf models, indicating that this model is preferred.

We can see that the endogenous effects are significant in both models: the network's microenterprise participation is significant in the microenterprise model (if the participation within your network is higher you are more likely to participate) and the network's decision to decrease land in agriculture is significant in the agriculture model (if more people in your network decrease agriculture you are more likely to decrease your own land in agriculture). Thus, by recognizing the simultaneous nature of decisions, we obtain stronger estimates of the social network effects. Comparing these estimates to the literature (Soetevent, 2006) we see that the marginal effects for the endogenous effects in my estimation are considered “large” but still within expectations.

In the microenterprise participation model none of the exogenous effects are significant, while in the agriculture model a network’s days not working is still associated with a lower probability of decreasing your land in agriculture. Considering own

household characteristics, area in agriculture in 2006 is a very significant determinant of whether a household decreased its agricultural land. Government transfers also encourage households to decrease land in agriculture, while having more children living outside of the community makes decreasing land in agriculture less likely. Proportion of life spent in the community is associated with a higher probability of participating in microenterprises, and participants in the ecotourism microenterprise in 2006 have a higher chance of remaining in the program in 2008. Finally, if the community has road access, households are less likely to engage in microenterprises and more likely to decrease agriculture, possibly indicating that these communities have more access to other sources of income.

4.7 Conclusions and policy implications

As part of a pro-poor environmental agenda, community-based natural resource management has become central to international conservation efforts in recent decades. While impact during project is more commonly assessed, it is necessary to evaluate the permanence of such projects beyond the external funding phase, since this is an important problem for international aid in general (Kareiva et al. 2008). My contribution in this chapter is to identify characteristics of households that indicate their permanence in the microenterprises and decrease in slash-and-burn agriculture (the objective of the microenterprises). As these projects emphasize cooperation among households in the

implementation phase, I hypothesize that social networks are also critical to the process of sustaining (or abandoning) activities promoted by the microenterprises.

I examine this hypothesis in the context of a pilot project to engage resident communities in sustainable management of a national forest in the Brazilian Amazon. In this chapter I evaluate the impact of social networks on a household's decision of how to allocate labor between agriculture and microenterprises. I discuss the motivation for focusing on social networks, and then estimate the impact of social networks on the decision to participate in the microenterprises and the decision to decrease land in agriculture. I find that social network effects are significant and behave as expected. Specifically, endogenous social network effect estimates are large (compared to literature) both in terms of determining which households participate in the microenterprises as well as which ones decide to decrease their land in agriculture. I also find heterogeneous effects regarding participation in different microenterprises in 2006, indicating that they have affected household decisions differently.

Such results have important policy implications in that extension agents and project implementers should be aware of how participation depends on the actual reference group and that this can drive household decisions regarding competing economic activities. The comparison between microenterprise participation and agriculture shows that households indeed make this decision simultaneously, while

considering the decisions of their peers and the opportunities available. Overall we see that social networks do play a significant role in determining labor allocation, which indicates following what other households in your network are doing is indeed an important way of deciding what to do. Another factor affecting the decision to decrease agricultural land are government transfers while constraints include poor health of networks and having children living outside the TNF. In terms of contituity in microenterprise participation, it is important to note networks do affect this decision and that different projects had differing effects. The results indicate that, similar to conventional wisdom, it is important to get weak ties (households who are part of many networks, as defined by Granovetter,1973) to participate since they are the households who might instigate others to adopt new livelihood strategies.

4.8 References

Adams, W. M., R. Aveling, D. Brockington, B. Dickson, J. Elliott, J. Mutton, D. Roe, B. Vira, and W. Wolmer. 2004. Biodiversity conservation and the eradication of poverty. Science 306:1146-1149.

Anselin, L. 2002. Under the Hood. Issues in the Specification and Interpretation of Spatial Regression Models. Unpublished manuscript.

Bacha, C.J.C., Rodriguez, L.C.E. 2005. The people of Tapajós. ITTO Tropical Forest Update v.15(4)

Bajari, P., Hong, H., Krainer, J., Nekipelov, D. 2006. Estimating Static Models of Strategic Interaction. NBER Working Paper 12013.

Bandiera, O., Rasul, I. 2002. Social Networks and Technology Adoption in Northern Mozambique. LSE STICERD Research Paper #DEDPS35

Bayer, P.J., Pintoff, R., Pozen, D. 2004. Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections. Yale University Economic Growth Center Discussion paper #864

Bayer, P., Timmins, C. 2005. On the equilibrium properties of locational sorting models. Journal of Urban Economics 57 (2005) 462–477

Berkes, F. 2007. Community-based conservation in a globalized world. PNAS 104(39): 15188-15193

Bramoulle, Y., Djebbari, H., Fortin, B. 2009. Identification of Peer Effects Through Social Networks. Journal of Econometrics 150 pp. 41-55.

Brock, W.A., Durlauf, S.N. 2007. Identification of binary choice models with social interactions. Journal of Econometrics 140 pp.52-75.

Brock, W.A., Durlauf, S.N. 2001a. Discrete Choice with Social Interactions. The Review of Economic Studies, v.68(2): 235-260

Brock, W.A., Durlauf, S.N. 2001b. Interaction-Based Models. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics Vol.5. North Holland, Amsterdam. Pp. 3297-3380.

Case, A. (1992) Neighborhood influence and technological change. Regional Science and Urban Economics 22: 491-508.

Conley, T.G., Udry, C.R. (forthcoming) Learning about a new technology: Pineapple in Ghana. American Economic Review.

Dasgupta, P. 2003. Social Capital and Economic Performance: Analytics. In: E. Ostrom and T.K. Ahn eds., Critical Writings in Economic Institutions: Foundations of Social Capital. Cheltenham, UK: Edward Elgar.

Dercon, S., Krishnan, P. 2003. Food Aid and Informal Insurance. Discussion Paper 2003/09 World Institute for Development Economics Research.

Dickinson, K. 2008. Indian Toilets and Tanzanian Mosquito Nets: Understanding Households' Environmental Health Decisions in Developing Countries. PhD Thesis available online at http://hdl.handle.net/10161/593

Dickinson, K.L., Pattanayak, K. 2009. Open sky latrines: Social reinforcing in the case of a (very) impure public good. Duke University unpublished manuscript.

Fafchamps, M., Gubert, F. 2007. The formation of risk sharing networks. Journal of Development Economics 83:326-350.

Fafchamps, M., Hill, R.V. 2005. Selling at the Farmgate or Traveling to Market. American Journal of Agricultural Economics 87(3) pp. 717-734

Fafchamps, M. (2003) Ethnicity and Networks in African Trade. Contributions to Economic Analysis and Policy, 2(1)

Fafchamps, M. and Minten, B. 2002. Returns to social capital among traders. Oxford Economic Papers 54 pp.173-206

Ferraro, P.J., Pattanayak, S.K. 2006. Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments *PLoS Biology* 4(4) Available: http://biology.plosjournals.org/perlserv?request=get-document&doi=10.1371/journal.pbio.0040105

Foster, A.D., Rosenzweig, M.R. 1995. Learning by doing and learning from others: human capital and technical change in agriculture. The Journal of Political Economy 103(6) pp. 1176-1209

Friedman, T.L. 2009. Truck, Trains and Trees. The New York Times. Available online at: http://www.nytimes.com/2009/11/11/opinion/11friedman.html?_r=1

Fujita, M., Krugman, P. , Venables, A. 1999. The Spatial Economy, MIT Press, Cambridge.

Garnett, S.T., Sayer, J., du Toit, J. 2007. Improving the effectiveness of interventions to balance conservation and development: a conceptual framework. Ecology and Society 12(1): 2. [online] URL: http://www.ecologyandsociety.org/vol12/iss1/art2/

Glaser, B., Haumaier, L., Guggenberger, G., Zech, W. 2001. The 'Terra Preta' phenomenon: a model for sustainable agriculture in the humid tropics. Naturwissenschaften 88(1)

Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78: 1360-80.

Hogset, H., Barrett, C.B. forthcoming. Social Learning, Social Influence and Projection Bias: A caution on inferences based on proxy–reporting of peer behavior. *Economic Development and Cultural Change*.

Ioannides, Y.M., Zabel, J.E. 2003. Neighbourhood Effects and Housing Demand. Journal of Applied Econometrics 18: 563-584.

Jensen, R. 2007. The Digital Provide: Information (technology), market performance and welfare in the South Indian fisheries sector. Quarterly Journal of Economics 122(3) pp. 879-924

Kareiva, P., Chang, A., Marvier, M. 2008. Development and Conservation Goals in World Bank Projects. Science v.321

Kremer, M., Miguel, E. 2007. The illusion of sustainability. Quarterly Journal of Economics 122(3) pp. 1007-1065

Loader, R. 1997. Assessing Transaction Costs to Describe Supply Chain Relationships in Agri-Food Systems. Supply Chain Management 2(1), pp.23-35.

Luke, N., Munshi, K. 2007. Social affiliation and the demand for health services: Caste and child health in South India. Journal of Development Economics 83: 256–279

Manski, C. 1993. Identification of Endogenous Social Effects: The Reflection Problem. Review of Economic Studies, 60: 531-542.

McShane, T. O., Wells, M. P. 2004. Integrated conservation and development? Pages 3-9 in T. O. McShane and M. P. Wells, editors. Getting biodiversity projects to work: towards better conservation and development. Columbia University Press, New York, New York, USA.

Moffitt, R.A. 2001. Policy Interventions, Low Level Equilibria and Social Interactions. In: Durlauf, S.N., Young, H.P (eds.). Social Dynamics. Cambridge, MA: MIT Press.

Munshi, K. 2004. Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. Journal of development Economics v.73(1):185-213

Neumann, R.P., Hirsch, E. 2000. Commercialization of non-timber forest products: review and analysis of research. Cifor, Bogor 175p.

Pattanayak, S.K., Sills, E.O. 2001. Do tropical forests provide natural insurance? The microeconomics of non-timber forest product collection in the Brazilian Amazon. Land Economics 77(4) pp. 595-612.

Perez, R.M., Arnold, J.E.M.(eds) 1996. Current issues on non-timber forest product research. Proceedings of the workshop “Research on NTFP”. Hotsprings, Zimbabwe 28 Aug-2 Sep 1995. Cifor and ODA, Bogor, Indonesia.

Pretes, M. 2002. Microequity and Microfinance. World Development 30(8): 1341-1353

Putz, F. E. 2000. Economics of home grown forestry. Ecological Economics, 32, 9-14

Quisumbing, A.R., McNiven, S., Godquin. M. 2008. Shocks, Groups, and Networks in Bukidnon, Philippines. CAPRi Working Paper 84.

Redford, K. H. 1992. The empty forest. BioScience, 42, 412-422.

Santos, P., Barrett, C.B. forthcoming. Identity, Interest and Information Search in a Dynamic Rural Economy. *Economic Development and Cultural Change*.

Soetevent A.R. 2006. Empirics of the Idenfitication of Social Interactions: and evaluation of the approaches and their results. Journal of Economic Surveys, 20(2): 193-228

Shanley, P., Luz, L., Swingland, I. R. 2002. The faint promise of a distant market: a survey of Belém's trade in non-timber forest products. Biodiversity and Conservation, 11, 615-636.

Schelling, T.C. 1969. Models of segregation, American Economic Review 59 (2) 488–493.

Schelling, T.C. 1971. Dynamic models of segregation, Journal of Mathematical Sociology 1 143–186.

Tiebout, C.M. 1956. A pure theory of local expenditures, Journal of Political Economy 64 416–424.

Van den Broeck, K., Dercon, S. 2007. Social Interactions in Growing Bananas: Evidence from a Tanzanian Village. The Centre for the Study of African Economies Working Paper #264.

Weinberg, B.A. 2006. Social Interactions and Endogenous Association. Unpublished manuscript.

Table 4.1 Descriptive statistics for household variables.

Definition	Variable	Mean	Std Dev	Min	Max
Dummy whether participates in microenterprises in 2008	projectpd	0.23	0.42	0	1
Dummy whether participates in microenterprises in 2006	emoc06	0.36	0.48	0	1
Dummy whether participates in ecotourism project in 2006	ptur	0.17	0.38	0	1
Dummy whether participates in oil project in 2006	poil	0.04	0.19	0	1
Dummy whether participates in eco leather project in 2006	pcouro	0.03	0.17	0	1
Dummy whether participates in furniture project in 2006	pmov	0.19	0.39	0	1
Area of agriculture in 2008	rocaq	3.30	2.14	0	16
Area of agriculture in 2006	lagr	3.85	3.03	0	22
Dummy whether decreased area in agriculture	agdec	0.51	0.50	0	1
Soil quality dummy (terra preta de índio)	lind	0.35	0.48	0	1
Number of heads of cattle owned by household	qcattle	1.73	5.14	0	46
Dummy if receives retirement pension	retire	0.30	0.46	0	1
Amount received in retirement pension (in 100s of R$)	retirep00	1.85	3.08	0	12.8
Dummy if receives government transfer	bolsaf	0.67	0.47	0	1
Amount received in government transfer (in 100s of R$)	bolsafp00	0.65	0.52	0	1.52
Time from community center	tbarrac	14.11	21.28	1	150
number of children living outside household	childout	1.53	1.89	0	9
Proportion of life spent in community	lifecom	0.79	0.32	0	1
Average age of household heads	agehhhm	46.62	14.15	21	81
Number of people in household	nhh	5.84	2.49	1	14
Proportion of household under 5 years of age	childm	0.11	0.14	0	0.57
Proportion of household over 10 years of age	adultm	0.73	0.23	0.25	1
Number of trips to city per month	citytrip	2.24	1.68	0	11
Number of days not working per month	DNW	10.83	15.72	0	60
Dummy whether community has all weather road access	road	0.83	0.38	0	1
Distance (in hours) by boat to community from Santarém	dboat	5.81	1.85	3.5	12

Table 4.2 Results of OLS regressions of household characteristics that explain the number of links for each type of interaction matrix (*** significant at 1% significance level; ** significant at 5% significance level; * significant at 10% significance level).

	Matrix	Neighbors		Association		Project		Choice	
	R^2 / F stat	5%	1.15	7%	1.63	10%	2.23	5%	1.11
Definition	Variable	Estimate	Std Error	Estimate	Std Error	Estimate	Std Error	Estimate	Std Error
	Intercept	2.05 ***	0.35	0.43 **	0.18	3.21 ***	0.92	6.28 ***	0.40
Wealth (principal component)	pcafactor	1.23E-02	5.39E-02	6.28E-02 **	2.77E-02	-4.16E-02	0.14	2.89E-02	6.26E-02
If head of household is migrant	migrant	-0.26	0.20	-6.58E-03	0.10	-0.30	0.53	0.20	0.23
If participates in project	projectpd	-5.28E-02	0.14	-4.31E-02	7.24E-02	0.42	0.38	7.96E-02	0.16
Land in agriculture in 2009	rocaq	1.25E-02	2.98E-02	-2.08E-02	1.53E-02	0.22 ***	7.91E-02	-2.68E-02	3.46E-02
Number of siblings in community	broc	2.97E-02 *	1.81E-02	-7.06E-04	9.26E-03	-0.15 ***	4.81E-02	-7.27E-03	2.10E-02
Household size	nhh	1.10E-02	2.61E-02	6.82E-03	1.34E-02	5.74E-03	6.93E-02	3.04E-03	3.03E-02
Age of heads of household	agehhhm	2.67E-03	5.31E-03	4.33E-03	2.73E-03	8.37E-04	1.41E-02	-1.50E-02 **	6.18E-03
Number of trips to the city per month	citytrip	-1.06E-02	3.70E-02	1.65E-02	1.85E-02	9.09E-02	0.10	6.32E-03	4.31E-02
Number of children living outside community	childout	-1.71E-02	3.85E-02	-2.61E-04	1.97E-02	-0.11	0.10	2.42E-02	4.48E-02
Time walking to community center	tbarrac	-5.52E-03 *	2.98E-03	1.92E-03	1.53E-03	-1.06E-02	7.91E-03	-2.06E-03	3.46E-03
Number of remittances received per month	gethn	8.00E-02	0.11	1.52E-02	8.02E-02	0.29	0.28	0.17	0.12

Table 4.3 Household models of decision to participate in microenterprises (projectpd as LHS variable) (*** significant at 1% significance level; ** significant at 5% significance level; * significant at 10% significance level).

Definitions	Variable	2SLS F value=1.42* R^2=13% Estimate	Std Dev	Brock and Durlauf Likelihood ratio ** Estimate	Std Dev
	Intercept	0.27	0.25	-0.48	0.88
Network's participation in project	Wyproj2S	0.54 *	0.30	1.08 **	0.45
Network's participation in project in 06	wsemoc06	3.32E-02	0.10	-6.67E-02	0.36
Network's land in agriculture in 06	wslagr	-1.57E-02	1.44E-02	-5.23E-02	6.00E-02
Network's days not working	wsDNW	1.17E-03	3.53E-03	1.12E-02	1.25E-02
Distance (in minutes walking) to community center	tbarrac	-1.34E-03	1.40E-03	-1.17E-02	8.15E-03
Area of agriculture in 06	lagr	-4.00E-03	1.02E-02	-1.55E-02	3.92E-02
Soil quality dummy (terra preta de índio)	lind	-4.22E-02	6.27E-02	-0.18	0.22
Dummy if receives retirement pension	retire	1.10E-02	9.38E-02	6.00E-02	0.33
Amount received in government transfer (in 100s of R$)	bolsafp00	-9.07E-02	7.05E-02	-0.24	0.25
Number of children living outside household	childout	1.21E-02	1.71E-02	3.71E-02	5.86E-02
Proportion of life spent in community	lifecom	0.25 ***	9.45E-02	1.01 ***	0.39
Average age of household heads	agehhhm	2.99E-04	3.82E-03	5.38E-03	1.35E-02
Number of people in household	nhh	4.44E-04	1.41E-02	-4.51E-03	4.98E-02
Proportion of working adults in household	workam	-0.19	0.18	-0.69	0.64
Proportion of non working members of household	nworkm	-0.49	0.38	-1.96	1.55
Number of heads of cattle owned by household	qcattle	5.32E-03	6.15E-03	1.23E-02	2.30E-02
Number of trips to city per month	citytrip	2.23E-02	1.76E-02	7.07E-02	5.91E-02
Number of days not working per month	DNW	-1.97E-03	1.92E-03	-8.64E-03	7.47E-03
Dummy whether participates in ecotourism project in 06	ptur	0.13 *	7.79E-02	0.52 **	0.25
Dummy whether participates in oil project in 06	poil	-8.54E-02	0.15	-0.28	0.51
Dummy whether participates in ecological leather project in 06	pcouro	-0.29	0.19	-4.84	234.20
Dummy whether community has land access	road	-0.11	8.76E-02	-0.49	0.31
Distance (in hours) by boat to community from Santarém	dboat	-1.06E-02	1.85E-02	-6.42E-02	6.31E-02

Table 4.4 Household models of decision to decrease land in agriculture (agdec as LHS variable) (*** significant at 1% significance level; ** significant at 5% significance level; * significant at 10% significance level).

Definitions	Variable	2SLS F value=2.27***		R^2=21%	Brock and Durlauf Likelihood ratio ***		Std Dev
		Estimate		Std Dev	Estimate		
	Intercept	9.44E-02		0.27	-0.69		1.01
Network's decision to decrease land in agricutlure	wsagdec	0.17		0.34	0.66	*	0.40
Network's land in agriculture in 06	wslagr	-1.87E-02		3.04E-02	-6.11E-02		5.65E-02
Network's days not working	wsDNW	-7.61E-03	*	4.15E-03	-2.32E-02	*	1.24E-02
Network's participation in project in 06	wsemoc06	0.11		0.11	0.24		0.34
Distance (in minutes walking) to community center	tbarrac	-3.70E-04		1.58E-03	-3.47E-03		4.99E-03
Area of agriculture in 06	lagr	5.70E-02	***	1.14E-02	0.30	***	5.75E-02
Soil quality dummy (terra preta de índio)	lind	9.24E-03		7.05E-02	-2.87E-02		0.21
Dummy if receives retirement pension	retire	6.75E-02		0.10	0.39		0.31
Amount received in government transfer (in 100s of R$)	bolsafp00	0.14	*	7.81E-02	0.45	*	0.24
Number of children living outside household	childout	-3.17E-02	*	1.91E-02	-0.13	**	5.81E-02
Proportion of life spent in community	lifecom	-5.64E-02		0.11	-0.24		0.31
Average age of household heads	agehhhm	1.07E-03		4.26E-03	3.66E-03		1.26E-02
Number of people in household	nhh	-5.89E-03		1.57E-02	-3.84E-02		4.67E-02
Proportion of working adults in household	childm	0.10		0.20	-0.94		1.03
Proportion of non working members of household	adultm	0.35		0.43	-0.28		0.71
Number of heads of cattle owned by household	qcattle	-1.52E-03		6.86E-03	1.88E-02		2.29E-02
Number of trips to city per month	citytrip	1.70E-03		1.97E-02	-1.33E-02		5.86E-02
Number of days not working per month	DNW	-1.76E-03		2.14E-03	-1.24E-03		6.56E-03
Dummy whether participates in ecotourism project in 06	ptur	-1.36E-02		8.59E-02	-7.68E-02		0.25
Dummy whether participates in oil project in 06	poil	0.33	*	0.17	0.89		0.62
Dummy whether participates in ecological leather project in 06	pcouro	9.49E-02		0.21	0.23		0.61
Dummy whether community has land access	road	0.17	*	0.10	0.47		0.30
Distance (in hours) by boat to community from Santarém	dboat	-2.88E-03		2.05E-02	-4.95E-02		6.34E-02

Table 4.5 Bivariate probit for simultaneous decision of whether to participate in microenterprises and whether to decrease land in agriculture (agdec and projectpd). (*** significant at 1% significance level; ** significant at 5% significance level; * significant at 10% significance level). Rho is not significant, but bivariate probit is better than separate models by AIC criterion.

Definition	Variable	Project participation model			Agriculture decrease model		
		Estimate		Std Error	Estimate		Std Error
	Intercept	-0.44		0.88	-1.26		0.81
Network's participation in project	wsemoc	1.11	**	0.46	-8.83E-02		0.41
Network's participation in project in 06	wsemoc06	-0.12		0.37	0.33		0.35
Network's decrease agricultural land	wsagdec	0.31		0.42	0.67	*	0.39
Network's land in agriculture in 06	wslagr	-8.43E-02		7.70E-02	-6.41E-02		5.57E-02
Network's days not working	wsDNW	1.10E-02		1.29E-02	-2.41E-02	**	1.16E-02
Distance (in minutes walking) to community center	tbarrac	-1.22E-02		8.33E-03	-3.65E-03		5.04E-03
Area of agriculture in 2006	lagr	-1.07E-02		3.96E-02	0.31	***	5.50E-02
Soil quality dummy (terra preta de índio)	lind	-0.20		0.22	-5.19E-02		0.20
Dummy if receives retirement pension	retire	0.13		0.34	0.34		0.31
Amount received in government transfer (in 100s of R$)	bolsafp00	-0.25		0.24	0.42	*	0.24
Number of children living outside household	childout	4.72E-02		5.85E-02	-0.12	**	5.85E-02
Proportion of life spent in community	lifecom	0.94	**	0.39	-0.18		0.32
Average age of household heads	agehhhm	4.61E-03		1.41E-02	2.33E-03		1.29E-02
Number of people in household	nhh	-6.32E-03		5.14E-02	-2.55E-02		4.70E-02
Proportion of working adults in household	workam	-0.83		0.65	0.14		0.60
Proportion of small children (<5yo) in household	nworkm	-2.20		1.58	1.13		1.47
Number of heads of cattle owned by household	qcattle	1.13E-02		2.38E-02	1.34E-02		2.33E-02
Number of trips to city per month	citytrip	7.88E-02		5.63E-02	-1.34E-02		5.96E-02
Number of days not working per month	DNW	-6.75E-03		7.56E-03	-1.57E-03		6.55E-03
Dummy whether participates in ecotourism project	ptur	0.48	**	0.25			
Dummy whether participates in oil project	poil	-0.17		0.52			
Dummy whether community has land access	road	-0.52	*	0.31	0.55	*	0.30
Distance (in hours) by boat to community from Santarém	dboat	-4.94E-02		6.21E-02	-4.01E-02		6.07E-02

APPENDICES

Appendix 4.A: Literature review on social network studies in development applications.

Table 4.A.1 Literature review of social networks in development applications in developing countries.

Study	Location of study	Empirics / theory	Problem	Theoretical modeling strategy	Definition of networks	Exog. ntwrk?	Empirical modeling strategy	IV
Case, 1992	Indonesia	Theory and empirics	sickle adoption as rice harvesting tool in Indonesia	Household's expected profit on new technology depends on neighbor's expected profits	district	yes	Probits of sickle adoption with and without district variables and compares log-likelihood of both.	none
Foster and Rosenzweig, 1995	India	Theory and empirics	Adoption of high yield varieties in India	Profit maximization where input decisions each year know the technology of production up to a random "target" for input use that has both a systematic and an idiosyncratic component.	village	yes	OLS, fixed effects, instrumental variables fixed effects	lagged village average
Besley and Case, 1997	India	Theory and empirics	Adoption of high yield varieties of cotton in India	Farmer maximizes expected payoff for sowing (dynamic problem). Individuals are forward looking and bayesian.	village	yes	simulations based on data of 28 farmers	none
Narayan and Pritchett,1999	Tanzania	Empirical	Social network capital effectively serves as a substitute for real capital in mediating economic mobility	none	Membership in associations	yes	OLS and OLS with IV	for income: trust in strangers, tribesman, cell leader, village chairman (government), district officials, central government.
Behrman et al, 2002	Kenya	Empirical	Family planning in Kenya	none	who you talk to	no	model that contraceptive use depends on own and network charac.	none

Table 4A1 Continued.

Study	Location of study	Empirics/ theory	Problem	Theoretical modeling strategy	Definition of networks	Exog. ntwrk?	Empirical modeling strategy	IV
Carter and Maluccio, 2002	South Africa	Empirical	Effects of private shocks on child development. Social capital as an individual asset conferring private	none	number of groups and informal assoc. in communities; geographic	yes	OLS with fixed effects	none
Fafchamps, 2003	Benin, Malawi, Madagascar	Theory and empirics	effect of ethnicity, gender and religion on trust	Game theoretical model: seller screns buyer and buyer decides to pay or not.	ethnicity; gender; religion	yes	Regressions of transaction technology on gender, religion, ethnicity and covariates	none
Romani, 2003	Côte d'Ivoire	Theory and empirics	Ethnicity and information flows	Technology adoption with production function that is increasing in knowledge	ethnicity	yes	CobbnoneDouglas production function	none
Fafchamps and Lund, 2003	Philippines	Theory and empirics	Using detailed data on gifts, loans, and asset sales, this paper investigates how rural Filipino households deal with income and expenditure shocks.	if risk is shared efficiently, individual consumption should be unaffected by idiosyncratic variations in income	Individuals on which they could rely in case of need or to whom the respondent gives help when called upon to do so.	yes	Whether gifts and informal loans serve to spread risk: test whether gifts and net informal borrowing rise when households face a severe shock, but fall when their partners face a similar shock.	For endogeneity of network: model with household fixed effects, which capture unobs. past shocks that continue to affect lending and gift giving in subsequent periods
Dercon and Krishnan, 2003	Ethiopia	Empirics	Study food aid distribution in rural Ethiopia. What determines the allocation rule of food aid in Ethiopia? What is the impact on consumption of food aid transfers?	none	whether receives aid	yes	Probit and OLS with village one time fixed effects	Real consumption is treated as endogenous using a two stage regression. Land per adult and livestock values per adult are identifying instruments.

Table 4A1 Continued.

Study	Location of study	Empirics/ theory	Problem	Theoretical modeling strategy	Definition of ntwrks	Exog. ntwrk?	Empirical modeling strategy	IV
Quisumbing et al, 2004	Philippines	Empirical	What shocks do households face and whether hh membership in networks helps mitigate shocks	none	Formal: number of group memberships; Informal: number of people in trust-based networks	no	OLS of shocks as a fcn of hh characteristics; Tobit for size of network (with IV); OLS with IV for income with networks size as variable in regression	Variables that affect whether households join groups, but which do not affect the size of the network: whether the household is a sugar producer, whether the head was Catholic, barangay heterogeneity indices, per capita expenditures on groups, and the barangay mean number of groups, excluding the household.
Conley and Udry, 2005	Ghana	Theory and empirics	Adoption of fertilizer for pineapple plantation	Choose fertilizer level to maximize expected profit at t+1, beliefs on fertilizer change over time.	Who you talk to; same intensity (different networks)	no	Fixed effects; logit for modeling probability on being on same network	lagged variables of network outcomes
Bandiera and Rasul, 2006	Northern Mozambique	Theory and empirics	Decision of whether to adopt a new crop: sunflower	Bayesian updating; learning by doing and learning from others.	family and friends; religion and village		Baseline regressions; fixed effects; nonnonelinear (quadratic), non parametric kernel regressions; uses village fixed effects.	number of adopters within networks
Kremer and Miguel, 2006	Kenya	Theory and empirics	Decision of whether to adopt deworking in randomized experiment	Adopt if private benefits > expected costs; Bayesian updating;	schools	yes, experimental data	Probit of whether treated	number of social links to treated schools
Munshi and Myaux, 2006	Bangladesh	Theory and empirics	Diffusion of contraceptive practices	Individual maximizes payoff, dependent on own and partner's actions, which determines the social sanction that she will face in that period.	religion	yes	OLS	none

Table 4A1 Continued.

Study	Location of study	Empirics/ theory	Problem	Theoretical modeling strategy	Definition of networks	Exog. ntwrk?	Empirical modeling strategy	IV
DeWeert and Dercon, 2006	Tanzania	Theory and empirics	Insurance against illness, weak ties. Who insures whom (link formation).	How households react with different coping strategies according to network structure they are in.	who you depend on for help or who depends on you for help	no	2SLS with fixed effects	Changes in demographic characteristics of network members during the survey period for a given network and remittance flows to network members from outside the network.
Mwakubo et al, 2006	Kenya	Empirics	Analyzes the influence of social capital on the farmers' perception of the soil erosion problem and the level of investments in soil conservation in marginal areas of Kenya	none	number of groups a housold is part of	yes	Heckman 2 stage; perception of soil erosion problem selection Probit and primary terracing intensity least square regression models	none
Moser and Barrett, 2006	Madagascar	Theory and empirics	Dynamics of smallholder technology adoption, with particular reference to a high-yielding, low-external input rice production method in Madagascar.	Decision to adopt/ expand/ disadopt in utility max framework; allow for learning as well as social norms	community	yes	probit model and symmetrically censored least squares estimation of a dynamic Tobit model to analyze the decisions to adopt, expand and disadopt the method	none
Adato et al , 2006	South Africa	Empirical	Suggests that the accumulation of social capital proves ineffective for households at the bottom of the economic pyramid in highly polarized economies	none	Class	yes	Returns to these core assets were estimated using a polynomial expansion of the basic assets	none

Table 4A1 Continued.

Study	Location of study	Empirics/ theory	Problem	Theoretical modeling strategy	Definition of networks	Exog. ntwrk?	Empirical modeling strategy	IV
Feder and Savastano, 2006	Indonesia	Empirics	Extent to which the attributes that make opinion leaders stand out in a community (social status, wealth, skills) can change their ability to disseminate information to an audience that is of a lesser status, lesser wealth and lesser skills.	none	village	yes	Probit of being selected to participate in training as indicators of farmers' opinion leadership capacity; Improvement in knowledge is modeled using twononesided tobit and ordered probit.	none
Luke and Munshi, 2007	South India	Empirical	Caste and child health investments	none	caste	yes	regressions for type of treatment	none
Jensen, 2007	South India	Theory and empirics	Assymetric information is eased by the use of cell phones in fisheries	Fishermen adopt cell phones if benefit>cost and take their catch to market with highest price.	who uses cell phones	yes	Non experimental program evaluation,	none
Fafchamps and Gubert, 2007	Philippines	Theory and empirics	Formation of insurance networks. Lit review of multiple pathways in which social networks might facilitate productivity growth, technology adoption and access to (informal) finance	cost vs benefits of establishing links	formation of networks: identify individuals on which it could rely in case of need or to whom they helps when called upon to do so.	no	logit dyadic regressions with corrected standard errors. If received gifts = 1, if received loans=1.	none
Fafchamps and Gubert, 2007	Philippines	Empirical	examine whether risk-sharing networks are formed so as to maximize the mutual gains from pooling income risk	none	formation of networks	no	OLS dyadic regressions with corrected standard errors. Y = log(value of gifts+1).	for total wealth: birthplace, number of siblings, and inherited wealth

Table 4A1 Continued.

Study	Location of study	Empirics/ theory	Problem	Theoretical modeling strategy	Definition of networks	Exog. ntwrk?	Empirical modeling strategy	IV
Van den Broeck and Dercon, 2007	Tanzania	Theory and empirics	Modeling the adoption of new banana varieties.	Bayesian updating	Kinship related groups, neighbourhoods and self-reported informal insurance groups	no	OLS and CobbnoneDouglas prod fcn (both with social effects)	none
Nepal et al, 2007	Nepal	Theory and empirics	Link between the strength and type of social networks and private forest conservation activity in rural Nepal. Test the effects of alternative social network indices on the number of trees planted on private land.	Net benefit maximization strength and type of social network affects the input requirement to produce the given outputs.	Forest user groups, farmers groups, water management associations, women in development groups, credit groups- then construct indices for participation	no	ZerononeInflated Negative Binomial (ZINB), ZerononeInflated Poisson (ZIP), and Negative Binomial (NB) models	household's caste/ethnicity, primary language at home (mother tongue), religion, and whether the household had experienced any major natural disasters (draught, epidemic, flood, and landslides) within the past five years
Dickinson, 2008	Tanzania	Theory and empirics	Malaria mosquito net adoption in Tanzania	household maximizes utility, others' decisions enter U fcn; sanitation has private cost.	who you talk to; same intensity	no	Probit of link formation and predicted value plugged into outcome probit.	none
Santos and Barrett, 2008	Ghana	Empirical	Role of identity in fertilizer adoption	links if cost<benefit; productive network embedded within social network	Who you talk to; same intensity	no	Probit of whether linked then test whether identity is significant	none

Table 4A1 Continued.

Study	Location of study	Empirics/ theory	Problem	Theoretical modeling strategy	Definition of networks	Exog. ntwrk?	Empirical modeling strategy	IV
Hogset and Barrett, 2008	Kenya	Empirical	Adoption of technologies in Kenya: comparing self with peer reporting	none	Who you talk to	no	Verifies if proxy peer reporting of adoption differs from own reporting. Finds that peer reporting is always significant while own reporting within network is not.	none
Cayers and Dercon, 2008	Ethiopia	Empirical	Whether receiving food aid depends on horizontal and vertical social networks.	none	size of the iddir the household was member of, whether the household has close associates holding official positions, and number of persons the household can rely on in times of need.	no	Probit with fixed effects. To evaluate endogeneity of networks uses older data to evaluate impact on present results and fails to identify reverse causality.	parent of the household head held an official position in the Kebele or elsewhere and a dummy for whether any parent was important in village social life as Ivs for political connections variable.
Comola, 2008	Tanzania	Theory and empirics	Link formation. How rural households form the links through which they provide and/or get economic support, and whether the connection structure of the community affects the formation of these links.	link if benefits>costs, heterogeneous costs and benefits	List of people who you can personally rely on for help and/or that can rely on you for help in cash, kind or labor?	no	dyadic regression (probit of links as a function of differences in agent's characteristics, with corrected standard errors)	none
Saha et al, 2008	Brazil	Theory and empirics	how social networks affect land use choices of farmers	Maximize expected profits based on information gathered from network	associations	no	spatial autocorrelation model	lagged variables

Table 4A1 Continued.

Study	Location of study	Empirics/ theory	Problem	Theoretical modeling strategy	Definition of networks	Exog. ntwrk?	Empirical modeling strategy	IV
Krishnan and Sciubba, 2008	Ethiopia	Theory and empirics	Demonstrate the role of both number of links and architecture in determining the impact of social networks on outcomes.	Stylised model for the formation of labournonesharing arrangements. Househol maximizes harvest subject to labor.	households were asked whether they had hosted a labour-sharing party in the main (meher) season, the number of people invited, and the identities of the people called.	no	Probit with village fixed effects. Tobit model for robustness checks.	Number of relatives living in the village, whether the household head was born in the village, whether she holds official office, average number of years of residence of the head of household in the village and whether the household head had ever travelle
Ligon and Schechter, 2009	Paraguay	Theory and empirics	Risk sharing among network members	Utility maximization with gifts over multiple time periods then derive expectations based on risk sharing behavior	Who you sent money to	no	Games	none
Dickinson and Pattanayak, forthcoming	South India	Theory and empirics	Latrine adoption in South India	household has a utility function that depends on private consumption, as well as the decision to adopt a latrine (which has social and private benefits)	village	yes	difference in difference; similar to Brock and Durlauf and Bajari et al.	none

Appendix 4.B:

Theoretical model

The model presented bellow is based on Munshi (2004) and adapted to the current context. I will start with a simple framework in which information on non-agriculture income is perfect, showing that this implies there is no role for social learning. I then change the model to account for imperfect information regarding non-agricultural income.

1. Perfect information on non-agricultural income sources

In this model I assume that each household has to allocate time to either agriculture (L_{ag}) or non-agricultural (L) activities to generate agricultural (y_{ag}) and non-agricultural (y) income. Agricultural income is certain, and I assume constant returns to scale (which is consistent with field observations). For non-agriculture:

i) $$y_{it} = y(X_i) + \eta_{it}$$

where $y(X_i)$ is the expected "wage" to be obtained from non-agricultural activities, X_i are characteristics of the household affecting their employability (such as transportation owned, distance from city, microenterprises in community) and η_{it} is a zero mean serially independent disturbance with variance λ_i^2. Also, $E(\eta_{it} \mid y(X_i)) = 0$. With perfect information $y(X_i)$ is known with certainty. The definition of y as wage or return to labor

will be used interchangeably throughout the text and represents the amount of money a household makes for an hour of work allocated to each type of activity.

Thus, the household chooses the optimal time to allocate to non-agriculture (L_{it}*) by maximizing its expected utility from the current profit V_{it} in each period:

ii) $$V_{it} = y_{it} * L_{it} + y_{ag}(T - L_{it})$$

where T is the household total time endowment. Therefore the household chooses L_{it} as to maximizes $V(Y,\Sigma)$ where $Y = (y(X_i) - y_{ag})L_{it} + y_{ag} * L; \Sigma = \lambda_i L_{it}$. Considering that households are risk averse, $V(.)$ is increasing in Y and decreasing in Σ.

Assuming an interior solution, an increase in $(y(X_i)-y_{ag})$ makes non-agricultural income sources more attractive (and therefore increases L_{it}), while an increase in λ_i decreases L_{it}. The optimal time allocation is an increasing function of $(y(X_i)-y_{ag})$ and a decreasing function of λ_i, both of which are constant over time, so the time subscript on L_{it}* can be dropped:

iii) $$L_i^* = L(y(X_i) - y_{ag}, \lambda_i)$$

Therefore, the household arrives at this optimal time allocation immediately and there is no role for social learning.

2. Imperfect information on non-agricultural income sources

In the case of imperfect information expected return to labor (or wage; $y(X_i)$) is uncertain, therefore $\Sigma = (\lambda_i + \sigma_{it})L_{it}$ where σ_{it}^2 is the variance of the household expected return to labor estimate, which is exogenously determined. Following this adjustment, equation iii) can be written as:

iv) $$L_{it} = L(y_{it} - \hat{y}_{ag}, \lambda_i, \sigma_{it})$$

where $\hat{y}$ replaces the certain y_{it} in part 1. Over time, $\hat{y}$ slowly converges to $y(X_i)$ as household gets more information. Also L_{it} converges to L_i* over time. The next step is to characterize the convergence of $\hat{y}_{it}$ to $y(X_i)$. The term σ_{it}^2 in equation iv) is important in that the risk-averse household is always interested in minimizing the uncertainty in $y(X_i)$ and to do this will use all the information on $y(X_i)$ that is available to her. The sources of information to the household are:

- An exogenous information-signal, perhaps from extension agents or microenterprise proponents about the value of his expected return to labor in each period. It is assumed that this signal provides an unbiased estimate of the yield that the household should expect on his own time and that all household in the community get signals of equal precision.
- Neighbor's decisions, used to infer their signals
- His and neighbor's wage (or returns to labor) results

The timing of the receipt of this information is as follow: At the beginning of each time period the household receives his private information signal. In a Bayesian setting the household combines that signal with his prior at the beginning of the period to compute the best-estimate of his expected wage. This in turn determines the time he allocated to non-agriculture in that period. Subsequently, he observes his neighbor's labor allocation decisions, which reveal the signals they received as well as the returns to labor that are realized in the community. With social learning the new information from neighbors' decisions as wells as their wage realizations is used to update the household's prior about the value of his expected return to labor for the next period.

The household's ability to learn from his neighbor's is complicated in the situation in which they do not share the same expected returns to labor. The signal that a neighbor receives or his subsequent wage realization can then no longer be directly applied to the household's own decision, needing conditioning for the differences between his own and his neighbor's characteristics when learning from him. This might not be possible if these characteristics are unobservable, in which case the household might choose to ignore information from his neighbors if the wage is sufficiently sensitive to the unobserved characteristics.

In the following sections I will consider two cases for social learning, one in which the expected wage is the same for all households in the community: $y(X_i) = y$ and the other in which the expected wage depends on unobserved household characteristics. The former case is ideal for social learning while in the second social learning could even break down if wages are sufficiently sensitive to these characteristics.

Social learning with imperfect information but constant expected yield across households

First, I will rewrite equation iv) to be consistent with a linear specification that I intend to estimate:

v) $$L_{it} = \pi_0 + \pi_1 \hat{y}_{it} + g(X_i, \sigma_{it})$$

where *Xi* includes y_{ag} and λ_i from equation iv). It also reflects the household's risk-aversion, which determines the weight that is placed on λ_i, σ_{it} in that equation.

Now I will describe how $\hat{y}_{it}$ is determined. Expected wages are assumed to be constant across households in the village so neighbor's information-signals and wage realizations provide an unbiased estimate of the household's own wage, without modifications. The information-signals that arrive in the community in a given period and the wages that area subsequently realized are pooled together by each household and applied to update his prior in the next period. All households in the village begin with a

common prior $\hat{y}_0$ in period 0. Since they utilize the same information to update their priors in each prior, all households have a common prior in subsequent periods as well. Over time these beliefs converge to the true wage y.

Following the timing of the various information sources that were outlined earlier, each household combines the common prior at the beginning of period t, $\hat{y}_t$ with the private information signal received at the beginning of that period, u_{it} to arrive at his best estimate of the expected returns to labor. Beliefs at the decision time vary across households in the community in each period. It is straightforward to verify that $\hat{y}_{it}$ will be a weighted average of $\hat{y}_t$ and u_{it}, when η_{it}, u_{it} are uniformally distributed, in a Bayesian setting.

vi) $$\hat{y}_{it} = \alpha \hat{y}_t + (1-\alpha)u_{it}$$

Now I will describe how $\hat{y}_t$ is determined. After all households have made their decision in period t-1, we can use the information on their decisions to extract the mean information-signal they received $\bar{u}_{t-1}$ at the beginning of that period. Since no one is systematically misinformed, the mean signal provides more information about the expected wage than any individual signal. Thus $\bar{u}_{t-1}$ supersedes u_{it-1} when it becomes available at the end of period t-1 and is used to compute the household's prior for the next period $\hat{y}_t$. Also, households' wage realizations also become available at the end of period *t-1*. Applying Bayes' Rule once more, the expression for $\hat{y}_t$ is consequently obtained as:

vii) $$\hat{y}_t = (1-\beta-\gamma)\hat{y}_{t-1} + \beta\bar{u}_{t-1} + \gamma\bar{y}_{t-1}$$

where $\hat{y}_{t-1}$ captures all of the information about the expected wage that was received in the community up to the beginning of period t-1 in equation vii). The village means, $\bar{u}_{t-1}$, $\bar{y}_{t-1}$ represent the new information that became available in that period[20]. Substituting from equation vi) into equation v):

viii) $$L_{it} = \pi_0 + \pi_1\alpha\hat{y}_t + \pi_1(1-\alpha)u_{it} + g(X_i, \sigma_{it})$$

Next, substituting the expression for $\hat{y}_t$ from equation vii) into viii):

ix) $$L_{it} = \pi_0 + \pi_1\alpha(1-\beta-\gamma)\hat{y}_{t-1} + \pi_1\alpha\beta\bar{u}_{t-1} + \pi_1\alpha\gamma\bar{y}_{t-1} + \pi_1(1-\alpha)u_{it} + g(X_i, \sigma_{it})$$

While equation ix) completely describes the household learning process, $\hat{y}_{t-1}$ and $\bar{u}_{t-1}$ are not directly observed by the econometrician. I therefore proceed to derive expression for these terms as functions of observed variables, to test for the presence of social learning. I will begin by the $\bar{u}_{t-1}$ term. Lagging equation viii) and taking expectations across all households in the community:

x) $$\bar{u}_{t-1} = \frac{1}{\pi_1(1-a)}[\bar{L}_{t-1} - \pi_0 + \pi_1\alpha\hat{y}_{t-1} - \bar{g}(X_i, \sigma_{it-1})$$

---

[20] The household will actually become more confident about the wage level over time and thus will put more weight on its prior. Thus, α will be increasing over time, while β, γ are decreasing over time.
Also, here I can use the average wage of different types of networks, not just neighbors.

$\hat{y}_{t-1}$ and $\bar{g}(X_i, \sigma_{it-1})$ are common knowledge at the beginning of period t-1, so each household can invert the (linear) time allocation function to compute $\bar{u}_{t-1}$ from the observed time allocation decisions $\bar{L}_{t-1}$ in period *t-1*. Substituting the expression for $\bar{u}_{t-1}$ derived above in equation ix):

xi) $$L_{it} = \pi_0\left(1 - \frac{\alpha\beta}{1-\alpha}\right) + \pi_1\left(1 - \gamma - \frac{\beta}{1-\alpha}\right)\hat{y}_{t-1} + \frac{\alpha\beta}{1-\alpha}\bar{L}_{t-1} + \pi_1\alpha\gamma\,\bar{y}_{t-1} + \xi_{it}$$

where $\xi_{it} = \pi_1(1-\alpha)u_{it} + g(X_i, \sigma_{it}) - \alpha\beta/(1-\alpha)\,\bar{g}(X_i, \sigma_{it-1})$ consists of terms that are unobserved by the econometrician. Now the only thing missing is expressing $\hat{y}_{t-1}$ in terms of observable variables to complete the derivation of the time allocation function. Lagging equation viii) by one period:

xii) $$\hat{y}_{t-1} = \frac{1}{\pi_1\alpha}[L_{it-1} - \pi_1(1-\alpha)u_{it-1} - g(X_i, \sigma_{it-1})$$

It is important to remember that all households have the same prior at the beginning of period t-1. So each household's time allocation decision L_{it-1} provides information (to the econometrician) about the $\hat{y}_{t-1}$ term. Substituting the expression for $\hat{y}_{t-1}$ in equation xi), I finally arrive at a specification of the time allocation function that can be estimated from the data:

xiii) $$L_{it} = \pi_0(\alpha + \beta) + \left(1 - \gamma - \frac{\beta}{1-\alpha}\right)L_{it-1} + \frac{\alpha\beta}{1-\alpha}\bar{L}_{t-1} + \pi_1\alpha\gamma\,\bar{y}_{t-1} + \varepsilon_{it}$$

with $\varepsilon_{it} = \xi_{it} - \left(1 - \gamma - \frac{\beta}{1-\alpha}\right)[\pi_1(1-\alpha)u_{it-1} + g(X_i, \sigma_{it-1})]$ consisting of all terms that continue to be unobserved by the econometricians.

Social learning when the expected yield depends on household's characteristics

In the case when the population is heterogeneous in observable and unobservable ways, neighbor's signals and wages provide a biased estimate of the household's own expected wage and therefore cannot be utilized without modification. Here, the information from a neighbor is less useful than the information that the household receives itself. In such a situation social learning declines immediately as I allow some of the characteristics differing among households to be unobservable or imperfectly observed.

The household may be prepared to accept transitory errors in his wage estimates; associated with the η disturbance term in equation i) or noise in the exogenous information-signals. However, mistakes that arise because he is unable to control for differences between his own and his neighbor's characteristics when learning from their yields are persistent, and therefore more serious. Take the case where all the neighbors' characteristics are unobserved by the household. She has two choices. She could rely on her own information-signals and wage realizations, ignoring information from her neighbors. Consistent but inefficient estimates of the expected wage would be obtained with such individual learning. Alternatively, she could continue to utilize information

from her neighbors, measured by the mean-wage as before. The efficiency of his estimates increased with social learning since more information is being utilized, but some bias will inevitably be introduced since the household cannot control for variation in the underlying determinants of the wage when learning from his neighbors. The household will ultimately choose between individual learning and social learning on the basis of the trade-off between bias and efficiency.