Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Economics

The Effect of Integrated Conservation and Development Programs in Protected Areas on Human Wellbeing

An Empirical Analysis of Brazil's Bolsa Floresta Programme

Elan Swartz

Master's thesis · 30 hec · Advanced level
European Erasmus Mundus Master Program: Agricultural Food and Environmental Policy Analysis (AFEPA)
Degree thesis No 996 · ISSN 1401-4084
Uppsala 2015

The Effect of Integrated Conservation and Development Programs in Protected Areas on Human Wellbeing
An Empirical Analysis of Brazil's Bolsa Floresta Programme

Elan Swartz

Supervisor:	Yves Surrey, Swedish University of Agricultural Sciences, Department of Economics
Assistant supervisor:	Jan Börner, University of Bonn, Department of Economics
Examiner:	Sebastian Hess, Swedish University of Agricultural Sciences, Department of Economics

Credits: 30 hec
Level: A2E
Course title: Degree Project in Economics
Course code: EX0537
Programme/Education: European Erasmus Mundus Master Program: Agricultural Food and Environmental Policy Analysis (AFEPA**)**
Faculty: Faculty of Natural Resources and Agricultural Sciences

Place of publication: Uppsala
Year of publication: 2015
Name of Series: Degree project/SLU, Department of Economics
No: 996
ISSN 1401-4084
Online publication: http://stud.epsilon.slu.se

Key words: Additionality, Amazon, Brazil, Conservation reserves, Economics, Human-wellbeing, ICDP, Impact Evaluation, Matching, Program Evaluation,

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Economics

Acknowledgements

I would like to thank my supervisor at the Swedish University of Agricultural Sciences Prof. Yves Surry and Prof. Jan Börner from the University of Bonn for their supervision and guidance. A special thanks to Elías Cisneros for his commitment and constructive feedback throughout my thesis. I thank my family and friends for making this process a very enjoyable completion to my studies. Finally I would like to thank the European Union which has made partaking in the AFEPA program possible.

Personal Declaration

I hereby confirm that to the best of my knowledge my thesis does not infringe upon copyrighted material nor violate any proprietary rights. To the extent that ideas, techniques, quotations, or any other material from the work of other people are included in my thesis, published or otherwise, are fully acknowledged in accordance with standard referencing practices. I declare that this is an independent work, which has been dually submitted in accordance to the stipulations of the AFEPA program at the Sveriges Lantbruksuniversitet (Swedish University of Agricultural Sciences) and University of Bonn.

SLU, July 17, 2015
Place and date of original submission

Elan Swartz
Student name

Student signature

Abstract

Considering the current debate over which strategies are most effective in promoting environmental and human welfare in protected areas, we focus on Integrated Conservation and Development Programs, an approach which seeks to find 'win-win' outcomes by supporting environmentally sustainable and profitable income generating sources. Counter, the predominant literature, we measure program effectiveness in terms of human welfare, accounted for by relative changes in household income and asset growth. Specifically, we seek to determine the additionality of Integrated Conservation and Development Programs in protected areas over the short run. Using a case study, additionality was determined by comparing two proximate reserves in the Brazilian Bolsa Floresta Programme that received the programme to varying degrees. An impact evaluation approach using parametric and non-parametric techniques was used, demonstrating key findings of no significant change in income and a large (65%) and significant effect on asset growth due to treatment under one matching approach. This dual result may be explained by the narrative where households reinvest program gains in productive assets. However, this narrative may overstate willingness to invest gains considering low levels of household education and risk aversion. Moreover, results in terms of asset growth are likely biased and overstated due to unobservable variable bias and remaining imbalance in key observables- particularly the amount of time participating in the direct cash transfer component of the Bolsa Floresta Programme. Evaluation at a later date when the program has had full potential to be implemented and adopted by households is recommended.

Abbreviations

APA	Rio Negro *Area de Proteção Ambiental* (Environmental Protected Area)
ATE	Average Treatment Effect
ATT	Average Treatment Effect on the Treated
BFF	Bolsa Floresta Familia
BFP	Bolsa Floresta Program
DID	Difference in Differences
FAS	*Fundação Amazonas Sustentável* (Amazonas Sustainable Foundation)
Gen	Genetic Matching
HH	Household
HHH	Household Head
ICDP	Integrated Conservation and Development Programs
MD	Mahalanobis Distance
OLS	Ordinary Least Squares
PES	Payment for Environmental Services
PS	Propensity Score
PSM	Propensity Score Matching
RDS	Rio Negro *Reserve de Desenvolvimento Sustentável* (Sustainable Development Reserve)
REDD+	Reducing Emissions from Deforestation and Forest Degradation
ZEF	Zentrum für Entwicklungsforschung (Centre for Development Research, University of Bonn)

Note from the author: The use of "we" and "our" or any related term should be considered as a stylistic device. This thesis is my own work and I am the only responsible author.

List of Figures

List of Tables

Table of Contents

Introduction

1.1 Problem Background

Poverty remains an important global problem where 1 in 5 people live on less than $1.25 per day, and many households **(HH)** live without access to electricity, drinkable water, sanitary conditions, and sufficient food (Shah, 2013). Furthermore, environmental degradation and climate change have remained at the forefront of policy debate as global emissions and pollutants push 'planetary boundaries' (Rockström *et al.*, 2009). Tackling these issues has mobilized resources on all scales ranging from international agreements such as the Millennium Development Goals and forthcoming Sustainable Development Goals to national, regional and local policies and initiatives.

In an attempt to mitigate climate change, international actors have seen opportunities to compensate for emissions through trading carbon offsets in international programs such as Reducing Emissions from Deforestation and Forest Degradation (**REDD**+)**.** These programs hold a global perspective as emissions and offsets may occur in spatially separate locations.

International donors have pledged significant finances for forest based climate change mitigation; however there is no clear strategy which is the most effective method to manage protected areas and improve the welfare of inhabitants (Nolte *et al.*, 2013). One reason for the strategy debate can be attributed to differences in accessibility which promote both ease of enforcement, but also greater environmental threat (Ferraro & Hanauer, 2014). Moreover, pressure for land is driven by demand for production of food, fuel, extractive resources and local population's need to support their own livelihoods (Butler & Laurance, 2008; Nolte *et al.*, 2013).

There is a body of literature surrounding strategies which promote environmental service provision. Approaches range from prohibiting land use, to incentivising positive outcomes through direct payment, to promoting alternative environmentally sustainable economic activities. Moreover, outcomes are not only a function of incentives and regulations; they are affected monitoring, enforcement and implementation of sanctions (Wunder *et al.*, 2008). Furthermore, in many cases, due to political and logistical barriers, sanctions are not administered, creating opportunities for nonconformity (*ibid*). Although, the causal linkage between poverty alleviation and achieving environmental goals is debated as outcomes are dependent on external factors such as national policies and long term funding (Sunderland et al., 2012); what is clear is that human livelihood strategies are spatially linked in rural environments. This particularly relevant in 'biodiversity hotspots' such as the Amazon (Naughton-Treves *et al.*, 2005).

Although many environmental service project designs account for local inhabitant wellbeing, there is debate over whether environmental and development objectives are complementary or conflicting (Sunderland *et al.*, 2012). On one hand, gains in income may promote more sustainable use of the environment and investments in clean technology and education. On the other, income gains may lead to increased resource demand and act as a driver of further deforestation (Naughton-Treves *et al.*, 2005). This was graphically illustrated by an example where HH use program resources to purchase chainsaws (*ibid*). Evidently income shocks have the potential to affect HH decisions which in turn affect environmental outcomes. Gaining a more clear understanding of specific conservation strategies and how they interact with human wellbeing is relevant.

In the context of protected areas, there is a body of literature on the effectiveness of different management strategies (Honey-Roses *et al.*, 2011; Nelson & Chomitz, 2011; Nolte *et al.*, 2013). However, this literature is solely focused in terms of initiative's primary environmental goals (Andam *et al.*, 2010; Sims, 2010). This thesis fits within this literature gap, and focuses on changes in human wellbeing due to environmental conservation strategies in protected areas. Furthermore, it follows an impact evaluation framework, which accounts for *additionality* of change. Additionality refers to the degree the program has contributed to achieving its desired goal relative to the base-case or unobserved counterfactual scenario where the program was not implemented (Persson & Alpízar, 2013). A case study approach will be followed which is useful in program evaluation as there is heterogeneity in environmental, socio- economic and political factors that affect program implementation and outcomes. However, as Börner et al. note, the contextual nature of case- study based program evaluation results in low external validity and extrapolation of results into broader conclusions (2013).

This thesis has focus on the effectiveness of one management strategy called Integrated Conservation and Development Programs (**ICDP**), in the context of protected areas. This approach seeks to use the synergies between poverty and environmental degradation to simultaneously tackle these issues (McShane *et al.*, 2011). The strategy promotes alternative and sustainable livelihood practices, which both increase environmental awareness as well provide greater return for participants. Thus, a 'win-win' situation is developed (Wunder, 2013). The ICDP narrative follows the logic that cash inflows to communities offer HH a reliable source of income with which to invest and cope with shocks to their environment (Arnold *et al.*, 2011). In some cases this investment has been found to "improve [poverty indicators and] household well-being through impacts on health, purchasing power, household productivity and resource allocation, asset consolidation, and reduce inequality" (Sills *et al.*, 2014, p 65).

The ICDP approach is intuitive as environmental preservation and human use are often not spatially mutually exclusive. They are linked in rural environments which despite conservation status often allows for human occupation (Naughton-Treves *et al.*, 2005). However, although the ICDP intuition is appealing, its track record has been questionable (Hughes & Flintan, 2001; García-Amado *et al.*, 2013; Bauch *et al.*, 2014). ICDP have been criticised for providing small incentives, short time horizon, bad targeting, low reliability and conflicting goals (Winkler, 2007; Blom *et al.*, 2010; Sunderland *et al.*, 2012; García-Amado *et al.*, 2013). Perhaps a more realistic view of the ICDP 'win-win' rhetoric is acknowledging the trade-offs between livelihood and conservation activities, and "looking for synergies between competing land use" (Sunderland *et al.*, 2012, p 237).

1.2 Research Question

Based on this debate, this thesis seeks to answer the following research question: *What is the additionality of community level ICDP programs in protected areas in terms of short run*[1] *changes to HH welfare measured by income and assets*.

One can hypothesis that there are strong barriers to actual accrual of these benefits over the short run, even when HH gain access to program support and expected new income generating sources have high returns. One barrier is the low flexibility of HH to commit their own resources over the short run. Another is that HH may be risk averse which is a function of knowledge and familiarity with practices, as well as preferences and resource constraints. Thus, perhaps it is optimistic to expect significant changes to HH welfare over a short time horizon, regardless of anticipated rates of return. These barriers are mitigated with training, education and time where local users can determine the viability of altering their resource allocation in light of their own constraints. In this light one can expect to witness a greater program effect or additionality over a longer time horizon.

This thesis contributes to the small body of quasi- experimental literature which uses micro- level quantitative evaluation techniques to determine the effectiveness of specific programmes implemented in protected areas (Weber *et al.*, 2011; Miteva *et al.*, 2012; Börner *et al.*, 2013). It answers the call for quantitative impact evaluation based on observed, real- world programs and policies (Ferraro *et al.*, 2011) which leads to better informed policy decisions (Persson & Alpízar, 2013). Conducting program evaluation on projects which target human and environmental well-being have implications on achieving these goals in their own right and influence future project development strategies, donor confidence and funding allocation (Smith, 2000; Börner *et al.*, 2013;

[1] The neoclassic definition of short run incorporates the idea that certain factors of production (often capital) are held constant. Firms or in this case HH, are unable to reallocate all factors to optimize profit due to restrictions in the form of prior investment, or contractual agreement with suppliers. Thus the short run is characterised by a lack of complete flexibility resource allocation (Perloff, 2008).

Davies *et al.*, 2014). Gaining a better understanding of the causal linkages between program designs and outcomes will help mitigate the structural misalignment of project goals and incentives between donors, implementers and evaluators (Ferraro, 2004; Salafsky, 2011).

1.3 Thesis structure

This thesis is structured into 9 chapters. The next section will provide a brief literature review and background to the Bolsa Floresta Programme (**BFP**) and the study area. Section 3 will discuss the theoretical framework which focuses upon the causal pathway of the ICDP on income and asset growth. Section 4 is the empirical approach which highlights the issue of selection bias in observational studies and ways of mitigating it. A discussion of standard parametric and non-parametric approaches and a list of general observable factors which we seek to control for will be given. Section 5 introduces the data used in this thesis and focus upon the specific control variables that allow us to capture the causal pathway discussed in section 3. In section 6 our results are provided. This section selects the best model based on bias reduction. Furthermore, program additionality is measured through the program treatment effect. Section 7 discusses the robustness of results based on remaining covariate imbalance as well as issues of unobservable bias. Section 8 is the discussion which highlights findings in relation to our conceptual model and potential issues with this narrative. In addition, limitations with the study will be discussed. The paper concludes with final thoughts, policy recommendations and opportunities for future research.

Background

2.1 Literature

The rigorous impact evaluation literature of program effectiveness in protected areas is largely measured in terms of environmental outcomes; however there are a few studies which measure protected area effectiveness in terms of poverty indicators (Andam *et al.*, 2010; Sims, 2010). These have found some evidence in contrary to the hypothesis that living in protected areas inhibits development and exacerbates local poverty. Rather, results demonstrate that protected areas can contribute to economic development and reduce poverty (*ibid*). However, these studies are criticised as they provide no recommendation of *how* and *which* method of investment in environmental and human wellbeing is best in achieving these goals within these areas (Bauch *et al.*, 2014).

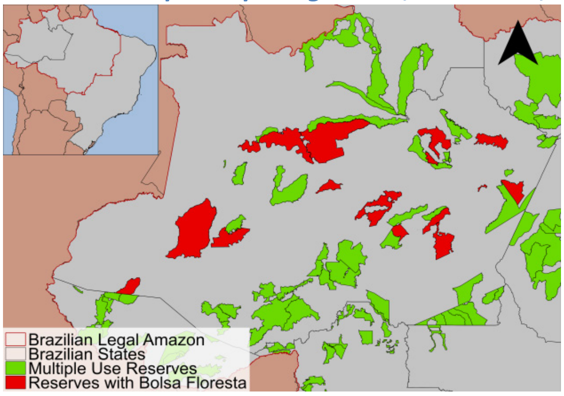
The literature focusing on the specific ICDP strategy within protected areas which using rigorous impact evaluation techniques is very small. One reason for this is the structural misalignment between funders, implementers and donors, making data availability problematic (Davies *et al.*, 2014). The two most relevant sources will be briefly summarized, as they relate directly to this thesis in terms of general location, quasi experimental methodology and focus on human wellbeing. Firstly, Weber et al.'s case study (2011) uses cross sectional data from 2006 to explore the effectiveness of an ICDP implemented in the Brazilian Amazonian state of Pará. Weber et al. (2011) model the HH labour allocation decision between income generating sources where the decision is affected by forgone wage endowment in land, HH labour supply and ability of members. Weber et al. (2011) use a combination of matching and regression analysis over a variety of dependent variables including assets, total income and cash with and without income generated from forest activities to check for robustness. Results demonstrate that participation increases cash and asset accumulation. Furthermore, there is some evidence that the ICDP promotes income diversification as the program adds an additional income generating source which HH use as insurance against shocks (Weber *et al.*, 2011).

Secondly, Bauch et al. (2014) use a panel data set (1997-2006) on the same case study as Weber et al. to explore the effectiveness of an ICDP. The authors use a combination of quasi experimental evaluation techniques including difference in differences (**DID**) and matching. DID is applied in order to control for unobservable time- invariant factors, while matching on baseline pre-program indicators allows comparison 'like with like' (Bauch *et al.*, 2014). Their analysis uses several different definitions of treatment, considering both community level and HH levels and implemented over different points in time. These definitions account for both direct and indirect improvements in welfare which can accrue to HH through the ICDP. Furthermore, these definitions capture differences in participation as a result of program targeting versus participant volunteering. In terms of the welfare dependent variable, key findings include positive point estimate impacts on HH income, but no change in HH assets (Bauch *et al.*, 2014). Using similar techniques, this thesis seeks to determine the impact of ICDP within protected areas. It contributes to the literature by adding a rigorous case study with unique socio- economic and environmental conditions. Broadening the literature base improves the understanding of the effectiveness of specific strategies over the short run.

2.2 The Bolsa Floresta Programme

In order to answer the research question, a case study of the BFP will be used. A brief background to the context of the program will be given. The BFP is an incentive based programme implemented in the state of Amazonia, Brazil as a means to protect environmental resources and improve rural livelihoods in the region. The program has been operated by the NGO *Fundação Amazonas Sustentável* (**FAS**) since its inception in 2007 and seeks to build capacity of local communities and reserves (Sills *et al.*, 2014). As of 2013, the BFP benefits over 7,000 families and 35,000 people in 15 state owned protected areas (refer to Figure 1) (FAS, 2013). These reserves have many communities of permanent residents, which must adhere to conditions relating to activities and management. The BFP is set to continue until 2050 (Sills *et al.*, 2014).

Figure 1: Reserves participating in BFP, Amazonia, Brazil



Source: Elias Cisneros, ZEF, University of Bonn

<u>BFP Components</u>

The BFP has four primary incentive components which operate on different scales. They are interrelated in their goals, however differ in approach.

a) *Bolsa Floresta Renda* (Income component) (Community level)
b) *Bolsa Floresta Social (Social Component)* (Reserve level)
c) *Bolsa Floresta Associação (Association Component*) (Reserve level)
d) *Bolsa Floresta Familiar (Family Component)* (HH level)

Furthermore, these incentive- components are not uniform in their implementation across reserves, communities and HH. Implementation is decided by FAS who choose how and when to implement them. Furthermore, actual participation on a HH level could be affected by conscious (motivation, or sentiments that the program constrains livelihood decisions) or unconscious (lack of information or awareness) motives. The four BFP components simultaneously affect local welfare and behaviour which in turn affects their physical environment. A brief discussion of each component and on how it acts as an inventive strategy will now be provided.

a) *Bolsa Floresta Renda* (Income component):

Under the Bolsa Floresta (**BF**) Income component, reserves receive R$140,000 (USD 84,916) annually which is distributed on a community level (Sills *et al.*, 2014). The funds are used for investment in environmentally sustainable income generating activities that "target on-farm processing activities for value added of existing products and non-timber forest product value chains, or alternative income sources such as eco-tourism, aquaculture, small-livestock breeding, and native honey production" (Börner *et al.*, 2013, p 13). These activities must align with the reserve's management plan and aims to diversify livelihoods (Agustsson *et al.*, 2014). BF Income reduces barriers to entry and costs of switching practices as the investment reduces infrastructure and capital costs, and provides education supporting the alternative income source (Newton *et al.*, 2012). As BF Income seeks to alter HH behaviour in a way that promotes both alternative income generation and environmental outcomes it can be considered as ICDP. [2]

b) *Bolsa Floresta Social (Social Component)*

This component operates at the reserve level with an amount of R$140,000 (USD 84,916) per reserve per year (Sills *et al.*, 2014). In conjunction with public sector institutions, it targets infrastructure improvements in transportation as well as education, sanitation and health services (Börner *et al.*, 2013).

c) *Bolsa Floresta Associação (Association Component*)

This component is also applied at the reserve level and corresponds to 10% of the Family allowance (see Family component) (Sills *et al.*, 2014). This component aims to strengthen local organization and participation (Agustsson *et al.*, 2014) and benefit reserve inhabitants (Börner *et al.*, 2013).

[2] The terms ICDP, treatment and BF Income will be used interchangeably throughout this thesis

The BF Social and BF Association components are unique as they are implemented on the reserve level. They are difficult to categorize as they do not affect environmental outcomes directly. Furthermore, they also do not directly promote alternative livelihood structures. In reality they act as grease to improve the functioning of local society, and provide opportunities for human, physical and social capital development.

d) Bolsa Floresta Familiar (Family Component)

The Bolsa *Floresta Familiar* (**BFF**) component is applied on the HH level, consisting of a monthly payment of R$50 (USD 30) conditional on agreement for adopting 'good forest management practices', which are additional to the basic conditions required to live in the reserve (Sills *et al.*, 2014). Since this payment is insufficient to live off, it acts as an incentive for compliance. BFF requires having been a resident in the reserve for 2 years (Börner *et al.*, 2013).

BFF can be considered as a strategy known as Payment for Environmental Services (**PES**). A brief discussion of PES and how it relates to HH welfare will be provided. PES relies on positive incentives or payments which are conditional on performance or service delivery. Although HH- gains may be relatively small, this income source may be quite effective in cash constrained areas (Wunder *et al.*, 2008). Moreover, PES provides other less tangible positive effects such as improving property right security by recognizing tenure of service providers (*ibid*). Furthermore, it appears that "incentives may contribute to asset building for land user's by improving access to education and capacity building, encouraging cooperation within communities and promoting infrastructure development" (Lima, 2014, p 54). Lastly, gaining program access acts to provide a sustained flow of services to those in need (Landell-Mills *et al.*, 2002). However, these outcomes cannot be considered the norm. There has been debate over whether and under what conditions PES is effective at improving livelihoods (Persson & Alpízar, 2013; Lima, 2014).

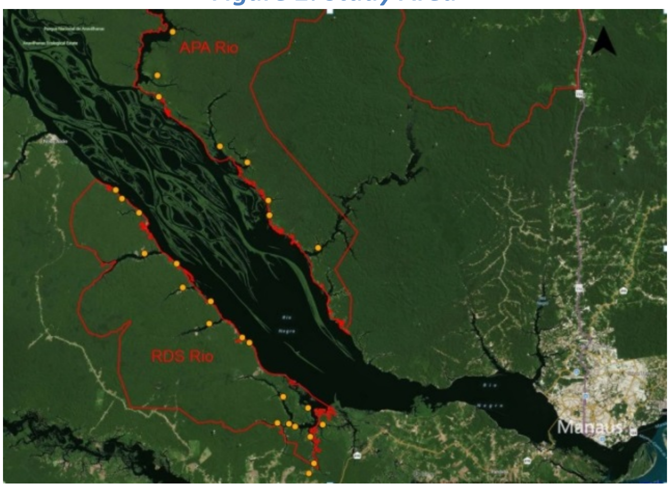
Furthermore, PES has been criticised over promotion of inequitable outcomes as PES targets groups that are not the poorest strata of society given they land owners (Grieg-Gran *et al.*, 2005; Agustsson *et al.*, 2014). Furthermore, PES may alter land value which may lead to appropriation (Grieg-Gran *et al.*, 2005). These issues of inequality have given rise to the idea of coupling PES with side objectives in terms of "poverty alleviation, regional development, employment creation" (Wunder *et al.*, 2008, p 849), which also improves political feasibility (*ibid*). Although these goals are worthwhile, explicit targeting may undermine the ability of PES programs to deliver on their primary goal of environmental service provision. Furthermore, as Wunder at al. (2008)note, targeting the poor is not a necessary condition for them to benefit. Lastly, although it could be argued that voluntary participation covers HH opportunity cost, PES may reduce income generating options which are important to HH sense of satisfaction in their daily activities. In this sense, PES can be thought of as pushing land owners to become 'conservation stewards' (Newton *et al.*, 2012), when in fact they seek allies.

Thus, there are several approaches simultaneously being implemented under the umbrella of the BFP. This multi- strategy approach is reasonable given the complicated nature of program design and how HH decision's and external forces unfold. Approaches are combined in order to gain from component design and follow a holistic approach in tackling desired goals of poverty alleviation and environmental service provision.

2.3 Study Area

This thesis will focus on 2 reserves of the 15 reserves where the BFP has been implemented. These reserves are called Rio Negro Sustainable Development Reserve (Reserve de Desenvolvimento Sustentável (**RDS**)) and Rio Negro Environmental Protected Area (Area de Proteção Ambiental (**APA**)) which are located in the 'deep amazon' (Viana, 2008). The Rio Negro **RDS** is located on the South bank of the Rio Negro river is 1,031 ha in size and has 19 communities and 525 HH (Luiza *et al.*, 2013). The Rio Negro RDS was founded in 2008, and the BFP was implemented in 2009 (Luiza *et al.*, 2013). On the other hand, the Rio Negro **APA** is located on the North bank of the Rio Negro, is 586 ha in size, has 16 communities and 1,300 HH (FAS, 2013). The two reserves are highly proximate to one another and equivalently 70 km northwest from the nearest city Manaus. Although the distance between the two reserves appears small, opportunities for trade are low due to the size of the river and difficulties in crossing due to land bars (refer to Figure 2).

Figure 2: Study Area

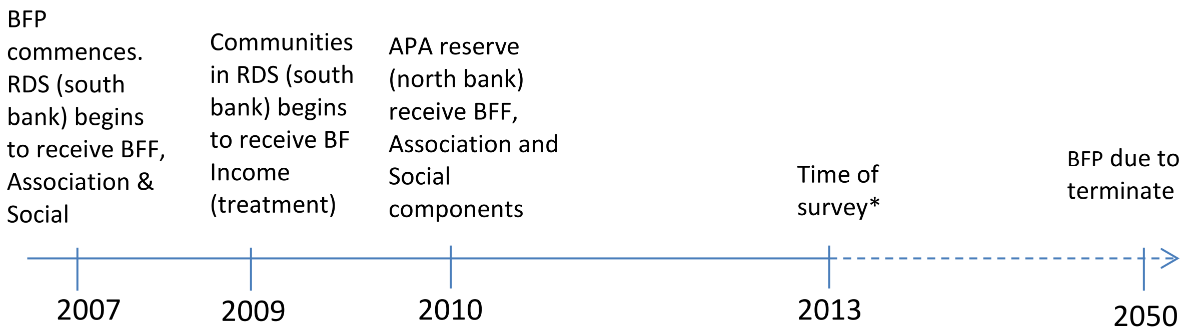


Source: Elias Cisneros, ZEF, University of Bonn

The difference between the reserves names APA and RDS reflects differences in reserve structure in terms of management and conditions for living within its borders. However, it can be argued that the two structures are similar as both have human inhabitants that are allowed to sustainably use reserve resources. Furthermore the lack of real management plan as implemented by the state of Amazonas (and administered by the Secretary of Environment and Sustainable Development) in either reserve adds to the validity of making comparisons between them.

One key difference remains between the APA and RDS which pertains to the degree the BFP was implemented at the time of survey (depicted in Figure 3). By 2013, the time of survey, the southern RDS reserve had received all four aforementioned program aspects of the BFP, while the northern APA had not received the BF Income component. This allows us to use APA as a counterfactual for the treated RDS. It should be noted that BF Income was not implemented immediately and uniformly across all communities in the treated reserve. Rather, starting in 2009 communities began to implement BF Income.

Figure 3: Program Timeline



Source: Own Depiction

*The majority of survey questions pertain to previous year (2012), except for asset value which was also included recollection values for 2007.

The BFP Narrative of Conservation

One of the primary goals of the BFP is to target environmental service provision and improve environmental outcomes, specifically through reducing deforestation. Although not strictly in line with our narrative, a brief theoretical background of the BFP approach to achieving gains in these outcomes will be provided.

The BFP seeks to reduce internally driven deforestation through the BFF PES component. Although BFF has extra conditions on HH activities relative to the standard reserve inhabitant agreement, as Börner et al. (2013) note, these conditions are not significantly different, and rather act as a means of gaining local public support for the program and the values that it holds. This is important when considering the long term approach held by the program. Rather, deforestation from external sources is seen as a more important threat. This is tackled by the program's ICDP and PES components which seek to gain local inhabitants as stakeholders in environmental conservation. It seeks to create local 'conservation allies' and promote stewardship built upon both intrinsic and extrinsic motives as a means to counter highly lucrative deforestation by those living outside the community (García-Amado *et al.*, 2013).

Measuring environmental outcomes are out of the scope of this thesis. Furthermore, one could argue that analysing this long term goal over short time horizons is inappropriate. Moreover, measurement is subject to strategic bias by HH to underreport changes in environmental outcomes. Lastly, we experience data restrictions in measuring these outcomes. For these reasons environmental outcomes will not be further discussed. The theoretical framework will focus on HH welfare as measured by the dependent variables of income and asset growth and how they are affected by the BF Income treatment.

Theoretical Framework

In the context of the BFP case study, my research question and hypothesis will be reiterated. We seek to determine the additionality over the short run of the BF Income component measured by a change to HH income and asset growth. Additionality of treatment is measured relative to the counterfactual where BF Income treatment was received in the RDS reserve, while APA had not at the time of survey. Furthermore, the short run signifies the period between 2009- 2013 over which treatment was implemented. The remainder of this section will discussion the causal pathways of how BF Income affects income and asset growth. These pathways can be categorized into four main component, change in labour allocation, change in demand, social capital development and economies of scale in production. A depiction of these pathways is portrayed in Figure 4.

a) Changes in Labour Allocation

The first way that BF Income affects HH income is through changes in labour allocation between income generating sources. This is an explicit goal of the BF income component, as participants move toward sustainable income generating sources. However there is nuance with regard to this factor for several reasons. Firstly, under the presumption that the BF Income source is more profitable per hour of labour[3], participant's labour allocation may be reduced as income required to support HH needs is satiated. This can be considered as an income affect. On the other hand, the opposite scenario may occur whereby higher per hour wage results in substitution for leisure hours in favour of labour. According to the literature, larger HH have the ability to offer more labour, and thus are more likely to participate to some degree (Weber *et al.*, 2011). Thus, it is unclear the direction which BF Income affects HH income. The overall effect of higher per hour wage in terms of change in labour allocation is determined by the relative magnitude of the income effect and substitution effects for each HH.

b) Changes in Demand

A further income and substitution effect can occur through changes to demand for goods and services. In terms of the income effect, HH may choose to spend additional income (holding labour allocation constant) on consumption goods such as durables that have a positive effect on future income. For example, investment in productivity changing technology such as access to electricity, purchases of equipment etc. may have positive feedback into income generating activities. On the other hand, due to the ICDP prices of goods may vary. For example, as more HH within a community

[3] This presumption is reasonable as BF Income has externally invested resources that should improve productivity. Furthermore, HH would not switch income generating practices if it were not more perceived to be more profitable than their current income generating activity.

produce chicken, the price of chicken may fall resulting in a reduction of HH spending ceteris paribus.

c) Social Capital Development

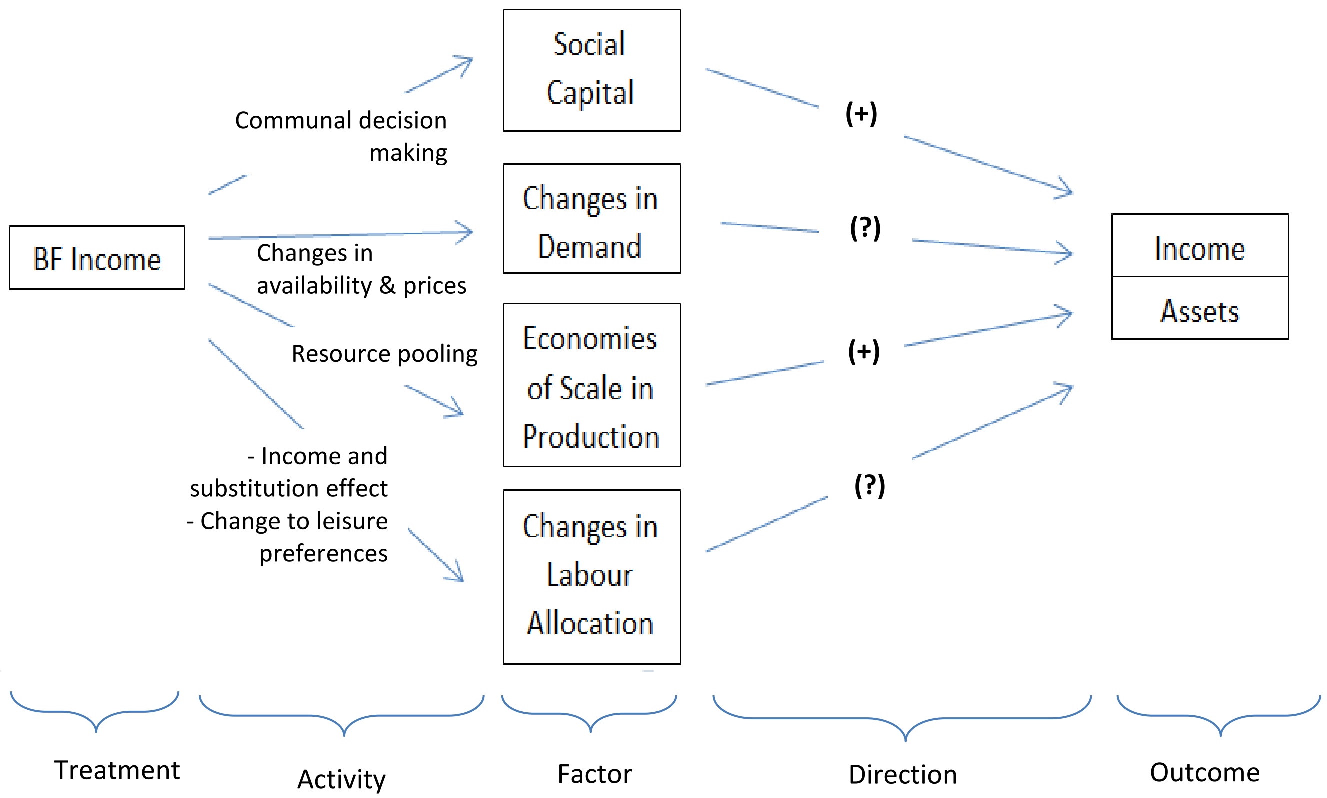
Social capital is important to consider and has positive effect on income for several reasons. Firstly, it affects opportunities which arise in the community. Furthermore, due to better communication, it reduces asymmetric information, which lowers transaction costs and improves markets for goods, credit, land and labour (Narayan & Pritchett, 1999). Moreover, social capital leads to adoption of informal insurance between community members, mitigating HH risk aversion and the probability to invest in and diffusion of productivity enhancing technology (*ibid*). In addition, it creates greater interest in investing in goods with positive externalities (*ibid*). Lastly, social capital helps determine which HH continue participating after program support has ended (Bauch *et al.*, 2014). Thus, social capital affects our dependent variables in many ways through their production and consumption decisions, and their ability to sell and purchase goods and services. In terms of this case study, BF Income affects social cohesion due to community involvement in the decision making process and implementation of the chosen ICDP. In addition, considering the entire community has access to the ICDP, one would expect knowledge sharing and communication of effective practices and modifications to occur between members.

d) Economies of Scale in Production

When many community members participate in the income generating activity, economies of scale can affect HH income by reducing costs.[4] Economies of scale can reduce cost through resource pooling which may lower the marginal cost of inputs and transportation of goods. In addition, resource pooling amongst participants allow for the opportunity to invest further resources (in addition to the ICDP investment) in technology and physical capital which positively affect productivity. Moreover, through labour specialization, efficiency in the production process may be improved. Thus, economies of scale generated by BF Income provide HH with opportunities to combine labour and financial resources as a means to reduce cost and achieve gains in productivity, both of which positively affect income.

[4] As will be explained in the section 5, income is calculated net costs.

Figure 4: Pathways that BF Income Affect Dependant Variables



Source: Own Depiction

Empirical Approach

One measure of additionality is determining the program, or treatment effect relative to the base case scenario or counterfactual, where the counterfactual can be defined as the "beneficiary's outcome in the absence of the intervention" (Khandker *et al.*, 2010, p 43). This section will discuss how the treatment effect can be measured. However, there is an inherent problem in actually observing a counterfactual as participants can only be in one state of being: either treated or untreated. To overcome this constraint, experimental design and statistical techniques are used to develop a valid counterfactual.

4.1 Treatment Effect

There are two relevant parameters which are used as measures of program effectiveness which include the Average Treatment Effect (**ATE**), and the Average Treatment Effect on the Treated (**ATT**). The ATE reflects the expected gains from treatment from a randomly chosen individual of the population (Khandker *et al.*, 2010). The ATT is a narrower measure as it compares average difference in treated and untreated individuals conditional on being in the treated area (*ibid*) .

ATE

Returning to our first measure of program effectiveness, the ATE is defined as the difference in expected outcome between treated and control group, which includes binary treatment status (D= 1 or 0), and outcomes Y(0),Y(1).

$$\tau_{ATE} = E[Y(1)|D = 1] - E[Y(0)|D = 0] \tag{1}$$

ATT

Due to uneven covariate balance between the treatment and control groups, the ATT is commonly measured (Sekhon, 2008). Equation (2) states that the ATT is the difference in mean outcome for the treated group if they were treated $[Y(1)|D = 1]$ and if they were not $[Y(0)|D = 1]$ (Caliendo & Kopeinig, 2008). The true treatment effect is measured in (2), however since the second term (the counterfactual) is unobservable (note the observable terms are highlighted in red); an observable proxy must be incorporated. Furthermore, although it is appealing to simply replace the second term of (2) with the untreated sample $E[Y(0)|D = 0]$, this is not valid when groups were not randomized.

$$\tau_{ATT} = E[Y(1)|D = 1] - E[Y(0)|D = 1] \tag{2}$$

Therefore, we add and subtract $E[Y(0)|D = 0]$ from the right hand side, creating (3), and then re-arrange, putting the different observed states of nature on the left side, creating (4). The two remaining terms on the right hand side depict selection bias. If the difference between these two terms is zero, as portrayed in (5), selection bias is not a problem and we can accurately measure the treatment effect (τ_{ATT}). Selection bias can be thought of as the difference between this true counterfactual (left hand side of (5)), and the observable one that we use as its proxy (right hand side of (5)).

$$\tau_{ATT} = E[Y(1)|D=1] - E[Y(0)|D=1] + E[Y(0)|D=0] - E[Y(0)|D=0] \quad (3)$$

$$E[Y(1)|D=1] - E[Y(0)|D=0] = \tau_{ATT} + E[Y(0)|D=1] - E[Y(0)|D=0] \quad (4)$$

$$E[Y(0)|D=1] = E[Y(0)|D=0] \quad (5)$$

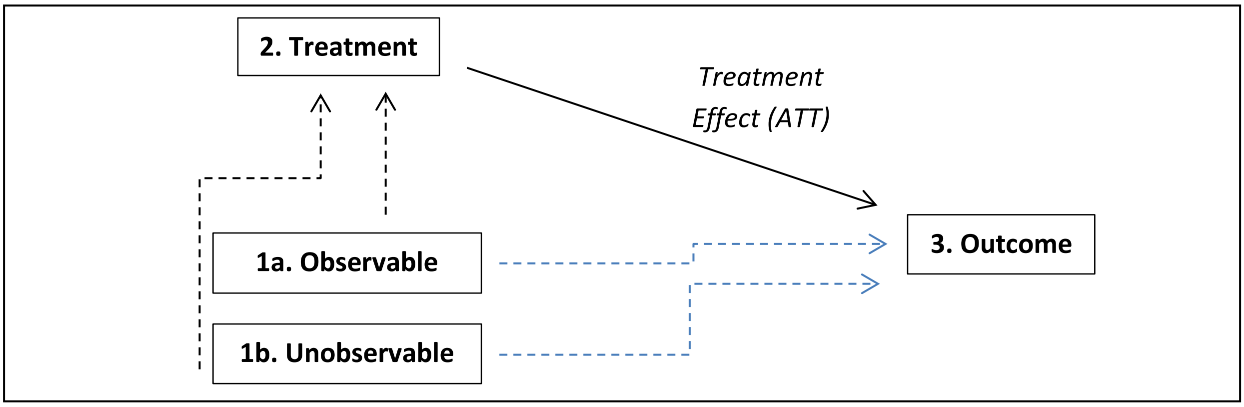
4.2 Selection Bias

Now that the effects of selection bias on treatment effect estimation have been depicted, it is necessary to understand how it occurs and why it is an issue. In particular, selection bias occurs when certain characteristics affect both an individual's likelihood to participate *and* realized outcomes (Heinrich *et al.*, 2010). This issue can be illustrated with a common example of positive selection bias. If individuals with higher levels of education choose to receive job training treatment, their wage outcomes will likely be higher than the untreated group not only because they received the additional job training, but because they have higher education (used as a proxy for productivity) as well. This issue can be based on both observed characteristics, or unobserved characteristics (such as motivation which may cause self- selection) which are found in the error term (Khandker *et al.*, 2010). Selection bias is key issue which causes estimates of program effect to be biased and impact to be miss-estimated (Heinrich *et al.*, 2010).

Randomization

One technique to solve for selection bias and isolate for the treatment effect is through a randomized experimental approach. This is the gold standard method in accounting for selection bias and accounts for both observable and unobservable in order to determine causality (Shadish *et al.*, 2008). Under a randomized approach, pre- program, participants are randomly selected for treatment and control groups. Thus, one can say that on average there are no observable or unobservable differences between these two groups (Smith, 2000). With a sufficient sample size, all individual's characteristics will be balanced between the two groups and the difference in outcomes can be attributed only to the effect of the treatment or program. Formally, this can be written as: $\{Y(0), Y(1) \perp D\}$, which is interpreted as outcomes $Y(0)$ and $Y(1)$ are independent of treatment status (D) (Andam *et al.*, 2010). To further explain our process; under randomization we do not need to control for any attributes as on average both observable and unobservable ones are equal between the two groups (Khandker et al., 2010). Thus the treatment effect can be directly estimated.

Figure 5: Depiction of Selection Bias:



Source: own depiction
Remark: note the direction of arrows, where feedback between treatment and outcome with 1) do not occur
- 1a includes all relevant factors which affect treatment and outcome. Thus under unconfoundedness 1b) is irrelevant by assumption. Controlling for 1a) on 3) allows estimation of the parameter of interest the ATT

4.3 Unconfoundedness/ Conditional Independence

However, random assignment is not the norm, whereby a commonly used approach is to 'replicate' randomization. This imitation relies on the unconfoundedness or conditional independence assumption which is developed by economic theory and applied to both parametric and non-parametric methods (Khandker et al., 2010). This assumption posits that *only* observable characteristics are relevant in determining treatment status and potential outcomes are independent of treatment status (Caliendo & Kopeinig, 2008). Therefore, by assumption, unobservable ones are insignificant sources of bias (Imbens and Wooldridge, 2009). Another way to think about it is that before the program was implemented, or if the program had not occurred, measures of welfare should be equal conditional on observable covariates. Mathematically, unconfoundedness can be expressed as follows $\{Y(0), Y(1) \perp D|X, \forall \ X\}$, where Y(0),$Y$(1) are outcomes, *D* is treatment status, and *X* are observable control covariates (Khandker et al., 2010). Referring to Figure 5, when controlling for observable characteristics (1a on 3), the treatment effect (2 on 3) can be estimated. After controlling for covariates *X*, treatment assignment is as good as random which mitigates selection bias. Under conditional independence, equation 5) can now be re-written as 6) holding the same interpretation as before, except outcomes are also conditional on observable covariates X.

$$E[Y(1)|X, D=1] - E[Y(0)|X, D=0] = \tau_{ATT} + E[Y(0)|X, D=1] - E[Y(0)|X, D=0] \qquad (6)$$

However, it should be noted that if one is unable to achieve complete pseudo randomization, selection bias may remain. Furthermore, due to the unobservable term in our selection bias equation, measuring selection bias remains difficult, although one can hypothesis the direction of bias.

4.4 OLS

This subsection will briefly discuss the use of the commonly used parametric Ordinary Least Squares (**OLS**) design in the context of selection bias. OLS has similar assumptions of unconfoundedness and selection on observable characteristics, making it a useful benchmark in ATT estimation. To estimate ATT with OLS we run a regression on outcome variables conditioning on observable covariates *X* and treatment status *D*. The OLS regression can be defined as follows: $y_i = x_i'\beta + (\tau_{ATT})D_i + \varepsilon_i$, where y_i is the dependent variable, x_i are the set of control covariates with corresponding β coefficents. D_i is the treatment dummy with coefficient τ_{ATT}, and ε_i is the error term which contains all unobserved characteristics that effect outcome. The error term may also contain unobservable characteristics which are related to both treatment and outcome. Thus, $cov(D, \varepsilon_i) \neq 0$, which creates unobservable bias in our τ_{ATT} estimate (Khandker *et al.*, 2010). This is not an issue for a well-specified model.

Some authors find that OLS performs relatively well in comparison to experimental outcomes (Shadish *et al.*, 2008), however there are a few circumstance where it may not be preferable. Firstly, since OLS is a parametric, imposing a restrictive functional form may not be satisfied. This is particularly an issue when covariate distributions are very different between treated and control groups (Ferraro & Miranda, 2014). Moreover, controlling for large number of covariates in OLS affects outcomes, especially when sample sizes are small. Given these caveats, non-parametric approaches such as matching which seek to improve covariate balance and mitigate the issue of dimensionality are available. Matching will be discussed in detail momentarily.

4.5 Quasi-Experimental Approach

There are several so called quasi- experimental approaches which seek to account for selection bias and allow for unbiased ATT estimation. These methods have been shown to be particularly useful in the fields of biodiversity and conservation (Miteva *et al.*, 2012, p 72). Deciding upon which quasi-experimental approach depends on which assumptions are most valid in the specific case and based on available data. A brief discussion of common quasi experimental approaches which include techniques such as difference in differences, instrumental variables, and regression discontinuity design will be held in Appendix A.1; however, for the moment, we can rule out these options based on data availability. Since in our case study, one time period of data is available, difference in differences is not an option. Moreover, lack of a strong and valid instrument negates use of instrumental variables. In addition, regression discontinuity design is not applicable as we are unclear about the criteria FAS used in their decision to implement BF Income in the RDS over the APA. Although single difference is an option, its assumptions are strong and analysis could be considered naïve as it assumes no selection bias. Thus, this thesis will focus on matching, which will be discussed in the next section.

4.6 Matching Approach

Matching is a non-parametric method of estimating causal inference (Sekhon, 2008) and seeks to replicate randomization. The difference is that while randomization creates balance over all observable and unobservable characteristics, matching only creates balance on observable ones (Heinrich *et al.*, 2010). Like OLS, matching also relies on the unconfoundedness assumption, which controls for observable characteristics which affect outcome. This can be an issue if the model is miss- specified and characteristics which are relevant in influencing treatment status are omitted. Thus, matching suffers from the same issue as OLS where selection bias from unobservable covariates remains an issue.

Matching can be thought of as a form of weighted regression. It differs from OLS as it is non-parametric and seeks to improve covariate balance. Where OLS uses the full sample, matching generally reduces the number of control observations and only accounts for ones that are 'good fits' for the treatment group. If balance between the two groups is adequate in the full sample, OLS and matching results will be similar. Returning to Figure 5, we remain interested in determining the treatment effect. As with OLS, under matching we control for 1a) on 2); however, now we also seek balance across these same covariates by treatment status (2). If matching creates balance over all observable covariates, treatment can be considered random. If this is not possible and poor balance remains, then treatment status is not entirely random based on observable characteristics and selection bias remains.

4.6.1 Common Support / Overlap

In addition to unconfoundedness, matching holds a testable caveat known as common support or overlap. This caveat relates to how well the control group fits the treatment group after controlling for observable characteristics. This condition ensures that individuals with the same outcome in terms of matching covariates have a positive probability of being in either the treatment or control group (Caliendo & Kopeinig, 2008). Mathematically common support can be expressed as follows: $overlap = \ 0 < P(D = 1|X) < 1$, where the probability (*P*) of receiving/ not receiving treatment (*D*) conditional on covariates *X*, lies between 0 and 1 respectively. Furthermore, a common problem is that matching techniques make balance on certain covariates worse (Sekhon, 2008). Although analysis must account for this change over specific covariates and their theoretical importance, there is no consensus in the literature how to best measure balance (Diamond & Sekhon, 2013).

4.6.2 Control Factors

There are two reasons to select control factors in matching. The first is due to selection bias. Since we do not know the criteria used to implement BF Income in the RDS reserve over the APA, we desire factors which we believe could affect both the decision to gain treatment and our outcome. For example, perhaps treatment was decided upon based on RDS having better infrastructure or members with higher average education. In addition, since treatment assignment was not random, we must control for other differences in community characteristics, which enables 'like with like' comparison. Before diving into the specific data and choice of actual covariates used in this study, it is worthwhile to discuss general factors which meet the unconfoundedness criteria and affect the dependent variables. These control variables will be categorized by how they affect the dependent variable, rather than by type of characteristic.

i) Ability/ Productivity

Firstly, it is important to account for ability and productivity within a HH. These factors translate into efficiency in using resources to produce an output, affect HH strategy decisions, and knowledge and understanding of the environment in which they live. These variables represent characteristics of individuals within the HH that are innate or developed overtime. One could expect HH with higher ability or productivity to have higher incomes.

ii) Social Capital

As discussed in the theoretical framework, social capital positively affects income through many paths; however its exogeneity with respect to treatment can be questioned as one could argue that treatment affects social cohesion due to community involvement in the decision making process and implementation of the chosen ICDP. To mitigate this issue, selected covariates to proxy social capital must remain constant throughout the implementation of the ICDP. Thus, these covariates will not be affected by the ICDP implementation.

iii) Endowment

Endowment in the form of goods is a proxy for wealth. It affects HH ability to gain income and access to income generating possibilities and technology that affects productivity. Furthermore, endowment can be seen as a proxy for unobservable 'past HH ability' to generate wealth, which is correlated with current HH ability. Endogeneity may be an issue with these factors, as income may have a return effect on HH productive assets. Endogeneity can be mitigated if we assume these factors are costly durables and take a long time to procure.

iv) Economies of Scale within HH

It is important to account for economies of scale within a HH. These factors affect demand for income or assets through task specialization and resource pooling.

v) Community Factors/ Access to Infrastructure

These factors affect HH access to infrastructure and institutions which relate to their ability to generate income and assets. It also alters time allocation between different sources and thus changes the opportunity cost of partaking in certain activities. Moreover, community factors include the innate community environment and its spill over onto specific HH. We should note that this group of variables are slightly different to the other groups. Firstly, as the name suggests community factors are measured at the community level rather than the HH level which gives them less variation. Moreover, it also characterizes differences in the overall community environment which we seek to control for.

4.6.4 Dimensionality and Balancing scores

Matching relies on controlling for observable characteristics that affect treatment. When one is only controlling for a single characteristic this is straightforward; however finding a valid match based on a high dimensional vector of characteristics is more difficult (Heinrich *et al.*, 2010). Introducing a balancing score reduces the issue of dimensionality and retains the same economic intuition of controlling for observable characteristics.

4.6.5 Distance Measures

Propensity Score Matching

Propensity Score Matching (**PSM**) as introduced by Rosenbaum and Rubin (1983) combines all covariates into a scalar. Rosenbaum and Rubin (1983) demonstrate that when it is valid to match based on covariates X (that meet conditional independence assumption), it is also valid to match based on propensity score (**PS**). This method of matching provides an unbiased estimate of the treatment effect, which is the average outcome difference between matched (based on a scalar PS), treated and controls (Caliendo & Kopeinig, 2008; Heinrich *et al.*, 2010). Post matching PS differences between treatment and control groups are an indicator of poor balance. This indicates model misspecification or too small a sample size as bias is asymptotically removed with PSM (Diamond & Sekhon, 2013). The PS is derived through running logistic regression on the treatment variable, and controlling for all exogenous characteristics (Caliendo & Kopeinig, 2008).

Mahalanobis Distance

Mahalanobis Distance (**MD**) makes comparison between treatment and control groups through taking the MD between each unit *i* and the closest unit in the opposite group. The average (scalar) distance for each covariate is reported (King *et al.*, 2011). The MD measure can be defined as follows:

$$MD(X_i, X_j) = \sqrt{(X_i - X_j)^T S^{-1} (X_i - X_j)} \qquad (7)$$

where S is the sample covariance matrix of X used as a normalization factor, and X^T is the transpose (King *et al.*, 2011). MD can be criticised as it regards all interactions among the covariates X as equally important. When X has high dimensionality, additional covariates has difficulty in matching a greater number of interactions (Stuart, 2010). Furthermore, non- normal distribution of covariates is also an issue (*ibid*).

Genetic Matching

Under all matching algorithms, researchers must decide on which matching covariates to include and how to test for balance. The Genetic Matching algorithm[5] endogenizes the process of determining optimal weights for chosen covariates through an iterative process (Sekhon, 2008). These weights are chosen in order to maximize overall balance for observed covariates (*ibid*). This approach is more general than PSM and MD, as results can replicate other approaches under certain circumstances. Although in reality neither of these situations are likely to occur, this generalized iterative form "dominates other matching methods in terms of MSE [mean squared error]" (Sekhon, 2008, p 7).

4.6.6 Further Matching Algorithms

The distance measures listed above can incorporate various matching techniques including replacement, nearest neighbour and caliper matching each which further affect both achieved balance and results. In general all model specification decisions were made in order to reduce bias. This occurred at the expense of efficiency or variance.

Firstly, we will consider matching with or without *replacement*. Replacement indicates that after a control unit was used, it was replaced into the set of controls for future potential matches. Replacement is intuitive as it allows for the closest possible match to be made which decreases bias because controls look more similar to treated individuals (Stuart, 2010). However replacement runs

[5] Available from the 'rgenoud' package in R

the risk that relatively few control observations are used to construct the counterfactual (Smith & Todd, 2005; Caliendo & Kopeinig, 2008; Khandker *et al.*, 2010). For our study, matching with replacement was conducted in order to reduce bias.

Nearest Neighbour (**NN**) selects the number of nearest control units used to determine the counterfactual. For this study, 1:1 matching was conducted, as it is intuitive that the single nearest neighbour provides the closest possible match for treated individuals. Although selecting multiple controls (or multiple NN) for each match decreases variance, it does not account for the fact some individuals may have many close matches while others do not (Stuart, 2010). This creates a trade-off between the degree of sample balance with the matched sample size (King *et al.*, 2011). For our study, 1:1 NN matching was performed in order to reduce bias.

This issue of match proximity can be accounted for using a *caliper*. Calipers imply that matches conducted must fall within a certain threshold in terms of standard deviation. When the caliper criterion is not met, treatment individuals are dropped from the sample. Specifying a small caliper may cause many observations to be dropped (Lunt, 2014). Narrower calipers results in closer matches, thereby reducing systematic differences and bias; "however it may also result in the reduction in matched subjects, thereby increas[ing] the variance of the estimated treatment effect" (Austin, 2011, p 150). In general, aside from Austin who recommend using a caliper which is 0.2 of the pooled standard deviation of the logit of the propensity score, a range of calipers are used in estimation (Lunt, 2014). For this thesis a variety of calipers on the PSM models were tested in order to improve fit.

Selected Matching Approach

Matching seeks to improve balance over our covariates. However, matching may also deteriorate balance over certain covariate (King *et al.*, 2011). If this occurs, the counterfactual cannot be considered random over treatment status, resulting in biased outcomes. Although it can be argued that excluding characteristics where balance is not achieved improves balance, failing to account for relevant characteristics results in model misspecification and unobservable variable bias. This thesis falls within the latter category, whereby all relevant characteristics were selected pre matching. The specific matching approach was determined in order to maximize balance, which may occur at the expense of covariate variance (Lee, 2013).

4.6.7 Match Quality

There is no unique measure of match balance, thus several will be explored in this thesis. First, a test of joint equality of means across all covariates was conducted for each algorithm. The null hypothesis is of joint equality for all tests. This is an issue when conducting matching with replacement. Second, t-tests were used to determine difference in means for each covariate, where a significant p-value is an indication of covariate imbalance. However, as Lee (2013) notes, both the F- test and t-test assume observation independence. Third, as suggested by Rosenbaum and Rubin (1985), the standardized difference in means[6] (also known as standardized bias) between treatment and controls is a strong indicator of balance. This metric is useful as it is unitless and thus accounts for differences in scaling between covariates. There are two specifications of standardized difference, including pooled and unpooled variance.[7] In this study, the former will be used. It can be described as the difference in sample means in treated and matched control as a percentage of the square root of average variances in both groups (Caliendo & Kopeinig, 2008, p 19).

$$\mathrm{SB_{Before}} = 100\,\frac{(\bar{X}_1 - \bar{X}_0)}{\sqrt{.5\,(V_1(X) * V_0(X))}} \qquad \mathrm{SB_{After}} = 100\,\frac{(\bar{X}_{1M} - \bar{X}_{0M})}{\sqrt{.5\,(V_{1M}(X) * V_{0M}(X))}} \tag{8}$$

SB represent the before and after standard bias, $\bar{X}_1$and $\bar{X}_0$ are the sample means of the treatment and control, and $V_{1M}(X)$, $V_{0M}(X), V_1(X)$, $V_0(X)$ are the sample variances for matched and unmatched groups. Rosenbaum (1985)suggest looking at variables that have at least 20% difference before matching, however we remain sceptical since having good balance in a covariate before matching does not indicate good balance after matching.

Moreover, the standard mean difference can be manipulated in order to obtain the percent bias reduction. This can be simply defined as $([1 - (\mathrm{SB_{Before}}/\mathrm{SB_{After}})] * 100)$. Thus, we are normalizing the change in standard difference by the original before matching difference. The final two tests are less important in determining model performance in terms of balance, but rather indicate overall model fit and variance. These include, Empirical Quantile- Quantile (eQQ) mean standard difference which is a non-parametric test that evaluates the rank of the observations (Ho *et al.*, 2007). As the mean eQQ difference shrinks to zero, so does the difference in distributions between the treatment and control groups.[8] Lastly, the variance ratio is a metric which indicates the variance of the treatment group relative to the control. A variance ratio of one indicates equal variance between the two groups (Sekhon, 2008). Thus, a variety of measures were used to calculate match balance. The models which performed best in this regard were preferred for further testing and analysis.

---

[6] The absolute standardized mean difference was taken for comparison purposes

[7] The standard difference unpooled normalizes the mean by the SD of the treated group

[8] The KS test statistic is another measure which captures differences over the entire distribution for a covariate. However, this statistic is not available for all covariates, and thus will not be used.

4.3.6 Testing for Robustness

Although we seek to find an adequately balanced sample which allows for unbiased inference into the ATT, remaining bias in observable covariates is an issue. Furthermore, the size of bias created by covariate imbalance cannot be directly tested. Rather the direction of bias can be speculated. One recommendation to handle remaining imbalance on key covariates is to perform exact matching (Lee, 2013), although this technique is out of the scope of this thesis.

Moreover, the ATT may be subject to hidden bias caused by unobserved factors which have explanatory power in determining treatment. In this regard, there is a model misspecification test known as Rosenbaum's Bounds which accounts for relevant yet unmeasured covariates. The test "measures the degree of departure from random assignment of treatment" (Keele, 2010, p 7). It seeks to determine the magnitude of hidden bias through adding an unobserved factor that confounds the observed treatment effect (Caliendo & Kopeinig, 2008). This unobserved factor signifies all unobservable characteristics in the model.

This test can be illustrated as follows: Under the no bias situation, individuals with the same covariate values have equal probability of receiving treatment. However, when hidden bias exists, individuals with the same covariate value may have different probabilities of receiving treatment (Keele, 2010). Rosenbaum's Bounds test uses an odds ratio captured by a sensitivity parameter Γ. There are two relevant tests which use the parameter Γ. Firstly, the Wilcoxon signed rank test compares how the p-value of the ATT changes under various thresholds of hidden bias (Γ) relative to the commonly used 5% significance level (Keele, 2010). A second test is the Hodges- Lehmann test which incorporates the additional effect of treatment using a confidence interval. This method "can be roughly interpreted as the difference in medians across treatment and control groups" (Keele, 2010, p 14).

The literature mentions the size of Γ indicates sensitivity to unobservable factors. Some authors state that $\Gamma<2$ indicates low sensitivity to unobservable characteristics (Duvendack & Palmer-Jones, 2012), however this appears highly cautious. Furthermore, even a low value of Γ does not necessarily indicate there is no effect of treatment. Rather it can be interpreted that the confidence interval for the effect includes zero, pointing to no significance of treatment (Becker & Caliendo, 2007). Therefore, low values of Γ indicate caution in interpreting results, especially when considering that matching was conducted under untestable unconfoundedness assumption, available covariates, and choice of match specification.

Data and Selected Variables

Primary data was used for this study collected by Zentrum für Entwicklungforchung (**ZEF**), the University of Bonn, and the FAS. The data was collected by 6 enumerators between November 2013 and February 2014 from the APA and RDS reserves in Amazonia, Brazil. Two separate surveys were conducted, including one on the HH and level and one on the community level.

5.1 Sources and Cleaning

The HH survey used a stratified sample with the pre survey goal to randomly obtain 15% of each treated RDS community and 45% from each APA control. In reality, all 19 communities in the treated RDS reserve were surveyed, obtaining a recorded survey proportion of 15% and 87 HH observations. On the other hand, from the control APA reserves only 8 of the 16 communities in the APA were sampled due to time constraints, and the conscious decision to leave out 3 communities due to known differences ethnic structure. Thus, a recorded survey proportion of 45% and 127 HH observations were obtained in the control communities. In sum, this lead to a total sample size of 214. The data was recorded in Microsoft access, and converted into Stata for analysis.

The HH questionnaire[9] covered the following topics:

1. Basic information about HH members
   - Number of members, gender, age, education, relationship to the HH Head (**HHH**) and ethnic and religious affiliation
1. HH Assets
   - Land (owned/controlled), housing characteristics and value for a range of durable goods in both 2007 and 2013
2. Bolsa Floresta Income and Family
   - Program participation, activities and benefits
3. HH income
   - Agriculture (production of specific goods, costs, hours worked)
   - Livestock and animal (production of specific goods, costs, hours worked)
   - Products from the forest and environment (production of specific goods, costs, hours worked, family forest use)
   - Salary from business or contractual labour (type, hours, value)
   - Community work (type, hours, value)
   - Other sources (type, value)

The second community survey, interviewed the leader of each community. Leaders were asked questions relating the community's demographic distribution, access to infrastructure (including education, health and markets), natural resource use and availability, as well as extreme events within the community, community benefits received through the BFP and the decision making

[9] Contact author, or ZEF for original Portuguese survey

process used to allocate them. In addition, self-reported characteristics of leadership were asked. This survey provided 19 observations from the treated sample and 8 from the control, providing a total sample of size 27 from the two reserves.

Data Cleaning

Missing values occurred if the respondent did not know the answer to questions, did not want to respond, were wrongly entered in the original questionnaire, or the question was left unanswered. Rigorous checks regarding inconsistencies in responses were conducted with reference to the primary 'paper' survey and the digitalized version.

Due to the small sample size, missing values were imputed from available information rather than dropped. In general this was only a significant issue for the factors which were used to generate the dependant variables however occurred for some independent variables as well. Furthermore, there were many inconsistencies in reported values, which required personal judgement in order to select the appropriate value for analysis. This can be illustrated using the example where HH were asked prices, quantities and total value for a specific good. Although the total value of output (revenue) should by definition correspond to the product of the price and quantity, this identity did not always hold. In each case, best judgement was used to correct issues. This allowed for a complete data set for to be obtained.

5.2 Dependent Variables

As per the literature, since poverty is a multidimensional concept (Alkire & Foster, 2011), several dependant variables were used to capture livelihood and wellbeing. Thus, for this study, two main continuous dependent variables, including 'log per capita income'[10] and 'change in log assets' or asset growth were tested in order to determine changes in HH livelihood and wellbeing.[11] Summary statistics for these dependant variables before and after matching will be presented in the next chapter (Table 11). Furthermore, log transformations were made in order to give lower weights to dependent variable outliers, capture decreasing economies of scale and for ease of interpretation.

Income can be used as a measure of welfare as it is one of the goals of the project, and provides HH with liquid assets for investment and livelihood needs. In the context of this survey it was measured for the year preceding the 2013 survey date. Thus, our treatment effect seeks to determine if income between our two study groups is different after controlling for relevant covariates. On the

[10] Referred to as income from now on

[11] Another measure could include cash income. This was not included as it focuses on wage labour which would be understated for subsistence HH and family derived labour.

other hand, the other in asset value depicts a change in the stock of HH wealth between the two time periods 2007 and 2013. It can be interpreted as percent asset growth.

Assets

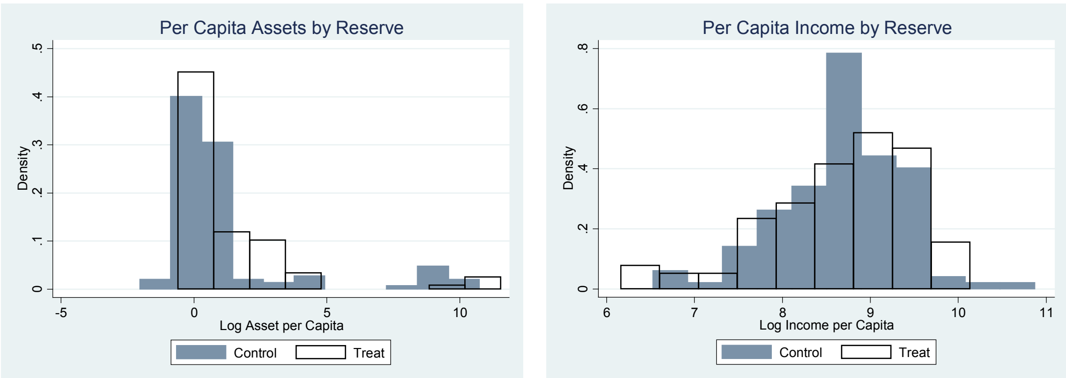
We will first provide details with the calculation of our asset growth dependant variable, starting with simply asset value. From equation 9), asset value ($V_{(asset)}$) is the sum of (J) assets at time period (t). Furthermore, as shown in equation 10), we take the log difference of asset value between each of the time periods (t_1, t_2). Moreover, 1 was added to asset value before taking the logarithm in order to account for zero asset value at a certain time period. Thus, our final dependant variable is log change in asset value or rate of asset growth.

$$V_{(asset)} = \left(\sum_{j=1}^{J} (v_{tj}) \right) \tag{9}$$

$$\Delta V_{log\ (asset)} = \log\left(\sum_{j=1}^{J} (v_{t_2}) + 1\right) - \log\left(\sum_{j=1}^{J} (v_{t_1}) + 1\right) \tag{10}$$

Before matching, asset growth has a distribution as indicated in the Figure 6 (left). It appears a majority of HH do not experience a change in assets, however the mean, and median are positive indicating positive changes in welfare. By visual inspection, one can speculate no treatment effect as the distributions of the dependant variable fit relatively well. However, this inspection does not include control covariates.

Figure 6: Distribution of Dependant Variables (Before Matching)



[12]

Source: own depication

[12] For illustrative purposes this graph omits 1 outlier for the control reserve that has a negative reported income (-1.36).

Income

Total annual HH income can be considered as the sum of direct income from production and subsidies from all HH members, minus the tangible costs of production (refer to Table 1). Income is generated from a variety of sources including agricultural products, livestock[13], meat and other products and services derived from animals, products collected from the forest and environment, income from family business, contract labour, community work, and other payments.

$$Y_i = \sum_{j=1}^{J}\sum_{k=1}^{K} r_{ijk} - \sum_{w=1}^{W}\sum_{x=1}^{X} c_{iwx} \tag{11}$$

Referring to Equation 11), the first term of right hand side can be interpreted as HH revenue gained (r_{ijk}), which is the sum of *J* income categories that have *K* components irrespective of which HH member produced the income. The second term are the total costs (c_{iwx}), which are the summed over (*W)* cost categories and (X) cost components. Combing these terms yield the desired the HH total income (Y_i).

Furthermore, we are interested in per capita income, which requires the adult equivalent HH size $(N_{adult\ eq})$ to be calculated (Equation 12). This equation seeks to explain gains from economies of scale within a HH. This equation can be illustrated using a sample HH. A single adult HH has an adult equivalence of 1, while 2 adults HH would have an adult equivalence of 1.7. Since children do not live alone, each additional child accounts for 0.5 adult equivalent members. Dividing Y_i by the number of adult equivalent members yields the desired result of per capita HH income (Y_{pc_i}). Lastly, the natural logarithm of the dependent variable was taken $(\log(Y_{pc_{HH}})$ as specified in Equation 13.

$$N_{adult\ eq} = 1 + 0.7(N_{adult} - 1) + 0.5(N_{child}) \tag{12}$$

$$\log(Y_{pc_i}) = log\left(Y_i / N_{adult\ eq}\right) \tag{13}$$

To reiterate $N_{adult\ eq}$ is the adult equivalent HH size, (Y_i) is the total HH income, and $\log(Y_{pc_i})$ is the logged per capita HH income. The distribution of the income dependant variable is illustrated in Figure 6 (right).

[13] It should be noted that there were certain measurement issues with regard to the value of livestock variable. In the robustness section, this will be further tested.

Table 1: Description of Income and Cost Categories

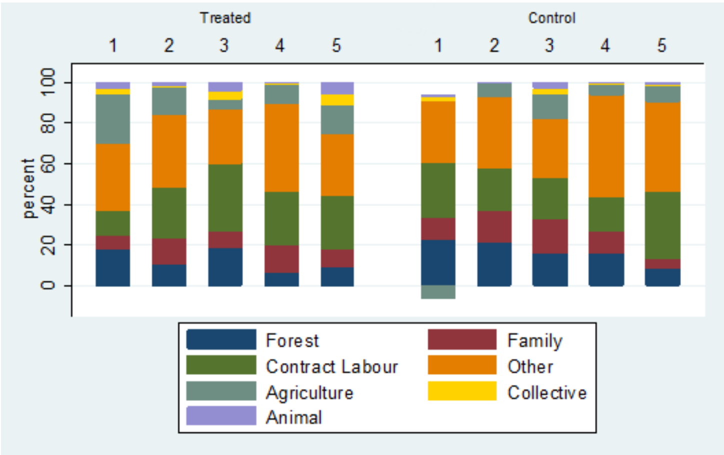
Income Categories	Description
Agriculture	Cereals (corn, rice, cassava (flour), tapioca flour, gum, potatoes, sweet potatoes); legumes (black beans, cowpea); vegetables (tomato, carrot, pepper, cucumber, garlic, lettuce, cabbage); fruits (mango, banana, acai, lemon, orange, watermelon, pineapple, coconut, soursop, coffee, cocoa); other crops (Tabaco, sugar cane, peanuts, maracuja)
Livestock	Including cattle, bull, calve, buffalo (adult, young), horse (adult, young), sheep, pork, duck, geese, turkey, fish, rooster, chicken, rabbit, bees Their products and services: milk, butter, cheese, egg, leather, manure, transport, honey, soap
Forest and Environment	Various wood varieties; non-wood products including flora for food, medicine, construction; and fauna for food, pets, sale; and different species of fish
Family	Including shop owner, vehicle mechanic, repair service , guide etc.
Contract Labour	Contract and salaried work from on and off-farm employment
Collective	Management of community forest, tourism, aviculture, Brazil nuts, fish, agriculture
Other Sources	Income from rent, insurance, maternity benefits, pension, inheritance, subsidies from BFF and other support programs

Cost Categories	Description
Agriculture	Input costs including seeds, manure, fertilizer, pesticides, tools, machinery, fuel, labour, transportation, land lease etc.
Livestock	Labour, materials, feed, rented material, production material etc.
Forest and Fish	Labour, transport
Family Business	Labour
Community work	Labour, material, transport and marketing

Source: HH Questionnaire and accompanying code book

Figure 7 demonstrates the low heterogeneity in mean per capita income source by quantile and reserve. HH generally obtain most of their income from 'Other sources' and 'Contract Labour'. Key differences between the two reserves include the negative percent of Agricultural profit in the poorest quantile of the control reserve. Moreover, as expected, the treated reserve has a larger proportion of income derived from 'Collective Sources' as it is directly attributable to treatment.

Figure 7: Per Capita Income by Income Source over Quantiles and by Reserve



Source: Own depiction

5.3 Treatment Variable

The treatment variable indicates that the community received the BF Income component at some point prior to survey. This decision to implement treatment was on the reserve level, where the RDS reserve implemented BF Income starting in 2007, while the APA had not at the time of survey. Treatment was applied on the community level, where the decision of which ICDP to pursue and how to allocate funds was communally made. For our purposes treatment is a binary variable applied to each HH and indicates if the community received the BF Income or not. As discussed in the theoretical framework, treatment can affect our dependent variable through changes in labour allocation, changes in purchasing of goods, social capital formation and economies of scale in production which allow for greater income source diversification.

5.4 Independent Variables

Twenty independent covariates were selected to match upon, including a range of binary and continuous variables as per the literature. They were selected as they meet the conditional independence assumption and are exogenous. For explanation purposes, covariates will be categorized by how they affect income in the same manner as in the theoretical framework. It should be noted that certain covariates can be used as proxies for multiple categories; however this is not an issue.[14] Refer to Appendix A.2 for further details concerning variable description.

i) Ability/ Productivity

In order to capture differences in HH ability and productivity, a range of variables relating to the HHH were used. These variables include *age, sex, level of education and weather they were born in interior*. Furthermore, the maximum level of HH education was considered. These variables are exogenous to treatment and have no issue of reverse causality with our dependent variables. They capture cultural norms, experience, connections within the community and environmental awareness which are used as proxies for productivity and ability.

ii) Social Capital

Moreover, variables which reflect social capital were accounted for. Although one could argue that social capital is affected by treatment, selected covariates are exogenous to treatment as they can be considered constant or changing at a constant rate over the course of treatment. Variables selected include *HH ethnic and religious majority status* and *distance inhabitants live from the community's centre.* These covariates capture ties within the community, business opportunities,

[14] In addition covariates were generated in a way which did not generate missing values or rather infinite values due to ratios with zero denominator values.

and development of business practices which are aligned with sustainable resource use regulations from the reserve.

iii) Endowment

Moreover, endowment in resources affects income generating possibilities. Consideration in choosing variables that remain exogenous to treatment was made. Covariates selected include *lagged asset value (2007), access to electricity, land owned* and *the HH participated in BFF.* Firstly, lagged asset value affects income possibility through physical capital, as well as capturing unobservable characteristics (business sense and ability to gain resources) that enabled prior HH asset accumulation. Secondly, access to electricity affects income through productivity and work hours available. Moreover, land ownership affects income through productive assets. By assumption, there is no issue of reverse causality since electricity requires time and significant resources to connect. Similarly, land can be considered constant throughout treatment as it requires significant saving and time to purchase. Lastly, BFF which strongly affects our dependant variables as it is a direct financial transfer was considered through two variables using both a both dummy variable for participation and a continuous one for the number of years the HH had participated. BFF can be considered exogenous as all HH in both reserves qualify.

iv) Economies of Scale within HH

HH earning capacity was considered through economies of scale within the HH. Weber at al. (2011) highlight the importance of HH labour constraints as it reflects the ability to invest time in alternative income generating possibilities. Since our dependent variable accounts for the number of members in the HH, other covariates including the *dependency ratio* and the *number of reported days sick were included.* These variables are exogenous to treatment and have no issue of reverse causality with our dependent variables.

v) Community Factors/ Access to Infrastructure

Ease of access to infrastructure is an important attribute to consider as it reflects forgone income. Variables which reflect the opportunity cost of travel include *distance to durables market* and *time to primary education and health facilities.* Although these variables may have changed over the course of implementation of BF Income, the change could be attributed to BF Social and Association which were equally applied to both treated and control communities. Moreover, the variables *environmental shocks and number of years since the community was founded* were included to proxy the community environment which affects HH welfare.

Results

This section will begin with a discussion of the issue of balance, as this is the primary testable means of measuring the effectiveness of matching. Given the covariates chosen based on theoretical grounds, various models will be tested (section 6.1.2). We are most interested in the ATT from the model which performs best in terms of balance, which will be presented in section 6.2. Further testing of the selected models will be conducted in the following robustness chapter. Analysis was conducted using both R and Stata.

6.1 Balance

6.1.1 Balance Before Matching

One measure of balance before matching is the pooled normalized difference in means adapted from Weber et al. (2011). This difference can be understood in terms of standard deviations, where a 0.1 standard deviation threshold is considered small. Referring to the final column of Table 2, as much as 16 out of our 20 covariates fall above this threshold and 12 lie above 0.2 standard deviations. The covariates relating land ownership and number of years the HH has participated in the BFF program appears to have the worst balance pre matching. In particular these differences in normalized mean provide evidence that using OLS on the full sample will provide biased results.

Table 2: Summary Statistics: Balance Before Matching

	Control		**Treat**		
	mean	SE	mean	SE	pooled difference
HHH sex	0.17	0.37	0.14	0.35	0.11
HHH age	47.34	16.82	46.61	15.47	0.06
HHH edu	3.97	3.48	5.06	3.97	0.42
max HH edu	6.93	3.58	8.36	3.73	0.55
dep ratio	0.36	0.27	0.33	0.26	0.17
days sick	13.43	29.77	11.58	21.98	0.10
interior	0.77	0.42	0.79	0.41	0.07
eth. majority	0.94	0.24	0.9	0.31	0.21
rel. majority	0.74	0.44	0.79	0.41	0.18
community year	24.23	8.67	30.06	15.48	0.67
env shock	0.09	0.28	0.28	0.45	0.72
health time	0.46	1.21	0.27	0.25	0.29
edu time	0.11	0.21	0.1	0.3	0.02
distance centre	1.25	2.36	0.87	1.99	0.25
distance mkt	59.09	24.9	43.5	24.46	0.85
asset 07	8.33	2.88	8.72	2.54	0.20
electricity	0.93	0.66	1.13	0.55	0.45
BFF	0.48	0.5	0.75	0.44	0.11
BFF year	1.14	1.31	2.8	1.86	0.77
land	22.8	52.44	27.53	70.12	1.34
Observations	127		87		

Source: Own calculation conducted in Stata

Binary Participation Models

Another measure of pre- matching balance can be determined from the binary regression model on the treatment dummy. The probit (from the normal distribution) and logit (logistic distribution) were used. The difference between these distributions is that the probit has relatively a fatter tail distribution, and thus is more likely to capture outliers (Caliendo & Kopeinig, 2008). For this reason the probit model will be considered in estimation. Coefficients and their signs were consistent between the two estimates for significant covariates (refer to Table 3).

Normally, the binary regression is used to develop a PS rather than for interpretive purposes, however from one perspective, significance indicates pre- match imbalance between covariates. From this study, significant variables include maximum HH education, total distance to market, as well the number of years the HH partook in BFF. In particular the sign and significance of these last two terms are surprising as it indicates that communities with a lower average time participating in BFF, and thus less contact with the BFP, are more likely to participate in BF Income treatment.

Table 3: Binary Regression Model on Treatment

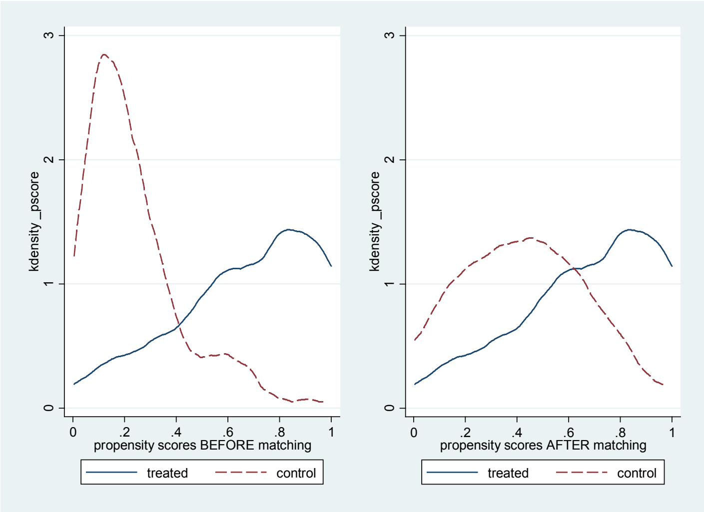
	Probit		**Logit**	
	Coef	SE	Coef	SE
HHH sex	-0.069	0.821	-0.194	0.732
HHH age	0.005	0.619	0.014	0.438
HHH edu	-0.011	0.783	0.007	0.921
max edu	0.087*	0.021	0.151*	0.023
dep ratio	0.088	0.871	0.288	0.766
days sick	-0.004	0.384	-0.006	0.518
interior	0.172	0.535	0.279	0.563
eth. maj	-0.154	0.712	-0.277	0.695
rel. maj	0.198	0.478	0.392	0.441
community year	0.034*	0.043	0.051	0.075
env shock	0.472	0.123	0.861	0.11
health time	0.05	0.754	0.053	0.847
edu time	0.675	0.155	0.756	0.393
distance centre	0.019	0.698	0.023	0.798
distance mkt	-0.016**	0.002	-0.030**	0.001
asset 07	-0.027	0.565	-0.027	0.749
electricity	0.297	0.132	0.428	0.219
BFF	-1.711***	0	-4.202***	0
BFF year	0.749***	0	1.658***	0
land	0.001	0.693	0.002	0.574
Constant	-1.873	0.058	-3.374	0.052
Observations	214		214	

Note: * p<0.05, ** p<0.01, *** p<0.001

Source: Own calculation conducted in Stata

These two indicators of pre match balance correspond to one another. Findings demonstrate that pre matching, control and treatment groups are quite different across several covariates. Furthermore, match imbalance can be illustrated graphically through plotting the density by propensity score (Figure 8). Here we see the pre matching PS distribution between the treated and control groups are quite different (left). Figure 8 (right) provides a preview of the potential gains in balance by performing matching.[15]

Figure 8: Propensity Score Distribution



Source: own depiction conducted in R and Stata
Note: Probit model, no caliper, with replacement

6.1.2 Balancing Tests Post Matching

This section will provide the results of various matching techniques in terms of balance. We are most interested in the ATT from the models that perform best in terms of balance. The matching techniques implemented include PS matching with both probit and logit specifications, with and without caliper, however only probit outcomes will be primarily reported. A variety of calipers were tested, where 0.1 SD were selected. This caliper balances the issue of loss of treatment observations with improving match quality. In addition, MD and the iterative Genetic matching algorithm were used. Furthermore, since there are no well-regarded unique and conclusive indicators of balance (Diamond & Sekhon, 2013), a variety of measures were used in comparing matching algorithms including:

a) Joint and unjoint difference in means
b) Standardized difference (pooled), and the percent bias reduction
c) eQQ mean difference
d) Variance ratio

[15] It should be noted that these distribution graphs are only possible for PSM as other techniques due not estimate a PS.

For a full set of results, refer to Appendix A.3. For the purposes of explanation, only visual interpretation and summary statistics will be presented in the text.

Number of Observations

Before looking at the post matching results, it is important to discuss how the number of observations changes under different approaches. The number of controls used is important as few controls may be important in developing the counterfactual and thus act as strong drivers of outcomes. Since our before matching balance is not very strong, one can suspect that relatively few observations (especially at the high PS range) are repeatedly used to match treatment individuals.

Considering Table 4, the weighted number of treatment observations is generally our treatment sample size of 87, unless a caliper was specified. In this case observations were dropped as no control was available within the specified bandwidth. The unweighted control accounts for multiple observations which have the same PS outcomes. The notion that few control observations drive outcomes is verified; where at most 50 different observations were used. Furthermore, under PS matching, single observations were used as much as 25 times or 29%[16] of the total required sample. MD and Genetic matching use greater number of unique controls and have a lower maximum, which are indicators of greater variance in developing the counterfactual.

Table 4: Control Observations under Various Matching Approaches

	Probit	Logit	Probit (caliper)	Logit (caliper)	MD	Gen Match
No. controls used	32	34	32	34	49	50
Max no. times single control used	25	27	17	22	10	5
Unweighted no. of Treated Obs	90	96	74	87	87	87
weighted no. of Treated Obs	87	87	71	72	87	87

Remark: Own Calculation conducted in R

a.i) F-test

Keeping differences in sample size in mind, various tests were conducted in order to determine post matching balance. Firstly, a test of joint equality of means (known as F-test or Hotelling test) across all covariates was conducted. The null hypothesis of joint mean equality for all models (including the comparison of pre matched controls with treatment) was not rejected as shown in Table 5. This indicates that in combination we cannot reject that covariate means are dissimilar. However, lack of observation independence for the F and t-tests remains an issue as matching with replacement was conducted (Lee, 2013). Thus, the results from these tests should not be overemphasised. Furthermore, although the F- test comparing pre matched treatment and control was not rejected,

[16] The single control observation is used as a match for 25 of the 87 treatment observations

this does not indicate lack of appropriateness of matching to further reduce bias as after all, our goal is to improve overall balance.

Table 5: P- value from F- test

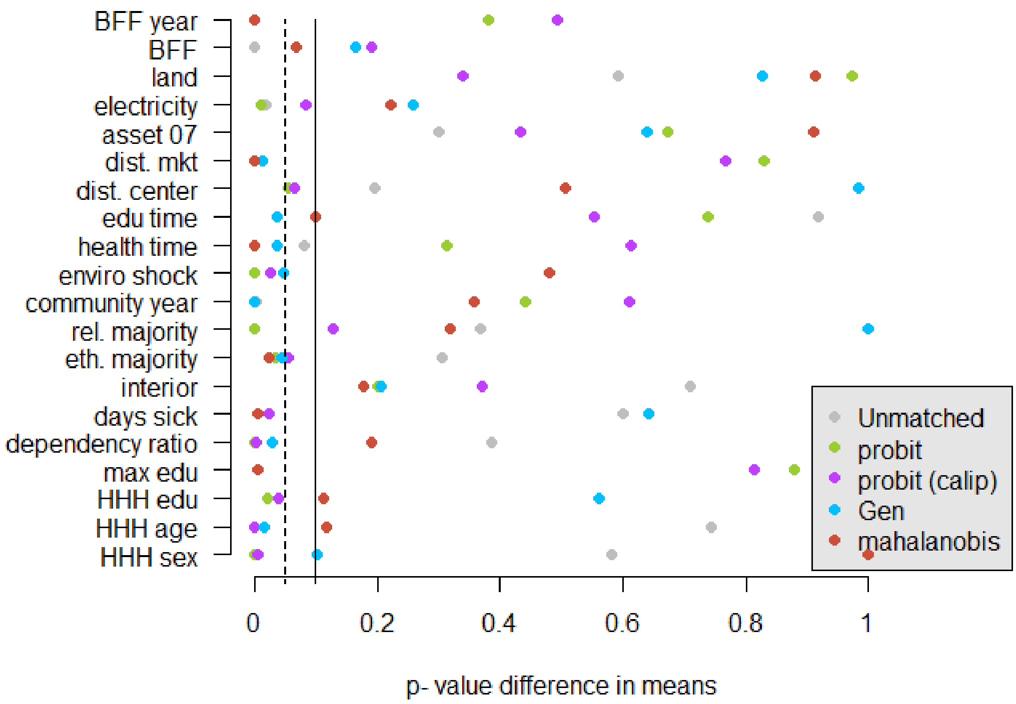
	before	probit	probit (caliper)	MD	Gen Match
p- value	0.8696	0.7295	0.7295	0.8942	0.516

Remark: Own Calculation conducted in R

a.ii) T-test: Difference of means

For each covariate a t-test was performed to determine difference in means, where a significant p-value is an indication of covariate imbalance. This metric demonstrates that matching has been somewhat unsuccessful as the number of significant variables in the Probit PSM approach (without calipers) has been made worse in comparison to the before matching base case. Use of calipers, as well as MD and Genetic matching appear to perform marginally stronger with fewer significant covariates (refer to Figure 9).

Figure 9: P- value of difference in Mean (T- test)



Source: own depiction

Remark: Significant variables which lie to the left of the 10% (solid line) and 5% (dashed line)

Table 6 provides summary statistics for this test across all 20 covariates. The count variables indicates the number of variables that fall within this range. In many cases, almost half of our covariates have remaining imbalance according to the difference in mean. As mentioned, not too much weight is put on these results due to lack of observational independence.

Table 6: P- value of difference in Mean (T- test) Summary Statistics

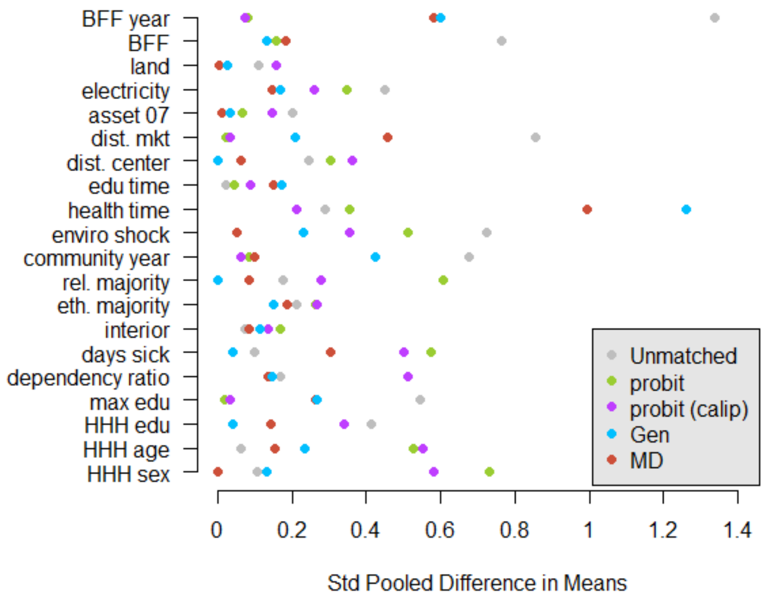
	Before Matching	Probit	Probit (caliper)	MD	Gen
count if < 0.1	9	10	9	7	10
mean	0.29	0.29	0.28	0.28	0.28
median	0.25	0.12	0.16	0.15	0.07
var	0.09	0.12	0.08	0.11	0.12

Remark: Own Calculation conducted in R

b) Standard Difference in Means (Pooled Variance)

The standardized difference in means between treatment and controls is a strong indicator of balance. Our results indicate that all matching specifications reduce absolute standard mean difference. As shown by Figure 10, a reduction in pooled difference in means under all matching approaches occurred. MD and Genetic matching algorithms perform best, with the most units lying proximate to zero. The latter is expected, as Genetic Matching explicitly seeks to maximize this metric of balance. Summary statistics for this metric, reported in Table 7 demonstrate MD and Genetic matching have the fewest variables with large remaining bias according to Rosenbaum's 20% threshold.

Figure 10: Absolute Standard Mean Difference (Pooled SD)



Source: own depiction

Table 7: Absolute Standard Mean Difference (Pooled SD) Summary Statistics

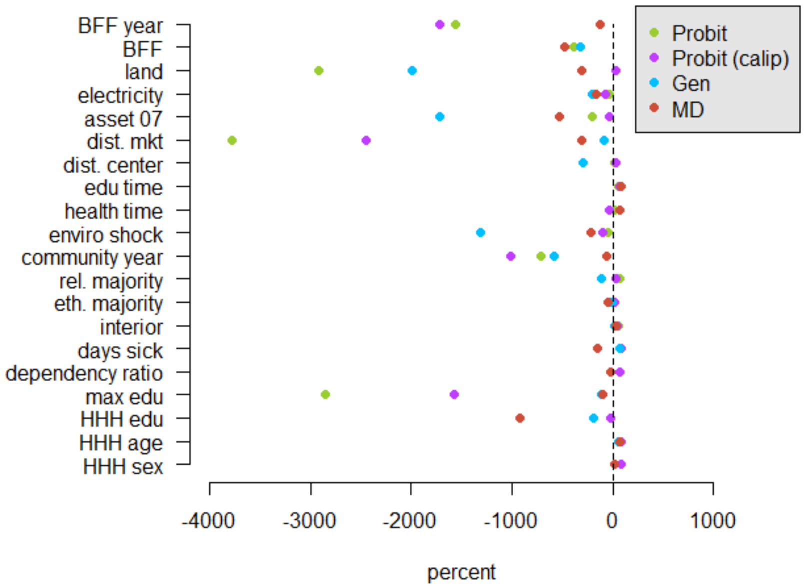
	Before Matching	Probit	Probit (caliper)	MD	Gen
mean	0.38	0.29	0.26	0.20	0.22
median	0.23	0.28	0.24	0.14	0.15
count if < 0.2	8	8	8	15	13

Remark: Own Calculation conducted in R

The standard mean difference can be manipulated in order to obtain the percent bias reduction. Results demonstrate that matching under all approaches reduces bias, which creates incentive for conducting this procedure (refer to Figure 11). Although bias reduction is quite substantial, one should be careful about simply minimizing mean bias as it can be driven by a few variables. For example, although the probit models has a large mean bias reduction, many covariates increased in bias relative to the no match situation. On the other hand, MD and genetic matching both reduced bias quite substantially and have fewer variables where balance deteriorated. When looking at the specific variables where imbalance remains (refer to Appendix A.3), although MD performs better in terms of number of balanced covariates, the key variable of market distance has much better balance in Genetic matching, with a pooled standardized mean difference of 0.21 compared to 0.46.

According to Table 8, MD and Genetic matching have the fewest increases in percent growth in absolute standard mean relative to the before matching case. Furthermore, they have achieved large overall bias reduction according to the mean and median. Once again, not too much weight should be placed on the mean, due to large percent bias reduction in specific covariates.

Figure 11: Percent Reduction in Bias (Pooled SD)



Source: own depiction

Table 8: Percent Change in Standard Mean Difference (Absolute/ Pooled SD) Summary Statistics

	Before Matching	Probit	Probit (caliper)	MD	Gen
count if < -20%	0	10	10	13	13
count if > 0%	0	10	10	5	5
count if > 20%	0	7	9	4	4
mean	0	-598	-339	-357	-1067
median	0	-2	-1	-108	-123

Remark: Own Calculation conducted in R

c) eQQ mean difference

This test represents overall model fit. On average, the only approach which improves the eQQ is MD, while Genetic matching also has a negative median eQQ . Furthermore, out of our 20 covariates, many, and in some cases a majority of covariates are made worse in terms of this measure of distribution equality. To a degree, this result can be attributed to reduction in sample size of the control. Summary statistics for this test are presented in Table 9.

Table 9: eQQ Summary Statistics

	Before matching	Probit	Probit (caliper)	MD	Gen
count if > 0	20	20	20	19	19
mean	2.08	3.26	2.71	1.97	1.99
median	0.49	0.48	0.51	0.34	0.42

Remark: Own Calculation conducted in R

d) Variance ratio

Lastly, the variance ratio is a metric which indicates the variance of the treatment relative to that off the control. Since variance reduction is not a goal in matching, results are not highly weighted in analysis. The results, demonstrate the variance ratio performs best under probit and Genetic matching specifications as specified in Table 10.

Table 10: Variance Ratio Summary Statistics

	Before matching	Probit	Probit (caliper)	MD	Gen
count if 1/2 < VAR < 2	5	10	6	5	5
mean	1.23	1.74	1.44	5.01	1.57
median	0.95	1.32	1.19	1.25	1.39

Remark: Own Calculation conducted in R

Overall, MD and Genetic matching appear to have had a relatively strong effect in terms of improving balance. Both metrics perform well in terms of number of unique observations in the matching procedure, with relatively low maximum number of times a single control was used. Moreover, both perform well in terms of standardized bias reduction. In terms of overall distribution, MD and Genetic matching had a mean difference eQQ close to zero. Thus, although bias is reduced, the difference in means and difference in standard bias tests indicate that there are still quite a few variables where imbalance remains, even under our preferred approaches. In some cases, matching made balance worse. Therefore, although we will now proceed with estimating the treatment effect, with specific interest in MD and Genetic Matching results, our findings should be interpreted with caution.

6.2 Summary Statistics for Dependant Variable

Before diving into results of our measure of additionality, we will briefly discuss summary statistics of our two dependant variables for our matched and unmatched sample over treatment status (refer to Table 11). We will ignore the dependant variable income2 for the moment which will be applied as a robustness test.

Starting with our asset growth dependant variable, the full sample mean of the two groups appear similar (1.24 for the control and 1.36 for the treated) with relatively high standard deviation. When controlling for observable characteristics, the mean for the control group reduces (for all models except probit with caliper) and the standard deviation decreases as well.

On the other hand, with our income dependant variable, considering the full sample, the mean of the two groups are similar (8.60 for the control and 8.66 for the treated) with relatively high standard deviation. Relative to this base case, post matching the mean income under all matching models except Genetic matching decreases the standard deviation. Thus, matching appears to alter the distribution of the treatment and control groups and generally lowers the variance relative to the full sample. The degree these distributional differences are significant will be tested with OLS and various matching approaches in the remainder of this chapter.

Table 11: Dependant Variable Summary Statistics by Treatment Status

	Full sample		**Probit**		**Probit (Caliper)**		**MD**		**Gen**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Control										
log diff assets	1.24	2.63	0.94	1.90	1.59	2.60	0.94	1.90	0.78	1.71
log per cap income	8.60	1.13	8.76	0.61	8.67	0.56	8.76	0.61	8.57	1.24
log per cap income2	8.62	1.13	8.78	0.61	8.72	0.58	8.78	0.61	8.59	1.24
Treat										
log diff assets	1.36	2.29	1.36	2.29	1.35	2.19	1.36	2.29	1.36	2.29
log per cap income	8.66	0.85	8.66	0.85	8.58	0.92	8.66	0.85	8.66	0.85
log per cap income2	8.68	0.85	8.68	0.85	8.60	0.91	8.68	0.85	8.68	0.85
Total										
log diff assets	1.29	2.49	1.15	2.11	1.47	2.40	1.15	2.11	1.07	2.04
log per cap income	8.63	1.02	8.71	0.74	8.63	0.76	8.71	0.74	8.62	1.06
log per cap income2	8.64	1.02	8.73	0.74	8.66	0.76	8.73	0.74	8.63	1.06

Source: Own calculation conducted in Stata

Remark: Refer to Table 4 for observation count

6.2 OLS (full sample)

As mentioned, OLS also relies on the assumption of conditional independence to matching (Bonin, 2014). However, OLS uses the full sample, while matching uses a subset of control observations and sometimes a subset of treatment observations if a calliper was specified. Since pre match imbalance appears to be an issue, results from full sample OLS should not be given too much consideration.

This regression analysis (Table 11) provides several interesting findings. Firstly, in terms of our measure of additionality, the treatment effect, the income dependant variable has a negative sign and is insignificant. Furthermore, there are several point estimates that have counterintuitive signs including maximum education, time to health facility and centre of community, HH assets in 2007 and participation in BFF. However, none of these variables are significantly different from zero. The significant variables include HHH age and the dependency ratio. The former can be interpreted as a one unit increase in HHH age causes a 1% increase in income.

In terms of our asset growth dependant variable, the treatment variable has a positive sign and is insignificant. Furthermore, HHH education and ethnic majority status have counterintuitive signs, although have insignificant p- values. As expected, lagged assets are a strong and significant predictor of asset change as lower asset values in 2007 create greater difference in asset value. Access to electricity is strong predictor of asset growth indicating a 33%[17] growth. Moreover, the maximum HH education has a positive effect of 7.5%.

Table 12: OLS Regression (full sample)

	Income		**Asset**	
	Coef	SE	Coef	SE
treatment	-0.26	0.18	0.27	0.15
HHH sex	0.06	0.19	-0.15	0.16
HHH age	0.013*	0.01	0.00	0.01
HHH edu	0.04	0.03	-0.01	0.02
max edu	-0.02	0.02	0.072***	0.02
dep ratio	-0.848*	0.33	-0.33	0.28
days sick	0.00	0.00	0.00	0.00
interior	0.21	0.18	0.06	0.15
eth. maj	0.13	0.26	-0.10	0.22
rel. maj	0.18	0.17	0.16	0.14
community year	0.00	0.01	0.00	0.01
env shock	-0.01	0.20	-0.04	0.17
health time	-0.06	0.08	-0.09	0.07
edu time	0.40	0.29	-0.03	0.25
distance centre	-0.01	0.03	0.04	0.03

[17] Calculated as the effect of dummy on *Y* is 100[*exp*(-*coef*) - 1] (Halvorsen & Palmquist, 1980). This method of calculating treatment effect will be used throughout the results section.

distance mkt	0.00	0.00	0.005*	0.00
asset 07	-0.01	0.03	-0.875***	0.02
electricity	0.07	0.12	0.283**	0.10
BFF	-0.29	0.30	-0.36	0.25
BFF year	0.17	0.09	0.12	0.08
land	0.00	0.00	0.00	0.00
Constant	7.882***	0.57	7.602***	0.48

Source: Own calculation

Note: p<0.05, ** p<0.01, *** p<0.001* p<0.05, ** p<0.01, *** p<0.001

6.3 Matching Results

Due to uneven balance between the treated and control group, the sign and significance of our ATT should be more accurate after performing matching. Recall we are most interested in the MD and Genetic matching models as they have demonstrate better balance across discussed metrics. Our ATT on the dependant variable per capita income is presented in Table 13. One can see that our ATT for the income dependant variable is insignificantly different across all matching algorithms. This can be interpreted as there is no significant effect of treatment when comparing the treated and matched control samples. Although the coefficient signs vary across matching algorithms, coefficients remain small and insignificant.

Table 13: Treatment Effect Income

	Probit	Probit (caliper)	MD	Gen Match
ATT	-0.01	-0.05	-0.10	0.05
SE[a]	0.24	0.19	0.16	0.14
p-value	0.96	0.80	0.53	0.73
N	87	71	87	87

Remark: Own Calculation conducted in R

[a] Abadie and Imbins Standard Error

In terms of the ATT of the dependant variable asset growth (presented in Table 14), our models demonstrate more significant results. The PSM models have negative coefficients and are insignificant. This can be in interpreted as there is no found significant change in assets growth between the treatment and control groups due to treatment. However, when applying MD, the ATT is significant at the 12% level indicating a 51% increase in asset growth due to treatment. The 10% significant level is commonly reported, thus MD falls slightly outside of this range. Genetic matching has both a larger estimate and is highly significant. In this case , treatment causes asset value to increase by 65%. It should be noted that the standard error is the so called Abadie and Imbens robust standard error.

Table 14: Treatment Effect Asset Growth

	Probit	Probit (caliper)	MD	Gen Match
ATT	-0.08	-0.23	0.41	0.50
SE[a]	0.85	0.64	0.26	0.16
p-value	0.92	0.72	0.12	0.00
N	87	71	87	87

Remark: Own Calculation conducted in R

[a] Abadie and Imbins Standard Error

6.4 OLS (matched sample)

Furthermore, post matching OLS was conducted which further controls for observable covariates and reduces standard error (refer to Table 15). The sample size of the control group is reduced to correspond with that of the treatment. For the income dependant variable the coefficient size of treatment gets larger (more negative). Under this specification, market distance has a counterintuitive sign and is significant. This implies further travel to market increases income. On the other hand, when considering our asset growth, the p- values of treatment increases making both models insignificant at the standard levels. This can be interpreted as no significant effect of treatment on asset growth. For our asset growth dependant variable, significant coefficients include maximum HH education, lagged asset value and access to electricity which corresponds to our full sample OLS results.

Table 15: Post Matching OLS

Gen Matching					MD			
	Income		**Asset**		**Income**		**Asset**	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
treatment	-0.23	0.15	0.17	0.16	-0.23	0.15	0.17	0.16
HHH sex	-0.04	0.17	-0.28	0.19	-0.04	0.17	-0.28	0.19
HHH age	0.01*	0.01	0	0.01	0.01*	0.01	0	0.01
HHH edu	0	0.02	0	0.02	0	0.02	0	0.02
max edu	-0.02	0.02	0.06*	0.02	-0.02	0.02	0.06*	0.02
dep ratio	-0.42	0.29	-0.42	0.32	-0.42	0.29	-0.42	0.32
days sick	0	0	0	0	0	0	0	0
interior	0.01	0.16	0.13	0.17	0.01	0.16	0.13	0.17
eth. maj	0.08	0.21	-0.32	0.24	0.08	0.21	-0.32	0.24
rel. maj	0.35*	0.15	0.14	0.17	0.35*	0.15	0.14	0.17
community-year	0	0	-0.01	0.01	0	0	-0.01	0.01
env shock	-0.08	0.14	-0.19	0.15	-0.08	0.14	-0.19	0.15
health time	0.26	0.34	-0.08	0.38	0.26	0.34	-0.08	0.38
edu time	0.41	0.24	0.14	0.27	0.41	0.24	0.14	0.27
distance centre	-0.01	0.03	0.03	0.03	-0.01	0.03	0.03	0.03
mkt distance	0.01*	0	0	0	0.01*	0	0	0
asset 07	0.04	0.03	-0.85***	0.03	0.04	0.03	-0.85***	0.03
electricity	-0.13	0.11	0.33**	0.12	-0.13	0.11	0.33**	0.12
land	0	0	0	0	0	0	0	0
BFF	-0.15	0.25	-0.45	0.28	-0.15	0.25	-0.45	0.28
BFF year	0.11	0.07	0.14	0.08	0.11	0.07	0.14	0.08
Constant	7.30***	0.52	8.32***	0.58	7.30***	0.52	8.32***	0.58
Observations	174		174		174		174	

Source: Own Calculation

Note: p<0.05, ** p<0.01, *** p<0.001

Robustness

7.1 Issues of Bias

As demonstrated from the balancing tests, although matching reduced standardized bias for observable covariates, it does not do this uniformly across all variables. Furthermore, imbalance remains even after matching. This discussion relates back to equation 4) and 5) of ATT. Our estimate will be biased if $E[Y(0)|X, D = 1] - E[Y(0)|X, D = 0] \neq 0$, where the first term is unobservable. For this reason, we can only speculate the direction of bias. The first sub section of this chapter will discuss the direction of remaining covariate imbalance in terms of creating bias in our results. Table 15 summarizes results of the discussion. The second section discusses issues of bias from unobservable variables. In combination, this helps illustrate if our ATT results are overestimated or underestimated.

7.1.1 Bias from Observed Variables

This discussion will include variables where a significant different remain post matching based on the t-test results and standardized differences. These results are reported in Table A.3 of the appendix. Furthermore a synopsis of the results from this section is reported in Table 16.

i) <u>Ability/ Productivity</u>

Innate ability or productivity was accounted for by four variables (HHH age, sex, HHH education, HHH born in the interior and maximum HH education). Age and maximum HH education had remaining imbalance and thus could promote bias. Firstly, the variable maximum HH education is positively correlated with income. Since our control group has a lower mean than the treated group, our ATT is overstated. Secondly, HHH age is also positively correlated with income. Since our control group has a lower mean than the treated group, our ATT may be overstated. Thus, ability or productivity is likely overstated.

ii) <u>Social Capital</u>

HH majority ethnic status within the community is a variable that captures social capital a factor also accounted for by 3 other covariates (religious status, distance from the community centre, and years since the community was founded). Being a part of the ethnic majority implies gains in social capital which is positively correlated with income. Post matching imbalance show that the control group has a significantly higher mean relative to than the treated group, thus income of the control is higher than it should be and our ATT could be understated.

iii) Endowment

Endowment is measured by lagged assets, electricity, land and BFF (2 covariates). BFF has an effect on income as this direct financial transfer allows HH to invest in productive assets, training and education, and enhances the range of available decisions for a HH. The variable where imbalance remains is the number of years the family has been participating in BFF. This variable does not have a strong effect on effect on current income; rather it is highly important for our asset growth dependant variable. Early or longer BFF participation could allow for savings and pooling of resources in order to obtain assets. Our results demonstrate that the control group has participated in BFF for significantly fewer years relative to the treated group. Thus, asset growth in particular in the control is understated, which results in overstating our ATT based on the endowment factor.[18]

v) Community Factors/ Access to Infrastructure

Readers should be reminded that this group indicates variables that were calculated on the community level and extrapolated down to the HH one for comparative purposes. Thus, these variables have fewer observations and less variation, making balance more difficult to achieve. There are several variables where post matching imbalance remains for community factors. Firstly, travel time reflects ease of access to institutions. Greater travel time imposes a cost (negatively correlated with income) HH reflecting forgone wage. Furthermore, time to health facilities in the control group has a significantly higher mean relative to the treated group. Thus, the control group's outcome is understated and our ATT is likely overstated. Similar logic can be applied with the market distance variable. Since the control group has a further travel distance relative to the treated, the control's opportunity cost of performing market transactions is higher than the treated group. Thus, income is understated and our ATT results are likely overstated.

Lastly, the number of years the community has existed can be categorized as an indicator of unobservable factors which relate to how a community functions. Older communities are expected to have a higher average income. Our results demonstrate that the control group has lower mean number of years than the treatment group, resulting in income being understated in the control group. This implies that our ATT potentially overstated. Therefore all these imbalances in observable community variables point toward these overestimating our ATT.

[18] Note, there was no factor in iv) where imbalance remained, therefore it will not be discussed

Table 16: Direction of Bias due to Observable Covariate Imbalance

Category Number	Variable with Unequal Mean post matching	Sign of theoretical correlation with Income	Observed Mean Control < than mean Treat	Direction of Bias of ATT
i)	HHH age	(+)	Yes	Overstated
i)	Max HH Education	(+)	Yes	Overstated
ii)	Ethnic Majority	(+)	No	Understated
iii)	Years participating in BFF	(+)	Yes	Overstated
v)	Time to health facilities	(-)	No	Overstated
v)	Market Distance	(-)	No	Overstated
v)	Year Community Exist	(+)	Yes	Overstated

Source: own depiction

As indicated in the prior discussion and summarized in Table 16, the overall bias resulting from remaining imbalance demonstrates that our treatment effect is likely overstated. This implies a decrease in significance in ATT.

7.1.2 Bias from Unobserved Variables

As illustrated in the empirical approach, Rosenbaum's bounds test enables us to test for the degree unobservable bias affects the ATT using an empirical measure of robustness. This test is particularly relevant for the asset growth which was found to be significant in terms ATT. Firstly, the Wilcoxen test determines the degree of sensitivity significant (upper bound) and insignificant (lower bound) results have to hidden bias. The upper bound is generally the more important indicator, where confirmed by our data, insignificant ATT from matching resulted in an upper bound of in Γ=1.[19] Genetic matching on the assets dependant variable is most interesting as it was found to be highly significant. In terms of Table 17, at the 5% significance level, Γ= 1.60. This can be interpreted as the following: if the odds of an individual receiving treatment increase by 60% due to differences in unobserved covariates (holding matched covariates constant), assets will no longer be significant. This indicates some robustness to unobserved characteristics. Conversely, the MD model is much less robust to unobservable variable bias with a 10% change in treatment status causing our inference to change. This low robustness to unobservable factors is unsurprising given that MD has a p-value ATT for asset growth of 0.12.

Table 17: Robustness to Bias Wilcoxen Signed Rank Test:
Value of Γ when significant at the 5% level

	Probit	Probit (Caliper)	MD	Gen
Income	(1.5)	(1.42)	(1.22)	(1.35)
Assets	(1.33)	(1.38)	1.10	1.60

Remark: Lower Bounds are in parentheses, while Upper Bounds are not
Own Calculation conducted in R

[19] Theoretically at Γ=1, the matched p-value should be replicated. Although our results do not demonstrate this precisely, significance does not change. Keele notes, this issue could which could be attributed to outliers (Keele, 2010).

Another test for unobservable selection bias is the Hodges-Lehmann test which considers the additive effect from treatment. It can roughly be interpreted as the difference in medians between treatment and control groups. Once again, Γ =1, indicates the variable had an insignificant ATT from matching, as indicated for all our approaches using the Income dependant variables. Referring to Table 17, considering the asset growth dependent variable under the MD and Genetic matching specifications, both of these become more robust relative to the Wilcoxen test. The interpretation is similar to above. For Genetic matching, one can conclude doubling the likelihood of treatment caused by an unobservable characteristic which also affects asset accumulation (while holding observed covariates constant) would have *no* effect on the significance of our dependant variable.

Table 18: Robustness to Bias Hodges-Lehmann test:
Critical value of Γ where confidence interval includes zero

	Probit	Probit (Caliper)	MD	Gen
Income	1.001	1.001	1.25	1.001
Assets	1.122	1.116	1.31	2.21

Remark: Own Calculation conducted in R

Overall, the dependant variable of asset growth seems to be reasonably robust to hidden bias under Genetic matching, and to a much lower degree under MD. Robustness further increases when outliers are controlled for in the Hodges-Lehmann test. Several unobservable variables that could cause bias could be savings and entrepreneurship ability (Duvendack & Palmer-Jones, 2012).

7.2 Alternative Dependant Variable

Furthermore, using an alternative dependant variable allows us to test for robustness of results. We considered a slightly different depiction of income, known as income2 with summary statistics in Table 11. This dependant variable removes the cost factor of animal production which was conducted due to issues in the questionnaire structure. Running our various matching algorithms on this dependant variable demonstrates a little change in significance of our ATT.[20] Under this dependant variable, our income ATT remains insignificantly different from zero with p- values of 0.52 and 0.75 for MD and Genetic matching respectively (refer to Appendix A.4 for details). Thus, there is additional evidence that our income is insignificantly affected by treatment.

[20] We expected a reduction in significance which can be explained as by the following. As we remove this cost component, total income increases. Therefore, BF Income as a percent of total income will fall ceteris paribus. Thus, gains from treatment will appear smaller reducing significance of our ATT.

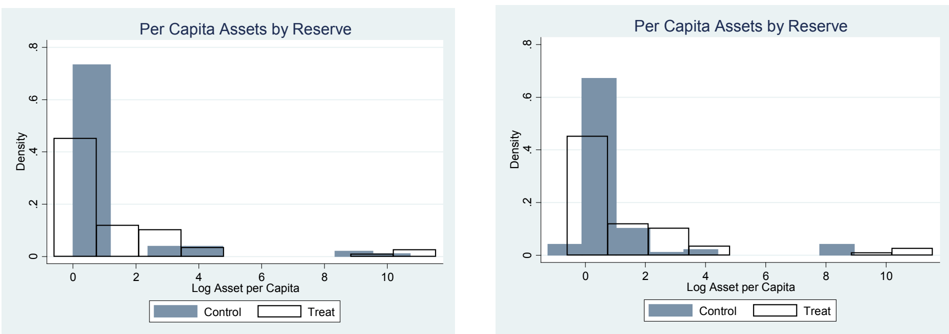
7.3 Outliers in the Asset Growth Dependant Variable

Although it is not advisable to drop outliers' ex-post, for robustness purposes this practice was tested for our asset growth dependant variable in order to determine if the significant ATT under Genetic matching was driven solely by outliers. Positive, unmatched outliers in the dependant variable for the treatment group may indicate that our ATT is overestimated. This hypothesis was developed due to the strong significance of our lagged asset covariate in pre and post matching OLS. Furthermore, visually, there appears to be a group of outliers on the right of our asset growth distribution (refer to Figure 13). Upon further inspection, the largest change in assets can be attributed to poor HH recollection of lagged assets, where the 15 observations that reported zero asset value in 2007, had some of the largest asset value growth.

Two tests were conducted by dropping different numbers of observations. Firstly, the single largest treatment observation was dropped. As expected, dropping this outlier (which reduces our treatment sample size) reduces the significance of our treatment on our asset growth dependant variable to the 15% significance level for MD, while with Genetic matching it remains significant at the 1% level (refer to Appendix A.5 for full results). A second test was conducted, dropping four treatment outliers, which lie to the right of 5 log unit difference in asset value (refer to Appendix A.6 for results). As expected significance of our ATT decreases. Under MD our ATT becomes insignificant with p-value of 0.33. Genetic matching still performs well, providing a highly significant treatment p-value of 0.01. Under Genetic matching the ATT estimate remains constant at a 65% increase in asset growth. Thus, it appears that although outliers in the treatment variable affect our results, ATT significance remains when controlling for poor recollection under Genetic matching. This provides some evidence that our significant ATT is not entirely driven by outliers.

Figure 12: Post Matching Distribution of Assets

MD Gen



Source: own depiction conducted in Stata

Discussion of Results

8.1 Findings Within the Conceptual Model

The ATT of the ICDP treatment on human welfare was calculated using matching techniques, where we are most interested in results from MD and Genetic matching as they perform best in terms of balance. Our results indicate an insignificant ATT on per capita income. In terms asset growth, MD finds no significant treatment effect, while Genetic matching finds a large and significant effect of 65%. Furthermore, upon running OLS with the matched sample, treatment significance disappears which demonstrates instability of results. Although significant ATT on asset growth appears to be robust in terms of unobservable variable bias for Genetic matching, post matching covariate imbalance provides evidence that the significance of our results are overstated. In relation to the literature, Weber et al. find that assets are significantly affected by ICDP treatment, although Bauch et al. (2014) find the contrary. Furthermore, Weber et al. (2011) find treatment causes a significant change in cash income and Bauch et al. (2014) find a positive effect on HH income.

Narrative

Taking the results of Genetic matching seriously, a narrative of significant asset growth and insignificant income attributed to treatment was developed relating back to our theoretical framework. Although we are unable to control for which path treatment affects outcome, there exists a plausible narrative of our witnessed results in relation to the 'changes in demand' factor. This narrative relates to how the dependant variables are defined. Since income is calculated net production costs (including investment in productive assets), HH revenue gained from the ICDP may be reinvested into the production process. This would produce no change in income, but positively affect asset accumulation. Reinvested ICDP gains in productive durable goods appears to be a plausible start-up strategy, as new income generating activities may require investment in tools, machinery etc., especially in the short run. This narrative can be criticised on two fronts. Firstly, that our results are driven by outliers in the treatment group, making the ATT overestimated and biased. Secondly that low HH education and risk aversion would inhibit this action, especially over the short run. The next part of this sub section will describe potential causes and issue of outliers in our asset growth variable.

Outliers in Assets

If asset growth is driven by outliers in the treatment group, then the ATT may not a valid indicator of treatment effect. Furthermore, if large asset accumulation were driven by 'non- productive investments' (which would not be counted as a cost for income generation), then our narrative does not hold. There is some evidence to support this idea as outliers in terms of asset growth corresponded to HH which purchased highly valuable assets such as accommodation.

In terms of outliers in asset growth, the treatment group appears to have greater quantity and larger outliers than the control groups. This indicates that our ATT is overstated. There are two drivers of outliers of asset growth. Firstly, they can be attributed to poor data quality. This can be demonstrated by the finding that all HH that reported an asset growth rate of greater than 5 log units also reported zero asset value in 2007. Secondly, large changes in asset accumulation could be driven by debt (negative asset) or savings (positive asset). The former does not appear to be a strong issue as the outliers HH in asset accumulation are not the same as those that have the largest net debt at the time of survey. On the other hands, savings could be considered as a driver for asset accumulation. This is plausible given the large proportion of HH income which falls in the 'Other' category including lump sum gains in the form of rent, social insurance benefits, inheritance, subsidies etc. It is difficult to analyse savings in the same manner as debt, as knowing if a HH had savings which they spent on asset accumulation, or never had savings to begin with is problematic with only current period savings data. This brief discussion demonstrates that there appears to be outliers which are unrelated to productive asset accumulation which provides evidence against our plausible narrative. Moreover, outliers created by poor recollection and savings may cause overestimation of results and bias in our ATT.

In an attempt to control for these outliers, several (both 1 and 4) were dropped. Results show the significance of the ATT for MD increases to 0.33, while using Genetic matching asset growth remains highly significant with a p-value of 0.01 and a consistent asset growth rate of 65%. This indicates that outliers are not the only cause of our significant ATT and rather there are in fact distributional differences between the two groups which could be attributed to treatment. The remainder of this sub chapter will discuss bias caused by observable and unobservable variables which biases our results.

Observable Characteristics

The next step in understanding the distributional differences in asset growth between the two groups was to consider remaining imbalance across observable covariates. Almost unanimously, our robustness results point toward an overestimation of our treatment effect, which reduces significance of our ATT. However from these observable characteristics, there is one key variable that warrants further discussion as it may be an important driver of differences in HH asset accumulation between the two groups. Specifically, this refers imbalance in the variable 'years participating in BFF', which captures HH endowment. One may recall that this variable refers to the number of years a HH has received a monthly cash transfer through BFF. It is very likely that since the treatment group participated in BFF for longer duration, asset accumulation would be greater as HH would have had a longer period of relatively higher income to purchase durables. This narrative provides evidence that our significant results for the asset growth variable are biased. Difference in asset growth could be not only attributed to treatment, but rather a result of this important confounding factor.

Unobservable Characteristics

With this key difference in observable variables in mind, findings that asset growth is fairly robust to unobservable variables that may cause changes in treatment and outcome under Genetic matching are not surprising. For Genetic matching, the Wilcoxen test shows that unobservable characteristics would require a 60% change in likelihood to participate to change our inference. The Hodges-Lehmann was even more robust requiring a 121% increase in participation to change inference. There are two candidates for these unobservable factors including lagged savings and entrepreneurial abilities. In terms of the former past savings could be a strong indicator of treatment and asset accumulation. In terms of the latter, ability would affect the motivation to participate and thus altering expected and realized gains from ICDP participation.

Overall, there are many causes of concern, which reduce confidence in the significance of our results under Genetic matching. Therefore extreme caution when interpreting the strong significance found in the ATT should be given. All other parametric and non- parametric techniques for our ATT with our asset growth dependant variable provided insignificant results at standard levels. Furthermore, outliers, unobservable characteristics and remaining imbalance in key observable covariates demonstrate an overestimation of our ATT. Moreover, our income dependant variable is subject to the same array of issues in terms of observable variable imbalance and unobservable variable bias pointing to overestimation of our ATT significance. However, it should be noted the narrative of years participating in BFF does not hold the same weight with income as it did for asset growth. Overall, in combination these tests points toward a relatively robust finding that per capita income is insignificantly affected by the ICDP program at this point in time.

8.2 Potential Reasons for Insignificant Effect

Although, the BF Income seeks to promote access income generating opportunities and positive financial returns for HH, no significant effect of the program is expected for several reasons. The most general argument is that at the time of survey, only a relatively short period (maximum of 4 years) had elapsed since the ICDP inception. Since ICDP are complicated, and require changing livelihood patterns, a short time horizon may be insufficient to witness an effect. This point can be given further detail. Firstly, at the time of survey we are unsure as to how far into actual ICDP implementation the process had evolved within each community. Moreover, there is heterogeneity in this process between treated communities, as each received independent in funding allocation and had autonomous decision making of how to best implement the ICDP.

This time frame issue is heightened by HH risk aversion which acts as a strong barrier to implementation of new practices. Risk aversion and low participation can be thought of as a function of the HH preferences, education, and wealth. Other roadblocks may include financial, legal, political and institutional restrictions. Thus, even after HH gain access to a new income generating activity, there are many barriers to actual implementation. These can be mitigated through training, education and time to experiment with new practices in light of their own constraints.

8.3 Limitations

General Limitations

There are a few common limitations which arise when conducting with- without experiments. These limitations may influence our results, and thus will be discussed briefly. Firstly, spill- over can be considered as the occurrence where treated HH affect non-treated ones and treatment effect is not isolated to the treatment community (Khandker *et al.*, 2010). This could be an issue if there were trade between reserves. If this were a problem it would make the difference between the two groups smaller, reducing our treatment effect (Bauch *et al.*, 2014). By assumption, this is not an issue as due to difficulties to travel between the two reserves. Another issue would involve HH migration to communities in order to gain from treatment. This would blur the delineation between treatment and control, and allow for self-selection into treatment. Once again, by assumption this not an issue as there is a two year prohibition on BFP participation, which is a strong incentive to prevent migration for solely program gains. However, indirect gains from other members of the community receiving the program could be achieved.

Treatment Variable

There are several limitations to this study with regard to the treatment variable. Firstly, an issue with treatment definition is that treatment was not homogenous in terms of value, type and time of implementation. One could speculate that higher value investment, and longer periods of implementation would lead to greater HH participation, and greater profitability. Furthermore, different ICDP investments may have different growth rates in participation and profitability.

Moreover, a heterogeneous treatment effect could arise even when holding treatment type constant. This could be attributed to HH investing different amounts of resources in the ICDP due to differences in opportunity costs. Furthermore, as Gebara illustrates, certain HH may benefit directly from the ICDP implementation, while others only gain indirectly from greater attention from monitoring and evaluation officers and tourism (Gebara, 2013; Bauch *et al.*, 2014). Moreover, perhaps greater monitoring may constrain illegal income generating activities which would otherwise have gone unnoticed. Furthermore, due to the communal decision making structure in the ICDP development, powerful members of the community may have greater influence in deciding which specific ICDP was implemented. Thus, community investment may directly benefit a select few community members which may have vested interests due to own preferences, experience, resources, connections etc.

It could be argued that defining treatment on the community level does not accurately reflect actual HH participation in BF Income. This issue was taken into account by Bauch et al. (2014) who considered treatment on both the community level and he HH level across different points in time. One alternative treatment variable could be percent of HH participating in the treated community which would reflect intensity of participation. A community level definition has the benefit of capturing all indirect costs and benefits through spill overs between community members (Bauch *et al.*, 2014). These can include changes in demand for certain goods due to changes in production practices, as well as indirect benefits and costs. Although a HH level definition of treatment may provide results of direct gain from treatment, a lack of correct designation from responses creates obscurity and biases our ATT results. This can be illustrated by an example where a community's ICDP promotes Brazil Nuts, and the HH reported partaking in this activity. If a HH attributes Brazil nut gain to their 'Agricultural activity' rather than 'Collective Income', a HH level of treatment may prove unreliable. A community level definition helps to mitigate this issue, as treatment status is independent of participation. Overall, although there are certain benefits for using a community level definition of treatment, using multiple definitions of treatment is recommended. In our case, the reason only a community level definition was chosen was due data constraints.

Conclusions and Recommendations

There are various approaches used to tackle environmental and human wellbeing in protected areas although there is no clear best approach. One strategy is an indirect ICDP approach which seeks to use the synergies between rural livelihoods in valuable natural environments to develop 'win-win' outcomes of environmental sustainability and gains in HH welfare.

The impact evaluation literature on management strategies largely focuses on environmental goals. Conversely, this thesis measures effectiveness in terms of the often secondary goals of human wellbeing. This thesis contributes to the small body of literature which uses rigorous program evaluation techniques to measure the additionality of ICDP implemented in protected areas with a focus on human wellbeing. Real world quantitative evaluation is necessary as it has strong ties to program and policy development, funding allocation and mitigating environmental damage and rural poverty. However, the contextual nature of case study based program evaluation results in low external validity and extrapolation of results into broader conclusions (Börner *et al.*, 2013).

The additionality of the ICDP treatment was evaluated using a case study of two reserves from the BFP. It was measured as the ATT relative to the counterfactual control reserve over the short run. The dependant variables used to measure human wellbeing included log income and log difference in assets or asset growth.

The primary issue that this paper seeks to address was that of selection bias, whereby characteristics simultaneously determine both participation and outcome. This is an important issue which plagues program evaluation. Given available non- randomized data, regression analysis and matching approaches were conducted in order to determine the ATT. Both approaches control for relevant observable factors which can be categorized as affecting productivity, social capital, endowment, economies of scale within HH and community factors. These factors control for selection bias under the assumption that these are the only relevant characteristics in determining both program selection and outcome.

Our balancing criteria in non- parametric matching demonstrate MD and Genetic matching perform best. In this regard, both selected matching algorithms and post matching OLS demonstrate no significant effect of ICDP treatment on the income dependant variable. On the other hand, MD and post matching OLS find no significance of treatment on our asset growth variable. However, Genetic matching found a highly significant ATT of 65% which was fairly robust to unobservable variables.

Taking the results of Genetic matching seriously, there is a plausible narrative for realizing this dual result which consists of HH reinvesting program gains in productive assets. Although this may seem like an intuitive and viable strategy for a start-up organization which seeks to grow and improve efficiency, perhaps it is overstating local sophistication especially when considering low levels of education and risk aversion. Other potential drivers of the significant ATT were inspected including outlier driven results, which could be attributed to poor HH recollection or omitted variables in the form of savings and debt. Removing outliers did not alter the significant effect on asset growth under Genetic matching. This indicates overall distributional differences between the treated and control communities in terms of asset growth.

Furthermore, issues of imbalance across key observable characteristics such as 'time participating in BFF' remain, which could be a strong driver of these differences in asset growth. Specifically, longer duration with additional income supported by BFF would allow HH greater opportunities to accumulate assets. This finding, in addition to imbalance amongst other observable covariates provides evidence that our significant asset growth result was overstated. Overall, insignificant results on the growth in assets dependant variable under MD and post matching OLS, in addition to potential unobservable variable bias due to covariates such as savings and entrepreneurial skill, and remaining imbalance in key observable characteristics indicate that one should be cautious when interpreting significance of our results.

In general, the ICDP strategy has been criticised for low effectiveness and inconclusive results. Although our results are no different, one should recall the context the BFP is applied. The BFP targets reserves which already have relatively low environmental pressure and levels of poverty. If the greatest immediate change to these goals were the sole criteria for program development, perhaps the BFP is misplaced. Rather, the BFP holds a long term approach which seeks to mitigate future environmental threats by developing local conservation allies. This long duration fits well with the ICDP strategy.

Over the short run expecting significant results is optimistic as even under strong incentives livelihoods take time to adjust. Firstly, relatively intangible gains in the form of participant empowerment and social capital formation (with positive externalities) may not affect measurable HH wellbeing over the short run. Furthermore, if the ICDP had not been successfully or fully implemented in the community at the time of survey, indicators of success may even decrease as HH may allocate scarce resources (physical, time, financial etc.) toward participation. As the project matures, and efficiency in production and marketing improve, one would expect project profitability to increase. Moreover, over a longer time span, HH risk aversion and preferences may change, resulting in greater direct gains from program participation.

Although the ICDP appears to have had an insignificant effect, this does not necessarily indicate its overall ineffectiveness. Rather, the process requires more time to bear fruit. Further evaluation at a later date where the program has had full potential to implemented and adopted by HH would be recommended. This is especially important as finding no significant effect over any time horizon implies program funds simply act as a subsidy for those living in the reserve. After all, the goal of the ICDP is to develop local capacity and build both environmentally sustainable and profitable income generating sources which are viable without indefinite external support.

Further Research

There are several recommendations for further research which would be possible with a richer dataset. Firstly, our community level definition of treatment is quite broad; although it has the benefit of accounting for spill- overs within treated communities. Another definition based on direct HH participation may prove to be interesting. Moreover, it would be interesting to study a heterogeneous treatment effect. This is plausible due to power dynamics within a community, which shape ICDP rollout favouring powerful community members. Furthermore, treatment designation in our study was considered based on receiving the community ICDP program irrespective of program type and amount allocated. Controlling for these factors would provide a better understanding of marginal returns to support. Lastly, post matching imbalance in key observable characteristics remained an issue in this study. Conducting exact matching on these covariates would reduce this source of bias.

Reference List

Agustsson, K., Garibjana, A., Rojas, E. & Vatn, A. (2014). An Assessment of the Forest Allowance Programme in the Juma Sustainable Development Reserve in Brazil. *International Forestry Review*, 16(1), pp 87–102.

Alkire, S. & Foster, J. (2011). Understandings and misunderstandings of multidimensional poverty measurement. *The Journal of Economic Inequality*, 9(2), pp 289–314.

Andam, K. S., Ferraro, P. J., Sims, K. R. E., Healy, A. & Holland, M. B. (2010). Protected areas reduced poverty in Costa Rica and Thailand. *Proceedings of the National Academy of Sciences of the United States of America*, 107(22), pp 9996–10001.

Arnold, C., Conway, T. & Greenslade, M. (2011). Cash transfers literature review. *London: Department for International Development*,.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, 10(2), pp 150–161.

Bauch, S. C., Sills, E. O. & Pattanayak, S. K. (2014). Have we managed to integrate conservation and development? ICDP impacts in the Brazilian Amazon. *World Development*, 64, pp S135–S148.

Becker, S. O. & Caliendo, M. (2007). Mhbounds-sensitivity analysis for average treatment effects.

Blom, B., Sunderland, T. & Murdiyarso, D. (2010). Getting REDD to work locally: lessons learned from integrated conservation and development projects. *Environmental Science & Policy*, 13(2), pp 164–172.

Bonin, D. (2014). Treatment Effects in Econometrics. *MIT Economics* [online],. Available from: https://www.google.se/search?q=Bonin+Treatment+Effects+in+econometrics+MIT&hl=sv&authuser=0.

Butler, R. A. & Laurance, W. F. (2008). New strategies for conserving tropical forests. *Trends in Ecology & Evolution*, 23(9), pp 469–472.

Börner, J., Wunder, S., Reimer, F., Bakkegaard, R. K., Viana, V., Tezza, J., Pinto, T., Lima, L. & Marostica, S. (2013). Promoting Forest Stewardship in the Bolsa Floresta Programme: Local Livelihood Strategies and Preliminary Impacts. *Rio de Janeiro, Brazil: Center for International Forestry Research (CIFOR). Manaus, Brazil: Fundação Amazonas Sustentável (FAS). Bonn, Germany: Zentrum für Entwicklungsforschung (ZEF), University of Bonn*,.

Caliendo, M. & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), pp 31–72.

Davies, T. E., Fazey, I. R. A., Cresswell, W. & Pettorelli, N. (2014). Missing the trees for the wood: Why we are failing to see success in pro-poor conservation. *Animal Conservation*, 17(4), pp 303–312.

Diamond, A. & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, 95(3), pp 932–945.

Duvendack, M. & Palmer-Jones, R. (2012). High noon for microfinance impact evaluations: re-investigating the evidence from Bangladesh. *The Journal of Development Studies*, 48(12), pp 1864–1880.

FAS (2013). *Results of a baseline study for a REDD+ pilot area in Brazil: the Rio Negro APA and the Bolsa Floresta Programme* [online]. London: IIED.

Ferraro, P. J. (2004). *Direct Payments to Protect Endangered Ecosystems and Experimental Methods to Estimate Payment Costs*. International Development Research Centre.

Ferraro, P. J. & Hanauer, M. M. (2014). Quantifying causal mechanisms to determine how protected areas affect poverty through changes in ecosystem services and infrastructure. *Proceedings of the National Academy of Sciences*, 111(11), pp 4332–4337.

Ferraro, P. J., Lawlor, K., Mullan, K. L. & Pattanayak, S. K. (2011). Forest Figures: Ecosystem Services Valuation and Policy Evaluation in Developing Countries. *Review of Environmental Economics and Policy*, p rer019.

Ferraro, P. J. & Miranda, J. J. (2014). The performance of non-experimental designs in the evaluation of environmental programs: A design-replication study using a large-scale randomized experiment as a benchmark. *Journal of Economic Behavior & Organization*, 107, Part A, pp 344–365.

García-Amado, L. R., Pérez, M. R. & García, S. B. (2013). Motivation for conservation: Assessing integrated conservation and development projects and payments for environmental services in La Sepultura Biosphere Reserve, Chiapas, Mexico. *Ecological Economics*, 89, pp 92–100.

Gebara, M. F. (2013). Importance of local participation in achieving equity in benefit-sharing mechanisms for REDD+: a case study from the Juma Sustainable Development Reserve. *International Journal of the Commons*, 7(2), pp 473–497.

Grieg-Gran, M., Porras, I. & Wunder, S. (2005). How can market mechanisms for forest environmental services help the poor? Preliminary lessons from Latin America. *World development*, 33(9), pp 1511–1527.

Halvorsen, R. & Palmquist, R. (1980). The interpretation of dummy variables in semilogarithmic equations. *American economic review*, 70(3), pp 474–75.

Heinrich, C., Maffioli, A. & Vazquez, G. (2010). *A primer for applying propensity-score matching*. Inter-American Development Bank.

Ho, D. E., Imai, K., King, G. & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3), pp 199–236.

Honey-Roses, J., Baylis, K. & Ramirez, M. I. (2011). A spatially explicit estimate of avoided forest loss. *Conservation Biology*, 25(5), pp 1032–1043.

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.

Keele, L. (2010). An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data. *White Paper. Columbus, OH*, pp 1–15.

Khandker, S. R., Koolwal, G. B. & Samad, H. A. (2010). *Handbook on impact evaluation: quantitative methods and practices*. World Bank Publications.

King, G., Nielsen, R., Coberley, C., Pope, J. & Wells, A. *Comparative Effectiveness of Matching Methods for Causal Inference*. (2011).

Landell-Mills, N., Porras, I. T. & others (2002). *Silver bullet or fools' gold?: a global review of markets for forest environmental services and their impact on the poor*. International Institute for Environment and Development London.

Lee, W.-S. (2013). Propensity score matching and variations on the balancing test. *Empirical economics*, 44(1), pp 47–80.

Lima, P. G. B. (2014). Strengthening Livelihood Flows on Payment for Environmental Services through Local Lenses: Evidences from the Bolsa Floresta Programme. *Journal of Sustainable Development Studies*, 7(1).

Luiza, L., Marostica, S. & Grieg-Gran, M. (2013). Local preferences for REDD+ payment.

Lunt, M. (2014). Selecting an Appropriate Caliper Can Be Essential for Achieving Good Balance With Propensity Score Matching. *American Journal of Epidemiology*, 179(2), pp 226–235.

McShane, T. O., Hirsch, P. D., Trung, T. C., Songorwa, A. N., Kinzig, A., Monteferri, B., Mutekanga, D., Thang, H. V., Dammert, J. L., Pulgar-Vidal, M., Welch-Devine, M., Peter Brosius, J., Coppolillo, P. & O'Connor, S. (2011). Hard choices: Making trade-offs between biodiversity conservation and human well-being. *Biological Conservation*, 144(3), pp 966–972 (The New Conservation Debate: Beyond Parks vs. People).

Miteva, D. A., Pattanayak, S. K. & Ferraro, P. J. (2012). Evaluation of biodiversity policy instruments: what works and what doesn't? *Oxford Review of Economic Policy*, 28(1), pp 69–92.

Narayan, D. & Pritchett, L. (1999). Cents and Sociability: Household Income and Social Capital in Rural Tanzania. *Economic Development and Cultural Change*, 47(4), pp 871–897.

Naughton-Treves, L., Holland, M. B. & Brandon, K. (2005). The Role of Protected Areas in Conserving Biodiversity and Sustaining Local Livelihoods. *Annual Review of Environment and Resources*, 30(1), pp 219–252.

Nelson, A. & Chomitz, K. M. (2011). Effectiveness of Strict vs. Multiple Use Protected Areas in Reducing Tropical Forest Fires: A Global Analysis Using Matching Methods. *PLoS ONE*, 6(8), p e22722.

Newton, P., Nichols, E. S., Endo, W. & Peres, C. A. (2012). Consequences of actor level livelihood heterogeneity for additionality in a tropical forest payment for environmental services programme with an undifferentiated reward structure. *Global Environmental Change*, 22(1), pp 127–136.

Nolte, C., Agrawal, A., Silvius, K. M. & Soares-Filho, B. S. (2013). Governance regime and location influence avoided deforestation success of protected areas in the Brazilian Amazon. *Proceedings of the National Academy of Sciences*, 110(13), pp 4956–4961.

Perloff, J. (2008). *Microeconomics*. 5th. ed Addison Wesley.

Persson, U. M. & Alpízar, F. (2013). Conditional cash transfers and payments for environmental services—a conceptual framework for explaining and judging differences in outcomes. *World Development*, 43, pp 124–137.

Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin III, F. S., Lambin, E., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J. & others (2009). Planetary Boundaries: Exploring the Safe Operating Space for Humanity.

Rosenbaum, P. R. & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), pp 33–38.

Salafsky, N. (2011). Integrating development with conservation: A means to a conservation end, or a mean end to conservation? *Biological Conservation*, 144(3), pp 973–978 (The New Conservation Debate: Beyond Parks vs. People).

Sekhon, J. S. (2008). Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *Journal of Statistical Software, Forthcoming*,.

Shadish, W. R., Clark, M. & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, 103(484), pp 1334–1344.

Shah, A. *Poverty Facts and Stats*. [online] (2013-01-07) (Global Issues). Available from: http://www.globalissues.org/article/26/poverty-facts-and-stats. [Accessed 2014-12-24].

Sills, E. O., Atmadja, S. S., de Sassi, C., Duchelle, A. E., Kweka, D. L., Resosudarmo, I. A. P. & Sunderlin, W. D. (2014). *REDD+ on the ground: A case book of subnational initiatives across the globe*. CIFOR.

Sims, K. R. E. (2010). Conservation and development: Evidence from Thai protected areas. *Journal of Environmental Economics and Management*, 60(2), pp 94–114.

Smith, J. (2000). *A critical survey of empirical methods for evaluating active labor market policies*. Research Report, Department of Economics, University of Western Ontario.

Smith, J. & Todd, P. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1–2), pp 305–353 (Experimental and non-experimental evaluation of economic policy and models).

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science : a review journal of the Institute of Mathematical Statistics*, 25(1), pp 1–21.

Sunderland, T. C., Sayer, J. & Hoang, M.-H. (2012). *Evidence-based conservation: lessons from the Lower Mekong*. Routledge.

Viana, V. M. (2008). Bolsa Floresta: um instrumento inovador para a promoção da saúde em comunidades tradicionais na Amazônia. *Estudos Avançados*, 22, pp 143–153.

Weber, J. G., Sills, E. O., Bauch, S. & Pattanayak, S. K. (2011). Do ICDPs Work? An Empirical Evaluation of Forest-Based Microenterprises in the Brazilian Amazon. *Land Economics*, 87(4), pp 661–681.

Winkler, R. (2007). *Why do ICDPs fail? The relationship between subsistence farming, poaching and eco- tourism in wildlife and habitat conservation* [online]. CER-ETH - Center of Economic Research (CER-ETH) at ETH Zurich. (07/76).

Wunder, S. (2013). When payments for environmental services will work for conservation. *Conservation Letters*, 6(4), pp 230–237.

Wunder, S., Engel, S. & Pagiola, S. (2008). Taking stock: A comparative analysis of payments for environmental services programs in developed and developing countries. *Ecological economics*, 65(4), pp 834–852.

Appendix

A.1

Quasi- experimental approaches commonly use two data designs; before and after the program/ treatment and with/ without. A combination of the two designs is also possible. A before and after design assumes a constant trend over time, whereby no external factors that affects outcome arises. On the other hand, a with- without design assumes that expected outcomes from the treatment and control groups in absence of implementation are the same.

The quasi- experimental approaches explained in Table 19 have certain assumptions which if are not met produce biased results. Providing motivation how these assumptions are met is necessary (Miteva *et al.*, 2012). It should be noted, several of these techniques can be used in conjunction.

Table 19: Description of Quasi- Experimental Approaches

	Difference in Differences(DID)	**Instrumental Variable (IV)**	**Regression Discontinuity Design (RDD)**
Description	DID removes unobservable differences by assuming they are time invariant. This assumption can be a problem if changes are a function of initiation conditions that influence participation (Bauch *et al.*, 2014). A much less robust approach is the single difference which uses the with- without design. This can be referred to as naive analysis, as it assumes no selection bias.	IV seeks to break the correlation between the error term and treatment. It uses a variable correlated with treatment, but which does not affect outcome directly. Thus, the only path the IV affects outcome is through treatment. The causal effect is estimated "by measuring how the outcome varies with the portion of the total variation in the treatment explained by variation in the instrumental variable" (Miteva *et al.*, 2012, p 73). One issue regarding this method is finding a strong and valid IV; one that is correlated with treatment and exogenous to outcome.	RDD follows a different approach to the other quasi experimental designs. Instead of making assumptions to develop a counterfactual, it rather assumes randomization over a small region. RDD requires a characteristic with a cut-off point that determines treatment status. Over a small region, one can assume no difference in covariates *X*, thus the only difference in outcome can be attributed to treatment (Khandker *et al.*, 2010).

A.2

Table 20: Variable Description

Variable	Type	Units	Survey	Description
Total pc income	Continuous	R$	HH	Total HH per capita income
Diff asset value	Continuous	R$	HH	Difference in asset value between 2013 and 2007
treatment	Dummy	1= Treated	Community	Participating in BF Income
HHH sex	Dummy	0= male	HH	Sex of HHH
HHH age	Continuous	years	HH	Age of HHH
HHH edu	Continuous	Years	HH	Education of HHH
Max edu	Continuous	Years	HH	Maximum education of HH members
Dependancy ratio	Continuous	Children/ HH member	HH	Dependency Ratio
Adult sick	Continuous	Days	HH	Number of days working aged members (>18) were sick
interior	Dummy	1= interior	HH	Born in Interior
Eth. majority	Dummy	1 = majority	HH	HHH or spouse belonging to ethnic majority
Rel. majority	Dummy	1= majority	HH	HHH belonging to religious majority
Community year	Continuous	Years	Community	Years since BFF implemented
Env. shock	Dummy	1= Shocked	Community	At least 2 environmental shocks occurred in the last year
Health time[a]	Continuous	Hours	Community	Time to health facility
Edu time[a]	Continuous	Hours	Community	Time to education facility
Distance centre	Continuous	Km	HH	Distance from house to centre of community
Distance Mkt[c]	Continuous	Hours	HH and Community	Total distance to market
Asset 07	Continuous	R$	HH	Total value of lagged assets
electricity	Dummy	1= Yes	HH	Access to electricity either by wired connection, or through owning a generator or solar power
BFF	Dummy	1= Yes	HH	HH participating in BFF
BFF year	Continuous	Years	HH	Years since family had participated in BFF
land	Continuous	Ha	HH	Amount of land owned

[a] Time was considered zero if the facility was located within the community

[c] Market distance was calculated as the sum of distance from the HH to the community centre and the distance from the community centre to large market

A.3

Table 21: Covariate Balance before and after Treatment

Covariate	**Mean**				**p-value (mean)**			**pooled Difference**			**Var Ratio**			**eQQ**		
	Treated un matched	Ctrl un matched	*Ctrl MD*	*Ctrl Gen*	*un matched*	*MD*	*Gen*	*un matched*	*MD*	*Gen*	*un matched*	*MD*	*Gen*	*un matched*	*MD*	*Gen*
HHH sex	0.14	0.17	0.14	0.09	0.58	1.00	0.10	0.11	0.00	0.13	0.87	1.00	1.42	0.02	0.00	0.05
HHH age	46.61	47.34	44.21	43.01	0.74	0.12	0.02	0.06	0.16	0.23	0.85	0.92	1.15	1.91	2.89	3.64
HHH edu	5.06	3.97	4.49	4.90	0.04	0.11	0.56	0.42	0.14	0.04	1.30	1.81	2.02	1.14	0.82	0.92
max edu	8.36	6.93	7.38	7.36	0.01	0.01	0.01	0.55	0.26	0.27	1.09	1.76	1.87	1.48	1.07	1.18
dep ratio	0.33	0.36	0.36	0.37	0.39	0.19	0.03	0.17	0.14	0.15	0.96	1.15	1.04	0.03	0.05	0.04
days sick	11.58	13.43	4.90	12.44	0.60	0.00	0.64	0.10	0.30	0.04	0.55	3.32	1.26	3.47	6.67	3.06
interior	0.79	0.77	0.83	0.84	0.71	0.18	0.21	0.07	0.08	0.11	0.94	1.15	1.22	0.02	0.03	0.05
eth maj	0.90	0.94	0.95	0.94	0.31	0.02	0.04	0.21	0.19	0.15	1.58	2.11	1.71	0.03	0.06	0.05
rel maj	0.79	0.74	0.83	0.79	0.37	0.32	1.00	0.18	0.08	0.00	0.86	1.15	1.00	0.06	0.03	0.00
Community year	30.06	24.23	28.54	23.48	0.00	0.36	0.00	0.67	0.10	0.42	3.19	6.97	2.76	5.97	6.39	6.71
Env shock	0.28	0.09	0.25	0.17	0.00	0.48	0.05	0.72	0.05	0.23	2.53	1.06	1.40	0.20	0.02	0.10
health time	0.27	0.46	0.02	0.58	0.08	0.00	0.04	0.29	1.00	1.26	0.04	65.86	0.03	0.50	0.25	0.60
edu time	0.10	0.11	0.06	0.05	0.92	0.10	0.04	0.02	0.15	0.17	2.09	3.43	3.76	0.07	0.05	0.06
Distance centre	0.87	1.25	0.99	0.87	0.20	0.51	0.98	0.25	0.06	0.00	0.71	1.35	1.05	0.48	0.44	0.25
Mkt distance	43.50	59.09	54.65	48.56	0.00	0.00	0.01	0.85	0.46	0.21	0.97	1.05	1.62	15.28	11.81	10.55
Asset 07	8.72	8.33	8.75	8.80	0.30	0.91	0.64	0.20	0.01	0.03	0.78	1.48	1.37	0.58	0.60	0.66
electricity	1.13	0.93	1.05	1.03	0.02	0.22	0.26	0.45	0.15	0.17	0.69	0.99	0.69	0.20	0.08	0.12
BFF	27.53	22.80	27.90	25.63	0.59	0.91	0.83	0.11	0.01	0.03	1.79	0.84	3.06	8.16	6.95	10.53
BFF year	0.75	0.48	0.67	0.69	0.00	0.07	0.16	0.77	0.18	0.13	0.76	0.85	0.88	0.28	0.08	0.06
land	2.80	1.14	1.72	1.69	0.00	0.00	0.00	1.34	0.58	0.60	2.01	1.89	1.98	1.68	1.08	1.12

Source: Own Calculation conducted in R

A.4

Table 22: Treatment effect with Income2 Dependant Variable

	PS probit	Probit (caliper)	MD	Gen Match
Estimate	-0.09	-0.08	-0.10	-0.05
SE[a]	0.25	0.19	0.16	0.14
T-stat	-0.35	-0.40	-0.64	-0.32
p-value	0.72	0.69	0.52	0.75

Source: Own Calculation conducted in R

[a] Abadie and Imbens standard error

A.5

Table 23: Treatment effect after dropping 1 asset outlier

	Probit	Probit (caliper)	MD	Gen Match
Estimate	0.11	0.00	0.40	0.79
SE	0.69	0.51	0.27	0.26
T-stat	0.16	0.01	1.49	3.06
p-value	0.87	0.99	0.14	0.00

Source: Own Calculation conducted in R

[a] Abadie and Imbens standard error

A.6

Table 24: Treatment effect after dropping 4 asset outliers

	Probit	Probit (caliper)	MD	Gen Match
Estimate	0.30	0.30	0.24	0.50
SE	0.34	0.24	0.24	0.19
T-stat	0.87	1.22	0.98	2.73
p-value	0.38	0.22	0.33	0.01

Source: Own Calculation conducted in R

[a] Abadie and Imbens standard error