

UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche  
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE  
CICLO XXIII

# THE SOCIAL AND GEOGRAPHICAL ISOLATION OF HOUSEHOLDS IN RURAL AFRICAN AREAS

**Direttore della Scuola:** Ch.ma Prof. MONICA CHIOGNA

**Supervisore:** Ch.mo Prof. ERICH BATTISTIN

**Co-supervisore:** Ch.mo Prof. ENRICO RETTORE

**Dottoranda:** SLAVICA ZEC

31/01/2012

# Riassunto

Il tradizionale approccio della valutazione delle politiche si concentra principalmente su un particolare parametro: in genere il *treatment status*, mentre una minore attenzione è stata volta verso i meccanismi attraverso i quali opera la politica in questione. Lo scopo principale di questa tesi è quello di acquisire di conoscenza dei meccanismi causali attraverso i quali una politica per lo sviluppo produce un effetto. Utilizzo il contesto delle piccole economie dell’Africa rurale, attraverso i due *case study* relativi ai progetti della Banca Mondiale (WB) nel campo dello sviluppo economico in Nigeria e in Gambia.

Nella prima parte della tesi si considera un progetto WB della costruzione delle strade e si propone una strategia per la derivazione del *marginal treatment effect* come un ingrediente della l’interconnessione tra la costruzione delle strade, l’accesso ai mercati locali e il benessere delle famiglie nelle piccole economie rurali. La mia ricerca parte dalla letteratura che crea i ponti tra *policy evaluation* e approccio strutturale alla Heckman (2010) e fornisce una sua applicazione nel contesto di economia dello sviluppo, dove la classica teoria economica spesso non riesce a spiegare il comportamento degli individui. Dopo aver derivato il *marginal treatment effect* dell’accesso al mercato al benessere delle famiglie, costruisco un *setting* per eseguire una valutazione *ex-ante* della componente infrastrutturale del programma.

Nella seconda parte della tesi, considero il contesto di *The Gambia Driven Development Program* per esplorare i meccanismi della formazione endogena delle reti sociali, definite dalle dimensioni dello scambio economico informale tra le famiglie nei villaggi. Gli scambi che consideriamo sono scambi tra le famiglie in termini di lavoro, credito, terra e mezzi di produzione. L’obiettivo della ricerca è duplice: condurre una ricerca empirica sfruttando i network sovrapposti dello scambio economico, cercando di ottenere una maggiore comprensione del ruolo del programma nella formazione delle reti sociali. L’obiettivo principale è quello di distinguere l’effetto del programma che opera indirettamente sulla probabilità di formare un link, attraverso le modifiche delle preferenze individuali verso lo scambio economico.



# Abstract

The traditional policy evaluation approach mainly focuses on one specific policy parameter, typically the treatment status, while less attention has been given to the mechanisms through which development program operates. On the other hand, in recent econometric literature there have been some significant developments in untangling the mechanisms of economic agents behavior, they shed light on preferences and technology parameters that are invariant to policy changes. The core of the thesis is to unlock the policy parameters in a unique World Bank programs in the challenging context of rural African economies, through acquiring knowledge about the causal mechanisms driving the change induced by the economic development policy. In the first part of the thesis, using a very unique setting of the World Bank infrastructure development program, I propose the strategy for the construction of building blocks that can be used as the ingredients of a structural system that explains the interconnection between the roads construction, access to markets and households' well being. I exploit the results from the innovative structural literature and apply them in under-developed economies setting, where standard economic theory often fails to explain the agents's behavior. The goal is to understand better the underlying developing economy principles, and to perform an *ex-ante* evaluation of the road construction on complex mechanisms of market exchange and welfare production function. This is interesting from both methodological and empirical point of view, as there are very few studies that apply newly developed structural econometric tools on real data. Moreover, in contrast to the carefully designed evaluations of many health and education initiatives in developing countries, our knowledge about the impacts of infrastructure projects has remained under-developed

In the second part of this thesis, I use the setting of The Gambia Community Driven Development Program, to explore the endogenous social network formation, that defines the dimensions of the informal economic exchange. The objective of the research is twofold: conduct an empirical research on a very rich dataset on social networks in rural Gambia, using the setting of the Community Driven Program Evaluation (CDDP) and focus on the endogenous link formation. The second goal is to disentangle the effect of the program that operates indirectly

on the probability of forming a link, through modifying the existing network attributes or modifying exchange preferences of individuals; thus find convincing way to identify the program effect on the distribution of individual and dyad fixed effects (preferences).

# Acknowledgements

I can never thank enough my supervisor, Prof. Erich Battistin, for all of his scientific support, generosity and understanding. He is the best motivator that I have ever known and has always been a pleasure to work with.

I am grateful to Prof. Enrico Rettore for all of precious advice and always illuminating point of view.

I would like to thank Prof. Jean Louis Arcand and Dany Jaimovich for giving me the amazing opportunity to work together on The Gambia project.

Many thanks to Prof. Alessandra Salvan and Prof. Monica Chiogna and all the lecturers from the first year - I have learned a lot, and I wish I could have learned even more. Thanks to all the administrative staff from the Statistics Department, they have all always been very efficient and reliable.

I would also like to kindly thank IRVAPP, our director Prof. Antonio Schizzerotto and all of my colleagues for being very accomodating and have generously met my needs in many circumstances in these last few months. Many thanks to my dear colleagues from XXIII cycle Marco and Anthony and especially my dear friend Nicola, it was great pleasure to study and work together. Thanks to other colleagues from different PhD cycles that I had the pleasure to meet and divide office space.

A special thanks to my talented and unique friend Michele De Nadai. Thanks for...just everything.

And finally, I wish to thank to my family and all of my dear ones who had to support me when I was unbearable and have believed in me - even when I did not.



# Indice

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Overview . . . . .	7
1.2	Main contributions of the thesis . . . . .	11
<b>2</b>	<b>The causal foundations of structural modeling and policy evaluation</b>	<b>12</b>
2.1	What are the goals of causal inference? . . . . .	12
2.2	Causal inference and policy effects . . . . .	13
2.3	Causal inference and structural parameters . . . . .	14
2.4	Notation and basic definitions . . . . .	15
2.5	Parameters of interest . . . . .	15
2.6	Why empirical research in development economics? . . . . .	17
<b>3</b>	<b>Rural roads access, markets and welfare</b>	<b>18</b>
3.1	Literature review . . . . .	18
3.1.1	Poverty and access to roads . . . . .	19
3.1.2	Access to markets . . . . .	20
3.2	Problem setting . . . . .	21
3.3	Estimation method . . . . .	23
3.4	Nigeria Commercial Agriculture Development Program . . . . .	24
3.4.1	Program description . . . . .	24
3.4.2	Data . . . . .	25
3.5	Empirical results . . . . .	26
3.5.1	Isolation and welfare . . . . .	27
3.5.2	Isolation and access to market - propensity score estimation . . . . .	28
3.5.3	Isolation and access to market - MTE estimation . . . . .	29
3.6	Future research . . . . .	32



<b>4</b>	<b>Social Isolation and Community Driven Development Programs</b>	<b>34</b>
4.1	Literature review . . . . .	34
4.2	A First Specification . . . . .	35
4.2.1	Basic definitions . . . . .	35
4.2.2	Specification at Household level . . . . .	35
4.2.3	Specification at Dyad level . . . . .	36
4.2.4	Dyads and basic inference . . . . .	36
4.2.5	One useful descriptive tool: Oaxaca-Blinder Decomposition . . . . .	37
4.3	Model specification . . . . .	37
4.3.1	Factor analysis . . . . .	37
4.3.2	Identification . . . . .	38
4.4	Model specification in our case . . . . .	38
4.4.1	Identification . . . . .	39
4.5	Likelihood specification . . . . .	40
4.6	Data and results . . . . .	41
4.7	Conclusions and Future Work . . . . .	41
	<b>Appendice A Nigeria Commercial Agriculture Development Program</b>	<b>42</b>
	<b>Appendice B Gambia Community Driven Development Program</b>	<b>45</b>

# Capitolo 1

## Introduction

### 1.1 Overview

The Development Impact Evaluation Initiative is a World Bank (WB) program that aims at generating knowledge on the effectiveness of different small and large scale development policies. So far, there has been a substantial attention upon rigorous impact evaluation using micro-data and exploiting widely accepted econometric tools through the traditional counterfactual approach. It resulted in improving the quality of WB operations through iterative learning and less risky scaling-up of the projects.

However, the traditional approach mainly focuses on one specific policy parameter, typically the treatment status, while less attention has been given to the mechanisms through which development program operates. The wide use of natural experiments in less developed countries (LDC) has allowed to obtain the estimates that are not invariant to changes in the characteristics of the policy itself and it is a ‘black box’ that rarely has the external validity in different environments.

On the other hand, in recent econometric literature there have been some significant developments in untangling the mechanisms of economic agents behavior, they shed light on preferences (or alternatively, on production technology parameters) that are invariant to policy changes. This so-called *structural approach* relies explicitly on the well defined underlying economic model and it offers tools that are used to simulate or predict changes in behavior, so that the results can be extrapolated to different policy environments. Todd and Wolpin (2003) draw a clear line between the definition of policy parameters and structural parameters using the setting of cognitive achievement production function. More interestingly, in Todd and Wolpin (2008) it is shown the structural parameters can be used in an *ex-ante* policy evaluation in widely know development project in Mexico (PROGRESA).

This thesis is an attempt to explore how structural reasoning can be applied in two typical

development policy evaluation settings. My focus mainly lies in the degree of isolation of rural households as a mechanism through which an external shock, such as WB development project, can modify preferences towards the economic exchange on one side and the cost ingredient of the agricultural production function, on the other side.

The first part of my thesis takes into consideration a geographical isolation of rural households in terms of their access to facilities, in particular access to markets. To address this question, I use the context of Nigeria Commercial Agriculture Development Project - a five year WB initiative that aims at strengthening agricultural production systems and targets small and medium commercial farms through capacity building, information diffusion and infrastructure building. The most interesting component of the project is the infrastructure component, that allows for impact evaluation through randomized phasing-in of roads and power plants building. I follow the pioneering work of Jacoby (2000) studies the relationship between road access to the households' nearest market and the value of land, using rural Nepal data. Since then, there were many attempts to address the question of the impact of road construction and market access on welfare (Van der Walle (2002) , Gonzales-Navarro and Quintana-Domeque (2010)). Overall, the road construction literature argues that access to markets boosts income diversification, raises technology adoption, lowers the cost of production and raises the competitiveness of the households closer to the markets.

To explore this issues, I start from the assumption that the road construction reduces the cost of reaching a market, as a locus of affairs, increasing its accessibility and raising the intensity of market exchange. To overcome the endogeneity of the households' choice to access the market place and identify its causal effect on agricultural revenues, I use the real GIS calculated distance from the households to the markets as an exclusion restriction for market access. This way, I construct a set-up of a structural model, that allows to identify the parameters of agricultural production function, that would mimic the impact of the newly constructed roads on the welfare outcomes, through improving access to market.

The second part of this thesis builds upon the recent empirical literature regarding social networks in development context. These suggest that one of the likely mechanisms through which social programs affect welfare is through breaking and establishing new social equilibrium. There were two main streams: one highlights the social processes behind risk sharing arrangements and another deals with the models of information diffusion. Recent empirical studies such as Fafchamps and Gubert (2007) and Dercon and Weerdt (2006) investigate whether the risk sharing methods are formed in order to maximize the mutual gains from pooling income risk. Krishnan and Sciubba (2009) examine the role of network structure in determining the impact of labor sharing networks on outcomes. In terms of social networks as the key factor in the transmission of information for technological innovations in agriculture, significant contributions have been provided by Conley and Udry (2008) and Bandiera and Rasul (2006), while Miguel and Kremer (2003) consider the use of innovative drugs in network-diffusion context. Here I consider the survey designed for the evaluation of The Gambia Community Driven Development Program (CDDP). It is a five year joint WB funded project with the government of The Gambia that is aimed at empowering and financing the communities to take lead in the identification and implementation of their own development projects. It is a randomized experiment designed survey, and very rich network data is collected both in treated and non-treated areas.

The starting point of my research is the endogenous formation of different economic networks, that typically overlap in rural economies. I build upon the pioneering work of Udry and Conley (2004), followed by Arcand and Fafchamps (2008) who exploit the properties of dyadic model framework to underline the importance in household similarities to engage in social interaction in terms of economic exchange.

The main idea is to integrate the latent component into the dyadic framework in order to account for the individual preferences over choices of social interaction. Data on different economic overlapping networks allow for using a set of measures for the latent concept of the propensity to exchange in the internal market (land, labor, tools and credit exchange). While identification of the standard policy parameters (such as the average treated effect on the treated on welfare indicators) can be obtained through perfect randomization, the structural parameters I consider (individual preferences towards the economic exchange) are identified non-parametrically. Integrating latent component into the model complicates the estimation and the GMM and the Conley (1999) variance estimator are abandoned in favor of semi-parametric procedure and maximum likelihood estimator.

## 1.2 Main contributions of the thesis

The core of the thesis is to unlock the policy parameters, through acquiring knowledge about the causal mechanisms driving change that a policy development project has on the set of defined outcomes. In this particular setting, the mechanisms are related to the degree of the household integration, both in social informal exchanges as well as market interactions. To my knowledge, due to data constraints, this aspect has not been studied in a consistent manner and no causality has been claimed.

In chapter 2, using the context of a very innovative infrastructure evaluation design within the Nigeria CADP, I provide an identification result exploiting the availability of GIS data on distances between the households and the markets, and contribute to filling the gap in the road evaluation literature, by estimating the causal effect of the isolation from the markets on agricultural exchange and revenues.

In chapter 3, using the setting of the Gambia CDDP, I provide a non-parametric identification result and I build a new estimation procedure that integrates the latent component into the dyadic framework in order to account for the individual preferences over choices of social interaction. I derive a semi-parametric estimator of the individual propensity towards the internal economic exchange, and compare their performance with the standard dyadic framework results.



# Capitolo 2

## The causal foundations of structural modeling and policy evaluation

### 2.1 What are the goals of causal inference?

The main tasks of of causal approach, as summarized by Heckman (2008), are:

Task	Description	Requirements
1	Defining the set of hypothetical/counterfactuals	Scientific Theory
2	Identification of causal parameters from hypothetical population data	Mathematical analysis of point or set identification
3	Identifying parameters from real data	Estimation and Testing theory

Empirical studies of the relationship between welfare indicators and anti-poverty instruments could be roughly divided in two main clusters: experimental and non experimental. Both deal with the identification of the direction and magnitude of causality, as defined by Granger (1969), with the necessity to test for the existence of causality in the first place. The definition of causality in economical framework has evolved since 1969. Granger's definition of causality implies that:

- Let  $X$  and  $Y$  be two linear non-deterministic covariance-stationary stochastic processes, such that  $X_t = a * u_t + b * v_t$  ,  $Y_t = c * u_t + d * v_t$ , where  $u$  and  $v$  are uncorrelated white



errors with unit variance, and  $a, b, c, d$  disappears for  $t < 0$ , then  $Y$  does not cause  $X$  if and only if  $a$  and  $b$  can be chosen identically 0.

- The assumption of *strict exogeneity*: If  $(X, Y)$  is written in an autoregressive representation,  $Y$  can be expressed as a lag function of current and past values of  $X$  with a residual that is not correlated with any past and future values of  $X$ , then  $Y$  does not cause  $X$  in Granger's sense.

Granger's definition has been extended subsequently to allow for conditioning on covariates and non-linear specification (see Chamberlain (1984) and Angrist and Kuersteiner (2004)). Recently, asymmetric causality testing has been suggested in the literature in order to separate the causal impact of positive changes from the negative ones (see Hatemi (2012)). However, although theory evolved to accommodate for more general approach, its essence has not changed. What results to be in continuous evolution even today is the focus on which parameters are to be studied and the distinction between so-called policy effects or production function estimates (See Heckman (1992 and 2000)). However, even though there are substantial differences in two paradigms, lately there have been many attempts to build bridges between the two streams in literature (See Heckman (2008) and Pearl (2012)).

This is exactly the econometric literature where the ideas that lie behind this thesis can be nested.

## 2.2 Causal inference and policy effects

The traditional approach to policy evaluation focuses on total effect of one specific variable (typically **the treatment status**) while little or no effort is made to understand the mechanism through which treatment effects operate. Instead of explicit economic theory underlying models, this approach exploits the paradigm of the **randomized controlled trials** and the object of interest are the effects of (quasi) experiments rather than parameters of economic models. It yields estimates such that the effect retrieved is **not invariant** to changes in the characteristics of the policy under investigation. It is more about *effects of causes* than it is about *causes of effects* (Heckman (2010)). Randomized trial is the gold standard paradigm, that allows the counterfactual to be defined as the potential outcome of a hypothetical treatment  $D = d$  that mimics the condition of the perfect experimental setting, ignoring the mechanisms that govern that outcome.

Hence, the causal problem is interpreted as missing data, where potential outcomes under the non treatment status are missing and potential outcome under the received treatments are observed. Therefore,  $Y(X)$  can be understood as a latent variable that expresses its value only

when  $D = d$ , and can be written:

$$Y = dY_1 + (1 - d)Y_0$$

Extrapolating the results driven from (quasi) experimental framework is only valid if the indirect effects are expected to be the same on average in the new population.

The fact that the experiments do not generally recover parameters of the production function is not always a limitation, as often the policy parameters answer directly the policy question, and provide a way to estimate the desired parameters without additional assumptions.

## 2.3 Causal inference and structural parameters

A sharp contrast lies between economic parameters featured in structured approach and the ones from the programme evaluation approach as they are derived from different paradigms. As illustrated in the previous subsection, not all the interesting causal questions can be answered from experimental studies. For that reasons, the representatives of structural models are after different specific type of parameters, that reflect **deep preference** and **technology** attributes that are **invariant** to policy changes. In particular, in policy evaluation setting these parameters would help to answer the question of how would the environment change if a specific policy is implemented.

This would guarantee an external validity of some (not all) parameters to policy changes, allowing them to be interpreted even in a different policy environment. It is used to predict or simulate changes in agents' behavior, and it relies explicitly on the underlying assumptions from the economic model. For example, if the goal is to estimate the effects of a new government policy of incentives to technology adoption, the problem can be framed in the setting where the policy affects the environment through lowering the price of production technology from  $p$  to  $(1 + t)p$ .

How does this work in practice? The core is in finding the adequate **variability**. Using the example of incentives to technology adoption, this variability can be searched in time/region/unit in which the price in the actual status is equal to the new price after the policy gets implemented. This can be used to make predictions about the effects of the program, all else being equal condition maintained and assuming no general equilibrium effects. An instrument like this provides the following answers:

- Evaluating the impact of already implemented interventions on certain outcomes
- Prediction of the impact of intervention implemented in one environment in different environments

- Forecasting the impact of interventions never historically experienced, by knowing the mechanisms through which it works.

## 2.4 Notation and basic definitions

In microeconomic literature, Roy model and its extensions have been widely accepted. It accommodates the idea of hypothetical outcomes, where the space of potential or so-called *counterfactual* outcomes is well defined by economic theory. Let us consider an economy where agents face, for simplicity, only two potential outcomes  $(Y_0, Y_1)$  with the distribution  $F_{Y_0, Y_1}(y_0, y_1)$ , where typically 0 refers to no treatment state and 1 stands for treatment state. We have that  $y_0, y_1$  is a realization of random variables  $Y_0, Y_1$ . Unlike the *Schrodinger's Cat paradox*<sup>1</sup>, economic agents cannot experience both potential outcomes at the same time. What is observed is either  $Y_0$  or  $Y_1$ , and this is what is called **the central policy evaluation problem**.

Roy model also defines agent decision rules, so that the observed outcomes are not necessarily a random sample of the  $F_{Y_0, Y_1}$  distribution. It could be that the agent selects into the treatment if  $Y_1 \geq Y_0 + C$ , where  $C$  is a cost of selecting into the treatment. In that case we have the decision rule:

$$D = 1(Y_1 - Y_0 - C > 0).$$

The the outcome for any person can be written as:

$$Y = DY_1 + (1 - D)Y_0.$$

It is important to notice that policy interventions operate either through shifting the potential outcomes  $(Y_0, Y_1)$  or through changing the cost  $C$ . Agents make subjective evaluation about the *ex-post* outcomes  $(Y_0, Y_1)$ , therefore Roy model also distinguishes between the *ex-post* outcomes and *ex-ante* outcomes  $E(Y_0|\mathfrak{S})$  and  $E(Y_1|\mathfrak{S})$ , where  $\mathfrak{S}$  is the information set that an agent possesses before the treatment. This implies that even if we were able to measure the individual treatment effect, the mere difference  $Y_1 - Y_0$  does not reveal anything about the ex ante subjective evaluation of ex post cost assessment.

## 2.5 Parameters of interest

Since it is almost impossible to come up with a plausible identification of the individual-level treatment effects, without making some extreme assumptions, econometricians typically focus

---

<sup>1</sup>Erwin Schrodinger uses a cat paradox to explain a theory of quantum mechanics called indeterminacy.

on identifying population level parameters, such as:

- Average Treatment Effect  $ATE = E(Y_1 - Y_0)$
- Average Treatment Effect on the Treated  $ATT = E(Y_1 - Y_0|D = 1)$
- Average Treatment Effect on the Untreated  $ATU = E(Y_1 - Y_0|D = 0)$

These parameters were widely accepted and used in policy evaluation literature. However, the parameter that has acquired a lot of attention more recently even in empirical responds to a question of marginal returns to policy. In the previously presented mode, it is identified by observing the agents that are indifferent between  $D = 1$  or  $D = 0$ . A detailed problem setting that conducts to the construction of marginal treatment effects will be explained in section 3.3.

## 2.6 Why empirical research in development economics?

Nelson Mandela in his famous speech in London's Trafalgar Square for the campaign against poverty in the developing world has said :*In this new century, many of the world's poorest countries remain imprisoned, enslaved and in chains. They are trapped in the prison of poverty. It is time to set them free. Like slavery and apartheid, poverty is not natural. It is man-made and it can be overcome and eradicated by the actions of human beings.*

Recently, many researchers engaged in empirical studies in development economics have put some substantial effort to disentangle the channels through which anti-poverty programs can reach the poorest ones. Personally, I strongly believe that it is necessary to put effort in this direction. The diversity of experience of the poor countries development programs is so heterogeneous that it makes it rather difficult to generalize the knowledge in any reliable economic theory. While for the richer countries, the agents' choices could be represented using standard microeconomic models that conserve external validity in many contexts - this is hardly the case for the underdeveloped areas, partly because the size of the population living in poverty is much larger and more heterogeneous in terms of their social, geographical and institutional characteristics.

In the empirical part of the thesis I will use the econometric instruments from the previously described structural policy evaluation approach to build a model structure that allows for better comprehension of the social and geographical isolation as a potential channel through which anti-poverty programs could be more successfully fostered and review what has been done so far in terms of causal inference.

# Capitolo 3

## Rural roads access, markets and welfare

Is access to markets a road to welfare in rural Nigeria? In this chapter, using a very unique setting of the World Bank infrastructure development program, I propose a strategy for the construction of building blocks that can be used as the ingredients of a structural system that explains the interconnection between the roads construction, access to markets and welfare indicators. I exploit results from the innovative structural literature, described in chapter 2 and apply them in a very challenging development setting, where standard economic theory often fails to explain the agents's behavior. This way, I will be able not only to understand better the underlying developing economy principles, but I can perform an *ex-ante* impact evaluation of the road construction on market exchange and, subsequently, on the well being of the households. This is interesting from both methodological and empirical point of view, as there are very few studies that apply newly developed structural econometric tools on real data. Moreover, in contrast to the carefully designed evaluations of many health and education initiatives in developing countries, our knowledge about the impacts of infrastructure projects has remained under-developed, and this makes it even more challenging.

### 3.1 Literature review

In most developing countries, highways and paved motorized roads compose only a small fraction of national road network, while trails, footpaths, tracks and other secondary roads are the ones most commonly used outside urban areas. These non motorized rural roads allow for transit of people, products and production factors, they are characterized by low quality standards and historically have received only marginal part of the national budget allocated to road construction and rehabilitation (see Escobal and Ponce 2002). Inadequate roads raise

transport costs and limit the access of households to local products and labor markets, education and health facilities, and this is especially true during the heavy rainy seasons when the most isolated households remain completely cut out from the community life and exchange. Only recently, both the attention of researchers and policy makers has shifted to small scale rural infrastructure projects with greater recognition of the relationship between high transport cost and poverty (World Bank (2007)).

### **3.1.1 Poverty and access to roads**

The evolution of econometric literature dealing with measurement of road benefits has gone through three major phases in the last two decades. It started with the macro perspective of the relation between welfare and rural infrastructure (see Donovan (1992) and Lipton and Ravallion (1995)). They claim positive impact South East Asian countries where a huge increase in rural infrastructure investments has led to economic growth and reduction in rural poverty.

After the 2000, there has been some growing interest in exploring road benefits at micro-level. Jacoby (2000) has developed a methodology for estimating road benefits at the level of households, exploiting the relationship between the value of farmland and its distance to agricultural markets, confirming the beneficial role of road networks on peasants welfare in terms of agricultural wages. He conducts an empirical analysis on data from Nepal to address the issue of the relationship either between the farmland value and distance to market or wages and distance to market relationship through semi parametric estimation of parameters. Fafchamps and Shilpi (2007) address the question about relative consumption of highly isolated Nepalese households with difficulties accessing the markets, arguing that markets foster competition between individuals, thereby undermining redistributive institutions and values (risk pooling mechanisms). Often, construction sites are selected in such manner to maximize social benefit of the investment, by selecting places with both high poverty and economic potential. Van der Walle (2002) provides a critical review on how the rural road investments are selected in Vietnam, within a cost effectiveness paradigm. This implies that often the road evaluation strategy suffers from endogeneity problem. This issue has been directly addressed by Dinkelman (2008) (see also Duflo and Pande (2007)), who shows that access to rural electrification has an impact on production technologies and market employment. He uses the two wave panel data on communities, and exploits the variation in the projects roll-out that occurred due to some technological constraints. While district's fixed effects control the differences in local labor market conditions, he instruments the project placement with the land gradient. The cost factors are central to identification strategy and the costs depend heavily on proximity and access to existing power

lines, population density of the settlement and finally, the land gradient and terrain. However, at this stage the literature has offered a very scarce evidence of the causal link between the infrastructure and welfare at the household level. One attempt to fill in the gap in the research about the merits of infrastructure can be found in Gonzales-Navarro and Quintana-Domeque (2010). They use the data from the infrastructure experiment in Mexico to evaluate the impact of street pavement on the professional and households' appraisal of property values. They find that the provision of street pavement not only raises the housing and land values but also influences household outcomes such as ownership of vehicles and durable goods. Moreover, they find evidence of a significant impact on credit markets, which acts as the channel through which the street pavement influences durable goods increase. They are the to first deal with the problem of selection bias in the infrastructure placement since infrastructure is likely to be allocated to places with high probability of largest returns, solving the endogeneity and obtaining identification through randomization. Escobal and Ponce (2008) focus on income and consumption as welfare indicators for rural households in Peru, finding an impact on income from wage-employment sources but not in consumption levels, arguing that the road quality improvement has been perceived as transitory.

### **3.1.2 Access to markets**

While in both theoretical and empirical literature it results pretty clear that there is evidence about the positive relationship between road construction and poverty reduction at the micro and macro level, it is not clear what is the mechanism through which road construction generates long run positive impact on household welfare. A very interesting theoretical article by Cannon and Liu (1997) addressed the question about the mechanisms through which road infrastructure generates poverty reduction and argue that it allows for lower production and transaction costs and encouraging economic exchange and foster production and labor specialization. On the other hand, market development has always been seen as a progress generator. Over the last couple of decades, the literature on rural development has highlighted the role of interlinked transactions between economic agents as a proxy of the underlying institutional setting, that explains the elasticity of the rural poor behavior to policy instruments. One of the seminal works that attempted to explain some paradoxes in peasants pervasive behavior can be found in Fachamps et al. (1991), where the non responsiveness of the peasants to price incentives to technology adoption has been explained through the structuralist missing markets model. They argue that the elasticity of peasant household response to price incentives is an important condition for any agricultural development program and they indicate several



channels through which this elasticity can be altered, one of which is the correction of market failure (in the extreme case - missing markets).

Finally, empirical literature, such as Mu and Van der Walle (2009), provides evidence of road rehabilitation on market development. In this particular case, standard counterfactual paradigm has been used through difference-in-differences setting combined with the propensity score matching to highlight significant average impact on the market growth in rural Vietnam. They also suggest that poorer zones have higher impacts due to lower level of initial market development. Evidence that the poorest commercial farmers would benefit the most from lowering cost of transportation and having better access to markets can be found in Fafchamps and Hill (2005).

To conclude, I have considered three waves of development literature - the one that deals with road construction impact on the local and households welfare, the one that deals with empirical evidence of how missing markets can be a reason for the failure of the public anti-poverty programs and finally the literature considering roads construction as an engine of market development. Yet, I found no research that binds the three pieces of evidence together in a unique cause and effect system, that allows for a clear interpretation of the mechanisms through which public expenditure on rural infrastructure helps alleviating the poverty in the isolation zones.

## 3.2 Problem setting

To fix ideas, we can consider  $Y_1$  as the level well being that would be encountered by the household if it involves in an economic exchange on the market ( $D = 1$ ), and  $Y_0$  if does not ( $D = 0$ ). Let  $V$  be household utility, assumed to be absolutely continuous. Let  $Z$  be an instrument such that it can be manipulated by policy maker (i.e. road construction), and implicitly it directly impacts the cost of reaching the market.

$$D = \mathbb{1}(\mu_D(Z) > V)$$

For the moment, the conditioning on a set of observable characteristics  $X$  is assumed invariant with respect to the policy implementation status, and is left implicit. We can also define a **propensity score**:

$$P[Z] = P[D = 1|Z] = P[\mu_D(Z) > V] = F_V[\mu_D(Z)]$$

and it follows that

$$D = \mathbb{1}(F_V[\mu_D(Z)] > F_V[V]) = \mathbb{1}(P[Z] > U_D)$$

where  $U_D$  is random variable normally distributed over  $[0, 1]$ , and  $V$  is assumed to be absolutely continuous. Note that  $P[Z]$  can be considered as a propensity score of agent accessing the

market. Hence we can define agents at the *margin of indifference* as those for whom it is true that:

$$P[Z] = U_D$$

and this would exactly replicate the traditional Heckman selection model. The variables  $Z$  come on top of  $X$  and satisfy the usual exclusion restriction, that is:

$$(Y_1, Y_0, V) \perp Z | X$$

In the setting like that, let us define  $D^b$  the market engagement choice and  $Y^b$  as individuals' welfare **in the *status quo*** (under the current environment and policy regime), where:

$$Y^b = Y_1^b D^b + Y_0^b (1 - D^b)$$

and similarly, for the well being of the compounds **after** the implementation of a new policy, that acts on  $D$  by facilitating access to markets by, say, reducing the cost to reach the market. Then we have that:

$$Y^a = Y_1^a D^a + Y_0^a (1 - D^a)$$

The policy effect that we are interested in is the *ex ante* effect on population well being that results from a marginal change in the policy instrument  $Z$  (i.e. road improvement that facilitates access to market).

Under the conditions stated, it is possible to obtain the **marginal treatment effect** (MTE):

$$\frac{\partial E[Y | P[Z] = p]}{\partial p} = E[Y_1 - Y_0 | U_D = p] \quad (3.1)$$

with  $E[Y_1 - Y_0 | U_D = p]$  being the return for the marginal agent who is indifferent to accessing the market once his propensity score is set to  $p$ . A very important implication is that  $Z$  affects the outcome variable **only** through acting on the propensity score. This means that, in the absence of general equilibrium effects, MTE is invariant to a set of alternative policies that could affect the values taken by propensity score  $P[Z]$ .

MTE is an important building block for many policy parameters (ATE, ATT and LATE, defined in section 2, can be derived from MTE), but above all, it is an essential ingredient for *ex-ante* evaluation. Suppose that the policy acts on  $Z$  by marginally reducing the cost of accessing the markets from  $\eta$  to  $\eta - \epsilon$ . Let  $F_{P[Z]}^b(p)$  be the distribution of  $P[Z]$  at the baseline, and analogously  $F_{P[Z]}^a(p)$  the corresponding distribution under the new policy regime. Then it is possible to show that:

$$E(Y^a - Y^b) = \int_0^1 MTE(p) [F_{P[Z]}^b(p) - F_{P[Z]}^a(p)] dp$$

If we consider a variant of a policy such that  $Z^a = Z^b + \epsilon$  then:

$$\lim_{\epsilon \rightarrow 0} E[Y^a - Y^b]$$

represents what would happen to the welfare of households if a road construction program reduces the cost of accessing the markets, holding everything else fixed.

### 3.3 Estimation method

For the estimation of single components of the setting described in the previous section (propensity score and MTE), I constructed upon the Robinsons's double residual methodology. If we considered a general model of the type:

$$y_i = \beta_0 + x_i\beta + f(z_i) + \epsilon_i \quad (3.2)$$

where  $y_i$  is the value taken by the outcome for the individual  $i$ ,  $x_i$  is a row vector of characteristics, and  $\epsilon_i$  is the disturbance assumed to have zero mean and constant variance. If we apply a conditional expectation on both sides, we have that:

$$E(y_i|z_i) = E(x_i|z_i)\beta + f(z_i) + \quad (3.3)$$

By subtracting (2) from (1) we have that:

$$y_i - E(y_i|z_i) = (x_i - E(x_i|z_i))\beta + f(z_i) + \epsilon_i \quad (3.4)$$

If the conditional expectations are unknown, they can be estimated by using some consistent estimators  $y_i = m_y(z_i) + \epsilon_{1i}$  and  $x_i = m_x(z_i) + \epsilon_{2i}$ . Therefore the final model to estimate is:

$$y_i - \hat{m}_y(z_i) = (x_i - \hat{m}_x(z_i))\beta + \epsilon_i \quad (3.5)$$

where  $x_i - \hat{m}_x(z_i)$  is a row vector of the differences between each explanatory variable and the fitted conditional expectation of  $x$  given  $Z$ . The estimated coefficients vector is therefore:

$$\hat{\beta} = \left( \sum_i (x_i - \hat{m}_x(z_i))'(x_i - \hat{m}_x(z_i)) \right)^{-1} \sum_i (x_i - \hat{m}_x(z_i))(y_i - \hat{m}_y(z_i))$$

with variance, if errors are iid with zero mean and  $\sigma_\epsilon^2$

$$Var(\hat{\beta}) = \sigma_\epsilon^2 \left( \sum_i (x_i - \hat{m}_x(z_i))'(x_i - \hat{m}_x(z_i)) \right)^{-1}$$

Since the parameter vector  $\beta$  is estimated, we can fit the non linear relation between  $y_i$  and  $z_i$  (4) non parametrically:

$$y_i - x\hat{\beta} = \beta_0 + f(z_i) + \epsilon_i. \quad (3.6)$$

## **3.4 Nigeria Commercial Agriculture Development Program**

### **3.4.1 Program description**

Nigeria's Commercial Agriculture Development Project (CADP) is supporting the Federal Government of Nigeria strategy options of diversifying into non-oil sources of growth and away from over dependence on oil and gas. It aims at strengthening agricultural production systems for targeted value chains among small and medium scale commercial farmers in five participating states in the country, and is being implemented in the states of Kaduna, Kano, Enugu, Cross River and Lagos. The total number of beneficiaries expected to directly participate in the project is estimated at 50,000 (10,000 per state) over a period of five years. Small and medium commercial farms will benefit directly through capacity building, technology demonstration and dissemination, access to matching grants and infrastructure, while households will benefit indirectly through access to farm and feeder roads and markets. One of the important sub-components of the program is rural electrification, that will support the provision of electricity to commercial farms where high transmission line exists. This project will finance the rehabilitation and maintenance of rural electricity, including provision of transformers and distribution, feeder lines, poles and accessories from the main transmission lines to commercial farmers facilities in collaboration with the Power Sector Reform Project.

The CADP will be constructing or rehabilitating about 500 kilometers of farm access roads and drainage structures per state in order to connect commercial farms and surrounding communities to the major road networks. It will provide resources for construction of new roads, rehabilitation of existing ones and maintenance of roads to communities and selected agricultural activities. The goal is to provide a better connection of the commercial farms to the local markets. This is obtained through Outputs and Performance Based Road Contracts.

CADP is very market oriented. Its future goal is the creation of domestic and export markets; this will support market development, awareness and knowledge sharing for commercial farmers. It will finance market information system, including market price surveys, website and information kiosks, market/financial linkages, (including promotion of products for supply chain financing, use of crops as collateral) quality control measures and standards; food safety, exchange programs, agricultural trade fairs and shows, local and international study tours.

### 3.4.2 Data

The data was collected through the *ad-hoc* survey conducted by National Bureau of Statistics, under the supervision of Research Institute for the Evaluation of Public Policies (Trento) and the World Bank. The collection went through two major phases: first the catchment areas were defined as areas around the road construction site where potential road beneficiaries live, fixing the radius around the treated area. It was followed by the so-called *listing survey* - a census of the households within the catchment areas, done in 2010. Then, during the following year, approximately one fourth of the households in five states was randomly selected to enter the baseline sample. Very rich and detailed information was collected about individual level demographics, education, schooling, labor and employment, time allocation to other activities, and access to markets and facilities, production, sales, expenditures and credit market access. Moreover, the survey collected tracking data for each household within the evaluation sample in order to follow households through the follow-up surveys, and global positioning system (GPS) information for each household and each of their plots. The final sample includes 4241 households and 27320 individuals.

Around 14 % of the questionnaires were collected in Kaduna, while in all other states this number is almost equally distributed and it amounts for about 21%. Most of the enumerated households cultivate land, in all the participating states, except Lagos state where the proportion of those not cultivating land exceeds land cultivators.

The primary source of electricity for domestic use is the public grid. While a very small proportion of households in Lagos state have no source of domestic electricity, the proportion of households without any source of domestic electricity is very high in Kano at 34%, and between 9% to 13% in the other states. A vast and overwhelming majority of households have no access to any source of electricity for their farms and enterprises. Over 85% of households in Enugu and over 90% of households in Kano, Kaduna and Cross River states have no source of electricity for non-domestic use.

In each state, a majority of households use the main road as the fastest means of accessing markets, irrespective of whether they visit the market daily, weekly or monthly. Households which are the most geographically isolated with respect to the time required to reach the nearest market are characterized by lower levels of the commonly employed indicators of well being. For example, using GIS data from Lagos we found negative association between consumption and income, and distance from markets. The same negative relationship with geographic isolation is found when energy adoption is considered. (see appendix A1 for the selected list of descriptive statistics.)

### 3.5 Empirical results

CADP baseline data provides a very rich ‘photography’ of characteristics and activities of rural households, however final sample contains very different dimension of farm production. Therefore, I consider the sub-sample of 1324 peasants involved in crop value chain production (maize, rice, cocoa, palm oil), and leave out the production of livestock and fisheries. This choice is motivated by the fact that the production process, its seasonality and production technology is rather heterogeneous between livestock and crops. There are many indicators of households’ well being that could be derived, however the welfare indicators I have considered are short term income, labor intensity and education expenditure. The reason for this selection stands in the nature of the policy that we are evaluating. The roads that are to be constructed are small rural roads - they connect the farms to the closest motorized roads, or to the nearest market. Most of the time the length of the road does not exceed 5km and sometimes roads are not even being constructed from zero - few of them were just rehabilitated. Therefore, I expect that this kind of intervention should not have general equilibrium effects (i.e. on prices of goods and services). The idea, supported by the literature review, is that the improved/constructed road access reduces the **time** and **cost** of going to the market, and this reflects on the short term labor market outcomes such as the hours worked or the total pay of workers. Also, I expect some intra household redistributive effects, for example the reduction in transport expenditure could allow for an increase in the education expenditure. In future research, other welfare outcomes will be taken in consideration.

As explained at the very beginning of this chapter, the *channel* I am considering is the market access. For rural households, markets are not just places of exchange of goods and services - these are usually meeting points, information exchange engines and are often situated near other types of local institutions (such as schools, hospitals, ecc.). In our sample, most peasants know where the market is and it is rather tricky to find a suitable indicators of market access. Even though we have the information on how often household members go to the nearest market, its use could be misleading, since individuals could declare to go to the market every day but this would NOT mean that they have commercial reasons of accessing it. Since we are interested in exploring the hypothesis of lowering the cost of accessing markets, this could be more transparent if market is defined as a place of some kind of economic exchange, related to farmers activity, than if we use the broader market access definition. It follows that the indicators of market access are the following:

- Dummy that indicates whether the household sells on the market
- Ratio of the quantity sold on the market over the total quantity harvested

- Ratio of the value sold on the market over the total harvest value.

While the first indicator is more straightforward in defining the market access, the second and the third are intended to be the proxies of market activity intensity measure.

The availability of GIS data on the **exact** latitude and longitude position of the compounds on one hand and the markets on another, has allowed us to obtain the measure of the **real distances** (in miles and minutes) that necessary if a peasant wants to arrive from compounds to the markets. Both motorized road and small rural roads are considered as possible paths in the calculus algorithm. This is important because the distances we calculate have to mimic the changes in road networks that will happen after the CADP full scale implementation. In order to control for the density of economic activity we use administrative area dummies, as well as the number of markets available in the range of 20miles of euclidean distance radius (see appendix A.1 for an example of the utility of the Global Positioning System).

### 3.5.1 Isolation and welfare

I begin by exploring the relationship between the selected welfare indicators and the distance from the markets. Since the preliminary graphical analysis has indicated that this relationship might not be linear, semi-parametric estimation seemed to be a reasonable choice, using Robinson double residuals method (as in section 3.3). I control for many pre-program observable characteristics, such as age, gender, literacy and formal education of the household head, household demographic structure, commodity interest group membership, crop, total area cultivated, the number of active markets within a 20 miles radius and controlling for administrative area effects.

The welfare indicators presented in Figures 3.1, 3.2 and 3.3 are, respectively, the individuals' last pay, labor intensity expressed in the length of the last working period and per child education expenditure. It appears that all three indicators are inversely correlated to the market distance. This means that further the compounds are from the market place, they have shorter working intervals/are less payed. Also, the education expenditure decreases with household isolation, as expected.

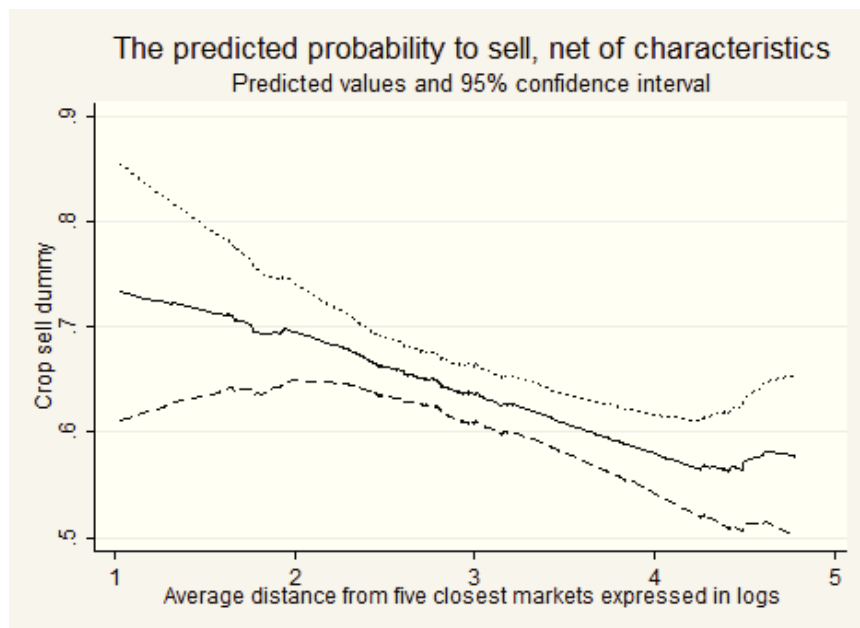


Figura 3.1: Semi parametric estimation of the correlation between the distance from the nearest markets and the last salary perceived by the HH head

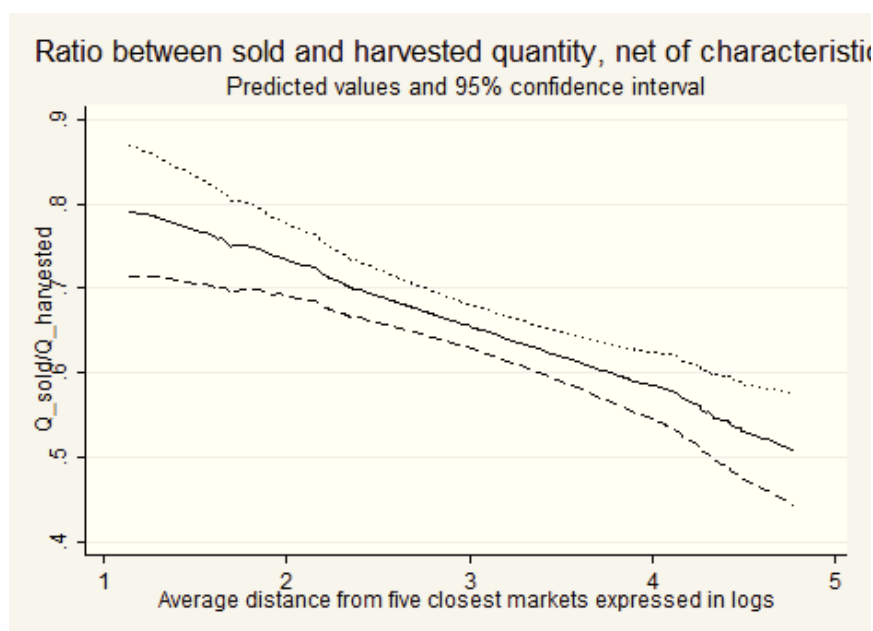


Figura 3.2: Semi parametric estimation of the correlation between the distance from the nearest markets and the length of the last payed working interval

### 3.5.2 Isolation and access to market - propensity score estimation

I start with the simplest indicator: a binary outcome that indicates whether the household, prior to the baseline, has sold on the market. Hence, as before, I begin by regressing semi-



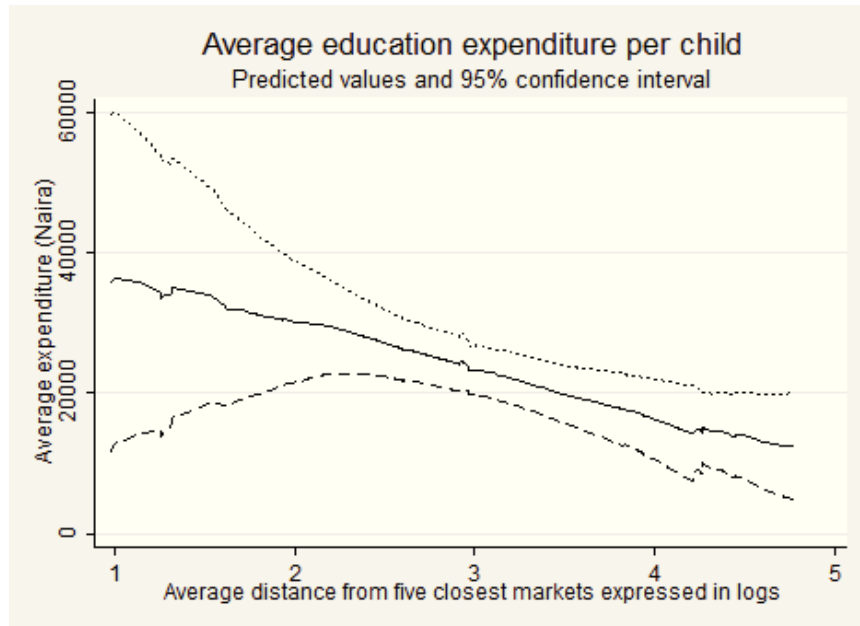


Figura 3.3: **Semi parametric estimation of the effect of distance from the nearest market and education expenditure.**

parametrically the decision  $D$  to sell to the market on the log of average distance from the 5 closest markets. We can see the result in Figure 3.4: as expected it is less likely that the most isolated households sell on the market. There is some rumour at the first and last quantiles of distances distribution, but the slope of the relationship emerges quite clearly. Note that what we have obtained here is the **propensity score** of selling on the market, given the observable characteristics and the distance.

### 3.5.3 Isolation and access to market - MTE estimation

Once that the propensity score has been estimated, I proceed with the semi-parametric estimation of the MTE distribution. I follow the procedure described in section 3.3, and I use the three welfare indicators as outcome variables. What I obtain can be seen in figure 3.5. We observe that higher the propensity score, that is higher the probability of engaging in market activity, higher are the labor market outcomes, and also per child education expenditure increases. This is pretty much what is expected. One of the possible explanation is that market place differentiate the labor engagement, so that it becomes more likely for the farmers to diversify the sources of income, that helps the income continuity and prevents from risk (crop loss, bad year, ecc.).

However, even though crop sell indicator is simple and intuitive, it suffers from the problem of scarce support, since the instrumental variation in distance  $Z$  does not span the propensity

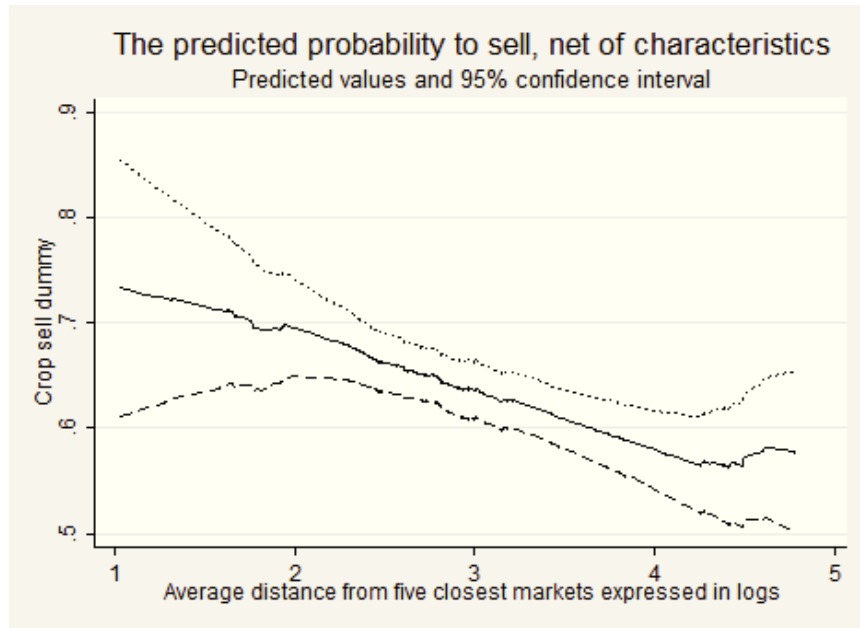


Figura 3.4: Semi parametric estimation of the effect of distance from the nearest market to probability to sell

score over the full unit interval.

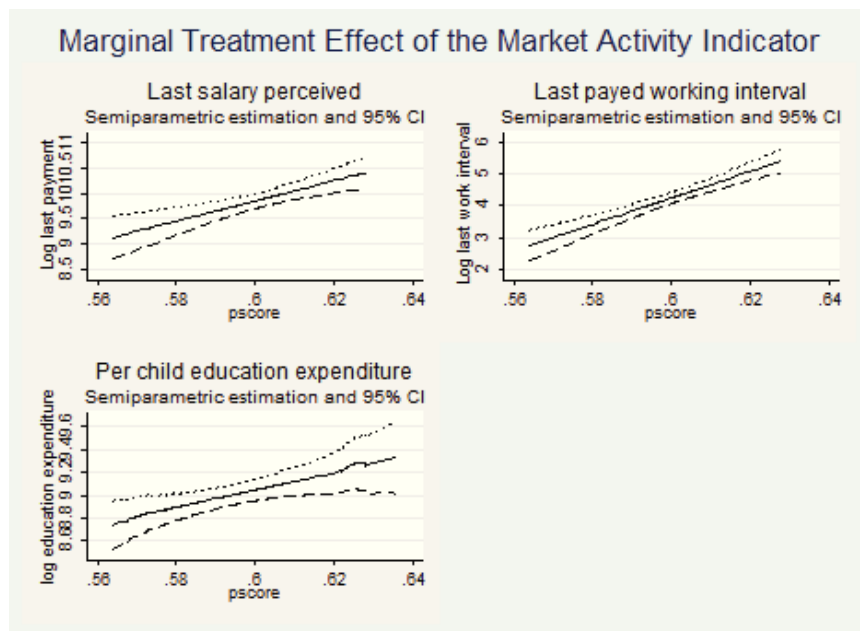


Figura 3.5: Semi parametric estimation of the marginal treatment effect of market access to selected welfare outcomes

The support problem is solved if we use another proxy for market activity (see figure 3.7). The first option is to use the fraction of quantity sold over total quantity harvested. The

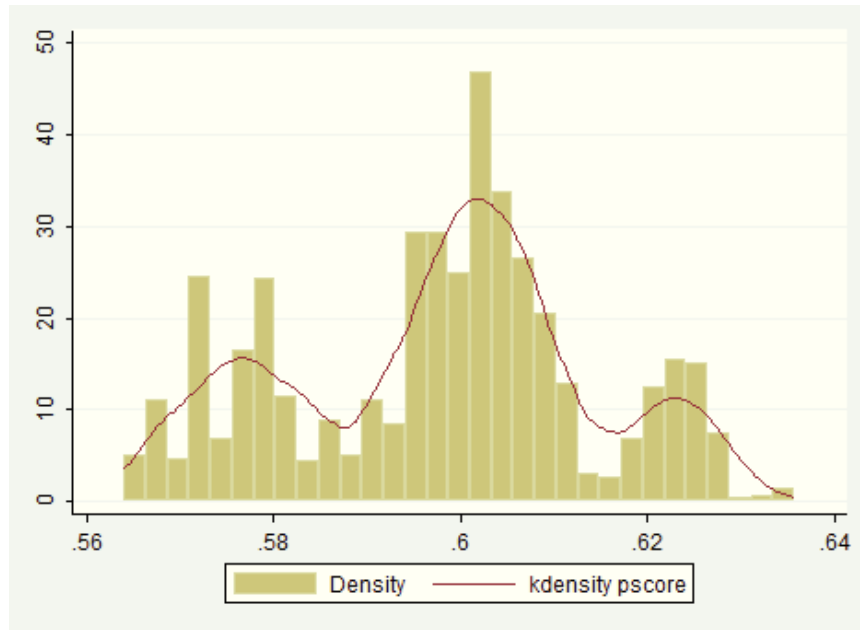


Figura 3.6: **The distribution of the estimated propensity score**

original indicator is continuous in a range  $(0, 1)$ , but it can be dichotomized in order to fit the original idea of market access: if the household sells more than about 60% of its harvest (the median value)- it can be considered as intense market activity and the farm can be defined as strongly market-oriented. If the sale share is less than the median, the compound can be considered as one with less access to market exchange. We can see from the Figure 7.8 that the slope is not so large in the case of last pay, while the MTE is larger in the case of length of the last working interval and education expenditure.

The second option of approximating the intensity of the household's activity on the market is to consider the ratio of the value sold (in Naira) and estimated value harvested. In terms of interpretation, it is rather similar to the previous indicator, what changes is the measurement unit. Assuming constant average prices in the catchment areas, even this variable can be dichotomised without losing any generality. I use the same rule with the cut-off of defined as the median ratio. The results follow in the Figures 7.9 and 7.10. The propensity score has a much better coverage of  $(0, 1)$  interval, while the MTE distribution is rather similar to the ones found in Figure 7.7.

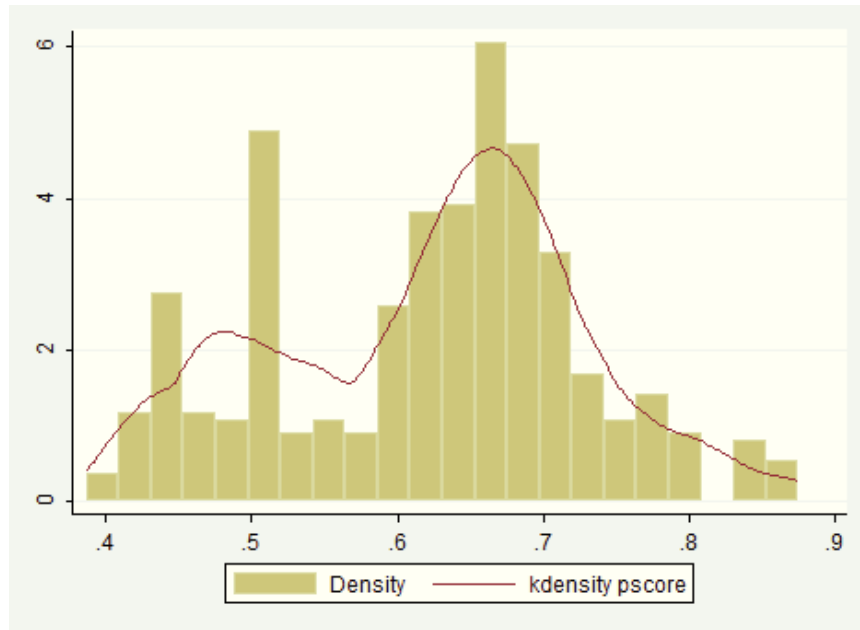


Figure 3.7: The distribution of the estimated propensity score

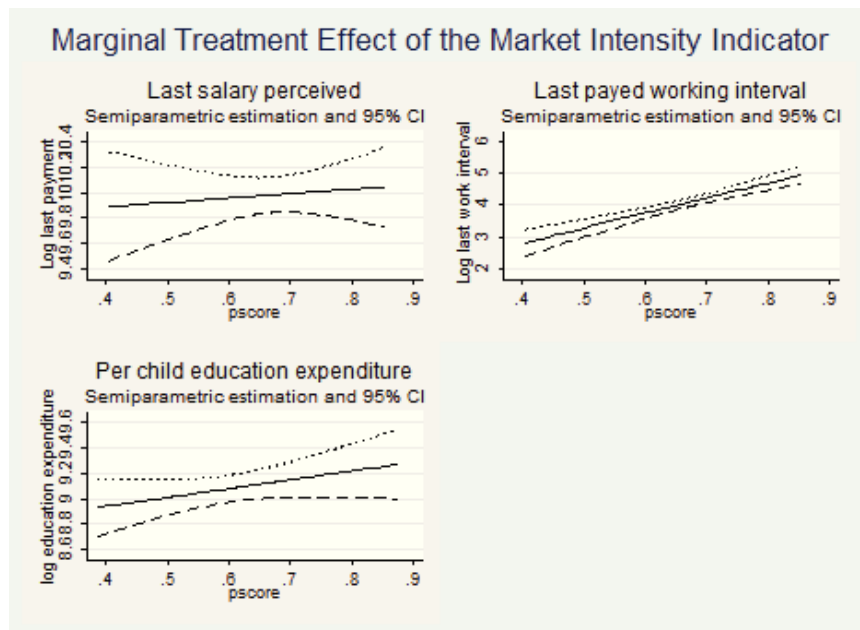


Figure 3.8: The distribution of the estimated MTE

### 3.6 Future research

Once that I have clean setting to refer to and coherent and complete data and once that I have a clear signal of structural relationship between the three variables of interes (roads, markets and welfare), next step would be to numerically derive the distribution of MTE and obtain the empirical counterpart of the *ex-ante* estimator. Later on, I would like to reproduce the findings

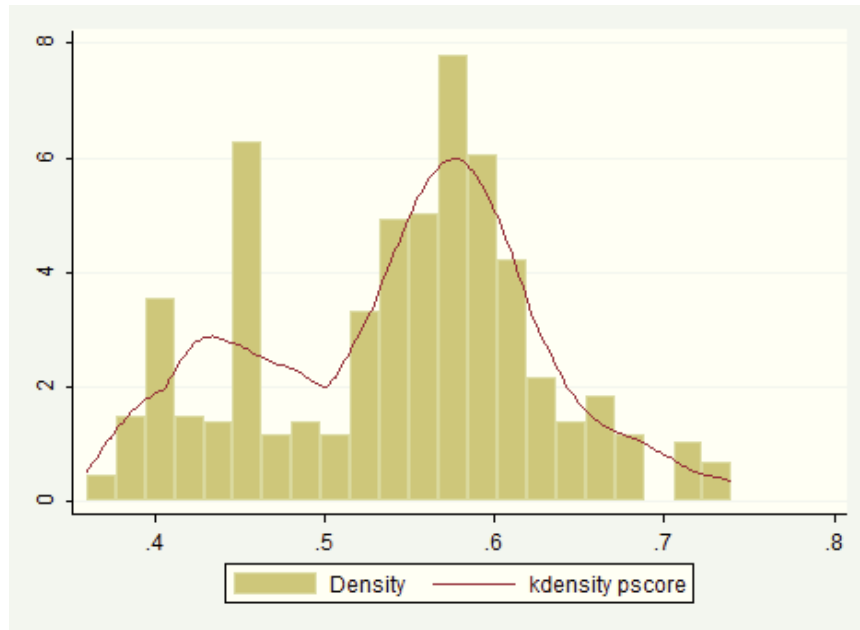


Figure 3.9: The distribution of the estimated propensity score

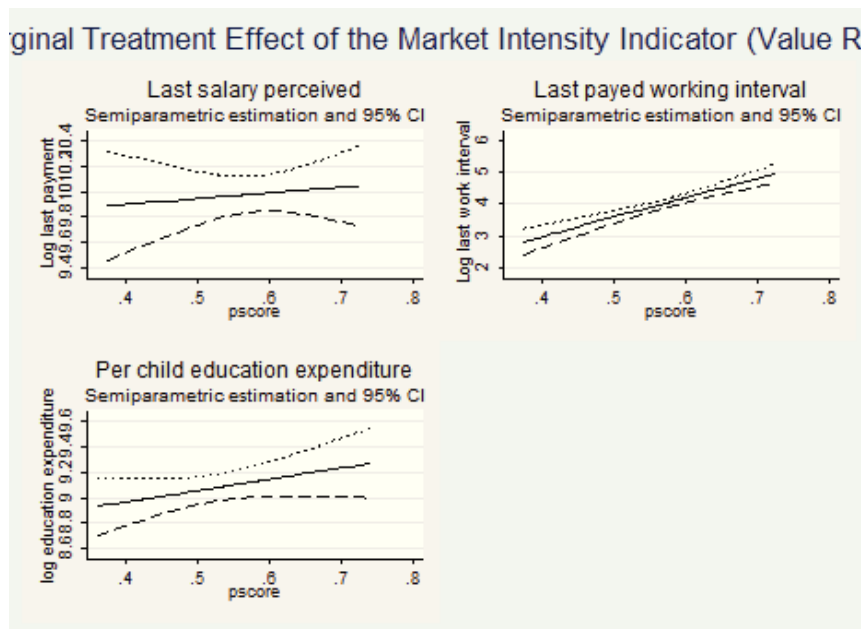


Figure 3.10: The distribution of the estimated MTE

in different settings of farmers production function, in order to assess the external validity of our results on a small scale. This is important, since a policy like this would soon be scaled-up and CADP policy makers have the urge to obtain the first results, without waiting for the follow-up survey. Hence, ex-ante evaluation could be a good compromise between the impact evaluation rigirousity and practical need of the stakeholders involved in the project.

# Capitolo 4

## Social Isolation and Community Driven Development Programs

Economic behavioral theory teaches that the choice of engaging in any transaction depends both on the observed and measurable characteristics of actors and their unobserved preferences. In this particular context, we distinguish between the individual propensity to engage in internal market transactions and the propensity of a specific couple of households to exchange. The objective of the research is twofold: conduct an empirical research on a very rich dataset on social networks in rural Gambia, using the setting of the Community Driven Program Evaluation (CDDP) and focus on the impact of program on the link formation. More importantly, the goal is to disentangle the effect of the program that operates indirectly on the probability of forming a link, through modifying the existing network attributes or modifying exchange preferences of individuals; thus find convincing way to identify the program effect on the distribution of individual and dyad fixed effects (preferences), by construction of a structural model that relies upon reasonable assumptions.

### 4.1 Literature review

In the last decade there has been a growing interest of studying the the embeddedness of economic activity in social settings has been fundamental to sociologists for some time, following the work of Granovetter (1985). Jackson (2008) draws a a parallel to the recent rise of interest in behavioral economics, which comes upon the realization that a deeper structural underpinnings of human behavior can help enrich the modeling of economic decision makers.

A very interesting social network flavored research has been done in the context of development economics. The topics regarding risk pooling and information and innovation diffusion have been of major interest, and both have the common ingredient - the network as a measurement of

social capital through which risk are shared or information about new technologies diffused. In correspondence to these two research interests, empirical literature divides in two main streams: one highlights the social processes behind risk sharing arrangements and another deals with the models of information diffusion.

Recent empirical studies such as Fafchamps and Gubert (2007) and Dercon and Weerdt (2006) investigate whether the risk sharing methods are formed in order to maximize the mutual gains from pooling income risk. Krishnan and Sciubba (2009) examine the role of network structure in determining the impact of labor sharing networks on outcomes. In terms of social networks as the key factor in the transmission of information for technological innovations in agriculture, significant contributions have been provided by Conley and Udry (2008) and Bandiera and Rasul (2006), while Miguel and Kremer (2003) consider the use of innovative drugs in network-diffusion context.

## 4.2 A First Specification

In this section we first define some basic social network analysis concepts, that would later be related with econometric literature and methodology.

### 4.2.1 Basic definitions

**Social network analysis** can be defined as mapping and measuring of relationships and flows between people, groups or other connected information/knowledge entities. The nodes in the network are the households while the links show relationships or flows between the nodes.

- Let us define  $l_{ij} = 1$  if there is a link between two households, and 0 otherwise.
- **Household degree** is defined as the count of the number of ties to other actors in the network  $d_i(g^m) = \sum_j l_{ij}$ .

### 4.2.2 Specification at Household level

Let  $Y_s^{(int)}$  be an indicator for the internal exchange across the dimension  $s$  and let  $Y_s^{(ext)}$  be an indicator for the external exchange. Moreover, let  $Z$  and  $W$  be the vectors of, respectively, individual and village level characteristics. Let  $u_i$  be the preference towards the internal exchange and  $e_{is}$  the idiosyncratic error term. For the internal exchange we have that:

$$Y_{is}^{(int)} = \alpha_0 + \alpha_1 z_i + \alpha_2 w_i + u_i + e_{is} \quad (4.1)$$

For the external exchange we have that:

$$Y_{is}^{(ext)} = \gamma_0 + \gamma_1 z_i + \gamma_2 w_i - u_i + e_{is} \quad (4.2)$$

### 4.2.3 Specification at Dyad level

Let  $Y_{ij}^{(s)}$  be the outcome undirected adjacency matrix for  $s$  economical exchange, where  $i$  and  $j$  are indices for individuals. Let  $Z$  and  $W$  be vectors of village and link related attributes, respectively and let  $u_i$  be the propensity towards the internal exchange, and  $u_{ij}$  be the  $ij$  specific exchange propensity. Let the following dyadic regressions be defined (as in Fafchamps and Gubert (2006) and Arcand and Fafchamps (2008)):

$$Y_{ij}^{(s)} = \beta_0 + \beta_1 |z_i - z_j| + \beta_2 (z_i + z_j) + \beta_3 w_{ij} + u_i + u_j + u_{ij} + \epsilon_{ij}, \quad (4.3)$$

The parameters  $\beta$  are informative about the structure of the network that is related to the matrix  $Y$ . If it weren't for the preferences, reduced-form regression (3) could be estimated by a simple OLS method/ML method, adjusting standard errors to account for the correlation across rows/columns of the matrix.

In this setting, we exploit the multidimensional nature of the available data, hence use jointly information different  $Y$ 's to derive sufficient conditions that allow identification, and thus to learn about any change induced by the CDDP intervention. The final goal is to use more flexible models in (1) and (2), i.e. use a non-parametric or semi-parametric approach to obtain the distribution of  $u_{ij}$  and  $u_i$ .

### 4.2.4 Dyads and basic inference

As far as the inference related to the dyadic regressions goes, the standard literature suggests the standard errors to be estimated using Conley's Cluster Covariance Estimator (CCCE) Cluster covariance estimators are used with data that has a group structure with independence across groups. Using individual level observation, we can write the estimator as:

$$AVar(\hat{\beta}) = \frac{1}{N - K} (X'X)^{-1} \left( \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N \left[ \frac{m_{ijkl}}{2N} X_{ij} u_{ij} u'_{kl} X_{kl} \right] \right) (X'X)^{-1}$$

where:

- $N$  is the number of dyadic observations.
- $X$  is the vector of all the regressors,  $X_{ij}$  is the vector of dyad specific regressors.
- $m_{ijkl} = 1$  if the dyads in question have common elements, and hence are correlated (if  $i = k$ ,  $j = l$ ,  $i = l$  or  $j = k$ ), and 0 otherwise.



As a non parametric alternative, quadratic assignment procedure can be used.

## 4.2.5 One useful descriptive tool: Oaxaca-Blinder Decomposition

Oaxaca-Blinder Decomposition allows for the first decomposition of the total effect into endowment effect (change in  $X$ ) and coefficients difference effect (change in  $\beta$ ). It is often used to explain gender gap in labor economics literature regarding career outcomes. Let the following potential dyadic regressions be defined for the matrix  $Y$

$$Y_{ij}^{(1)} = \beta_0 + \beta_1|z_i^{(1)} - z_j^{(1)}| + \beta_2(z_i^{(1)} + z_j^{(1)}) + \beta_3w_{ij}^{(1)} + u_{ij}^{(1)},$$

$$Y_{ij}^{(0)} = \tilde{\beta}_0 + \tilde{\beta}_1|z_i^{(0)} - z_j^{(0)}| + \tilde{\beta}_2(z_i^{(0)} + z_j^{(0)}) + \tilde{\beta}_3w_{ij}^{(0)} + u_{ij}^{(0)}.$$

It is easy to show that:

$$\begin{aligned} \Delta Y = (\beta_0 - \tilde{\beta}_0) &+ \beta_1|z_i^{(1)} - z_j^{(1)}| - \tilde{\beta}_1|z_i^{(0)} - z_j^{(0)}| \\ &+ \beta_2(z_i^{(1)} + z_j^{(1)}) - \tilde{\beta}_2(z_i^{(0)} + z_j^{(0)}) \\ &+ \beta_3w^{(1)} - \tilde{\beta}_3w^{(0)} + \epsilon \end{aligned}$$

In this setting, I will use it to decompose the total effect of the exogenous shock provided by the World Bank CDDP onto the propensity to exchange - into the change due to the households characteristics and change due to different drive ( $\beta$ ) that the program has onto the outcome. (see Appendix A.2 for details.)

## 4.3 Model specification

### 4.3.1 Factor analysis

The main idea that lies behind the factor model is to use a set of measures for the latent concept of the propensity to exchange in the internal market. The measures we can use are the actual exchange along the four economic dimensions. The joint variation among there measures is divided into the common part  $\theta$  and the random part  $\epsilon$ . All the economic questions depend in a different manner on the latent factor - therefore each question can have a different factor loading, which can be interpreted as a coefficient of the latent factor in the regression of a link on the latent factor.

Let  $L$  be the dummy that indicates whether the individual  $i$  has an external link, while  $Y$  is a dummy that indicates whether a pair of individuals in the same village have a link. Let  $*Y$  and  $*L$  be their latent analogous quantities. There are repeated measures on these individuals/pairs represented by  $k = 4$  different dimensions of economic exchange (land, labor, production tools and credit). Let  $\theta$  be a latent factor, that can be interpreted as a propensity to exchange in the internal market, and

let  $\delta$  be the pair specific propensity to exchange. We can construct the following simple simultaneous equation model for the internal links:

$$Y_{ij}^k = 0, 1$$

$$*Y_{ij}^k = \lambda^k(\theta^i + \delta_{ij}) + \tau_{ij}$$

as well as for the external links:

$$L_{ij}^k = 0, 1$$

$$*L_i^k = \lambda^k(-\theta^i) + \epsilon_i$$

where  $\epsilon$  and  $\tau$  are simple uncorrelated idiosyncratic errors.

### 4.3.2 Identification

the identification of factor models is based on the covariance matrix of the items:

$$Cov(*Y) = \Lambda \Sigma_f \Lambda' + \Omega$$

and

$$Cov(*l) = \Lambda \Sigma_f \Lambda' + E$$

where  $\Lambda$  is a matrix of loadings,  $\Sigma_f$  is a variance-covariance matrix of the factors, and  $\Omega$  and  $E$  are diagonal matrices of ‘uniqueness’ variances of idiosyncratic terms. The assumption I make are:

$$\theta \perp \epsilon, \omega$$

$$\epsilon, \omega \approx N(0, \sigma)$$

Therefore, we need to identify  $L = 2 * k = 8$  factor loadings for  $L = 8$  items and a variances of the factor  $\theta$  and  $\theta + \delta$ , and the two variances of the error terms - which makes 12 parameters to estimate. Therefore we have

$$\frac{L(L-1)}{2} \geq 12$$

which is the identification condition, and actually we have that:

$$\frac{8(8-1)}{2} \geq 12.$$

## 4.4 Model specification in our case

Let  $k$  denote different exchange types,  $(i, j)$  are the households, and  $Y_{ijk}$  is a dummy that indicates whether the household  $i$  and  $j$  exchange over the  $k$  dimension.

$$Y_{ijk}^* = \lambda_k(\alpha_i + \alpha_j + \alpha_{ij}) + U_{ijk}$$

where:

$$Y_{ijk} = I(Y_{ijk}^* \geq 0)$$

In the simplest formulation, we can assume  $(\alpha_{ij}, \alpha_i, \alpha_j)$  are independent across individuals and copulas and from  $U_{ij}^k$ . Moreover,  $U_{ij} \perp U_{rs}$  for  $i \neq r$  and  $r \neq s$ . Under these conditions, standard results imply that the distributions of  $U_i^k$  and  $U_i$  are non-parametrically identified, by Kotlarski theorem.

#### 4.4.1 Identification

**Theorem 4.4.1.** (Kotlarski's Theorem) *If  $T_1^s = \theta_1 + v_1$  and  $T_2^s = \theta_1 + v_2$  and  $\theta_1 \perp v_1 \perp v_2$ , the means of all three generating random variables are finite, and the conditions of Fubini's theorem are satisfied for each random variable, and the random variables possess nonvanishing characteristic functions, then the densities of  $(\theta, v_1, v_2), f(\theta), f(v_1)$  and  $f(v_2)$ , respectively, are identified.*

If I average the outcomes for the individuals  $i$  over the individuals  $j$ , I obtain the expression for the so-called **degree**:

$$\begin{aligned} \bar{Y}_{i1} &= \alpha_i + \bar{U}_i^1, \\ \bar{Y}_{i2} &= \lambda_2 \alpha_i + \bar{U}_{i2}, \\ \bar{Y}_{i3} &= \lambda_3 \alpha_i + \bar{U}_{i3}, \\ \bar{Y}_{i4} &= \lambda_4 \alpha_i + \bar{U}_{i4}, \end{aligned}$$

Note that in the first equation there is a constraint of  $\lambda_1 = 1$ , and the system can be rewritten in the following manner:

$$\begin{aligned} \bar{Y}_{i1} &= \alpha_i + \bar{U}_i^1, \\ \frac{\bar{Y}_{ik}}{\lambda_k} &= \alpha_i + \bar{U}_{*ik}^*, \end{aligned}$$

where  $\bar{U}_{*ik}^* = \frac{\bar{U}_i}{\lambda_k}$  and  $k \neq 1$ . Thus, we can non parametrically identify the densities of  $f(\alpha_1)$  and  $f(U_k)$ .

For the moment I allow for the following distribution assumptions:

$$\begin{aligned} \alpha_i &\stackrel{iid}{\sim} N(0, \sigma_\alpha^2) \\ U_i &\stackrel{iid}{\sim} N(0, \sigma_u^2) \end{aligned}$$

The conditional distributions of the outcome variable for  $k = 1$  is therefore assumed to be:

$$Y_{ik} | \alpha_i \sim N\left(\alpha_i, \frac{\sigma_u^2}{n-1}\right)$$

and for  $k \neq 1$

$$Y_{ik}|\alpha_i \sim N(\lambda_k \alpha, \frac{\sigma_u^2}{\lambda_k^2(n-1)})$$

Note that in this case I assume uncorrelated errors across different measures.

## 4.5 Likelihood specification

The likelihood function can be defined from  $(^1\bar{Y}_i, ^2\bar{Y}_i, ^3\bar{Y}_i, ^4\bar{Y}_i)$ . The joint likelihood function for the  $i$ th individual is given by:

$$f(y_{i1}, y_{i2}, y_{i3}, y_{i4}) = \int_{-\text{inf}}^{+\text{inf}} f(\alpha_i) \prod_{k=1}^4 f(y_{ik}|\alpha_i) d\alpha_i$$

where  $f(\alpha_i) = \frac{1}{\sigma_\alpha} \phi \frac{\alpha_i}{\sigma_\alpha}$  and  $f(y_{ik}|\alpha_i) = \frac{1}{\sigma_{u^*}} \phi(\frac{y_{ik} - \lambda_k \alpha_i}{\sigma_{u^*}})$ . Note that  $\sigma_{u^*} = \frac{\sigma_u^2}{n-1}$ .

The second piece of the integrand can be developed as follows:

$$L(\bar{Y}_{ik}; \sigma_u^2, \lambda_2, \lambda_3, \lambda_4 | \alpha_1) \approx \left(\frac{\sigma_u^2}{n-1}\right)^{-1/2} \exp\left(\frac{(y_1 - \alpha_1)^2}{\left(\frac{2\sigma_u^2}{n-1}\right)}\right) * \left(\frac{\sigma_u^2}{n-1}\right) \prod_{k=2}^4 \frac{1}{\lambda_k^2} \exp\left(\frac{(y_k - \lambda_k \alpha_1)^2}{\frac{2\sigma_u^2}{\lambda_k^2(n-1)}}\right)$$

the log likelihood function for the  $i$ th individual is the following:

$$\begin{aligned} l(\bar{Y}_{ik}; \sigma_u^2, \lambda_2, \lambda_3, \lambda_4 | \alpha_1) &\approx \frac{4}{2} \log \sigma_u^2 - 2 * 4 \log \lambda_k + \sum_{k=1}^4 \frac{(y_i - \lambda_k \alpha_i)}{\frac{2\sigma_u^2}{\lambda_k(n-1)}} \\ &\approx -2 \log \sigma_u^2 - 8 \log \lambda_k + \frac{\lambda_k(n-1)}{2\sigma_u^2} \sum_{k=1}^4 (y_i - \lambda_k \alpha_i) \end{aligned}$$

Where  $\lambda_k = 1$  if  $k = 1$ .

It follows that:

$$\begin{aligned} l(\bar{Y}_{ik}; \sigma_u^2, \lambda_2, \lambda_3, \lambda_4 | \alpha_1) &\approx \\ &\approx -2n \log \sigma_u^2 - 8n \log \lambda_k + \frac{\lambda_k(n-1)}{2\sigma_u^2} \sum_{i=1}^n \sum_{k=1}^4 (y_i - \lambda_k \alpha_i) \end{aligned}$$

Where  $\lambda_k = 1$  if  $k = 1$ .

## 4.6 Data and results

Data collection strategy was done through Rapid Rural Appraisal Methodology, broadly described as a systematic but semi-structured activity out in the field by a multidisciplinary team and is designed to obtain new information and to formulate new hypotheses about rural life (see Chambers (1994)). The compounds, households and individuals were asked to reply jointly on three main sections of the questionnaires: basic village info, household information and interactions along different socio-economic dimensions (exchange of labor and production tools, credit and land market, marriage and kinship relationship and neighborhood information). Information was collected under the form of census from 60 randomly drawn small villages, out of which 30 treatment and 30 control villages, that have population  $< 1000$  according to 2003 census) were involved in the sample, that amounts for 3320 households. See appendix A.2 for the descriptive statistics of the sample at micro and macro level.

From the dyadic regressions, it results that there is assortative matching along the individual and households characteristics. It can be seen that there is a strong preference towards male-male households exchange. Also large land owners are more likely to engage in land exchange with those who have small portion of land, especially if the land sum is higher (meaning the richer the landing land owner). The complete results of dyadic regressions can be found in appendix A.2

Results from Oaxaca Blinder decomposition of the WB external shock suggest that there is no difference in endowments found while there are differences in coefficients through which the external shock operates (see appendix A.2 for details). This is a strong signal to continue on the research and underpin the mechanisms through which the WB program works.

## 4.7 Conclusions and Future Work

Apart from working in an under-explored development country context, this research is quite challenging because of pretty new dyadic framework, multi directional causality scheme (with at least 4 networks involved), binary outcome and complex unobserved component structure. In my future research, I intend to use jointly information from adjacency matrices and derive empirical counterparts of the previously shown model, that would allow me to pin down different channels through which an external shock alters the existing equilibria in rural areas of the Gambia. I also intend to compare the results obtain by the above constructed strategy with the results from the existing standard dyadic regression approach.

# Appendice A

## Nigeria Commercial Agriculture Development Program

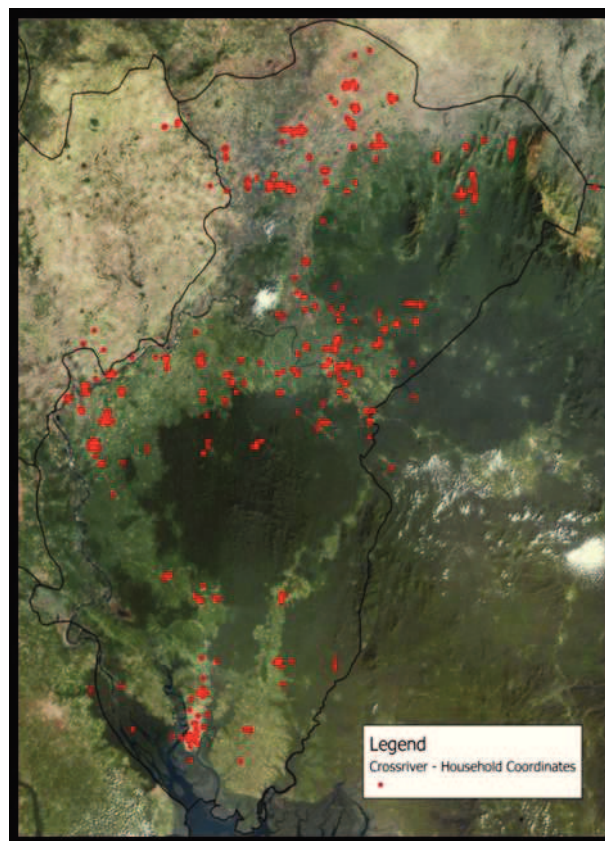


Figura A.1: Geographical dispersion of the households in the state of Cross River - GIS satellite image

Tabella A.1: Summary statistics

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>
household size	6.195	2.925
number of minor	2.249	2.135
female HH head	0.119	0.324
age	47.594	12.792
%literate	0.852	0.355
% no edication	0.079	0.27
% primary school	0.262	0.44
% secondary	0.273	0.446
% quranic	0.075	0.263
% owns a car	0.195	0.396
% owns a moto/scooter	0.376	0.485
% hours worked per week	29.504	32.915
yearly income (naira)	330606	1580462
% cultivates rice	0.124	0.33
% cultivates oil palm	0.046	0.208
% cultivates cocoa	0.086	0.28
% cultivates fruit tree	0.071	0.257
% cultivates maize	0.327	0.469

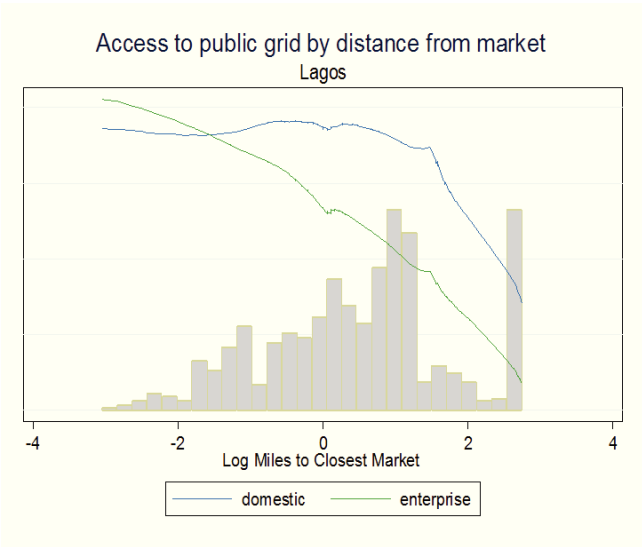


Figura A.2: Access to electricity and distance to the nearest market (in miles)

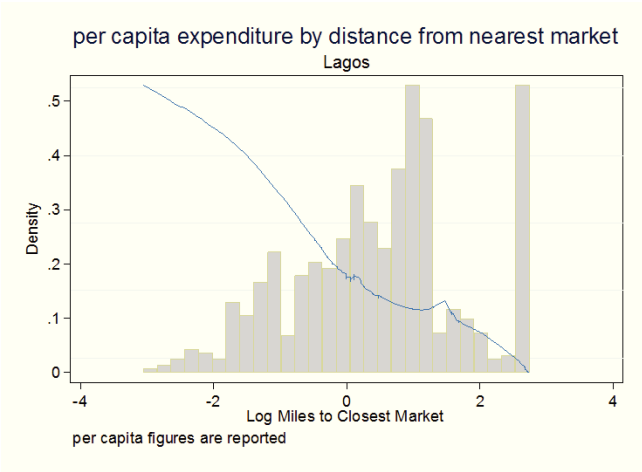


Figura A.3: Per capita expenditure and distance to the nearest market (in miles)





# Appendice B

## Gambia Community Driven Development Program

Tabella B.1: Household Level Summary statistics

Variable	Mean	Std. Dev.	N
Household Size	12.801	12.518	2866
The age of household head	3.899	0.321	2865
% Female HH head	0.066	0.249	2885
% Formal education	0.146	0.353	3254
Polyhamous households	0.402	0.49	3254
Non muslim	0.031	0.173	3254
Emigrants in the hh	0.544	0.498	3254
Kamanyango pachuko	0.047	0.212	3254
Land per workers (in hectars)	2.306	7.425	2858
Per capita income (in thousand dalasies)	3.57	5.134	2851
proportion of agricultural income	0.12	0.237	2852

Tabella B.2: Village Level Summary statistics

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>
village populatation	548.459	177.142
% Illiterate people	0.368	0.21
% not having access to electricity	0.971	0.038
% using public wc	0.371	0.293
% not having access to improved water	0.859	0.235
WB poverty index	0.642	0.097
% in semiurban area	0.112	0.316
N		3254

Tabella B.3: Dyadic regressions

	Land	Labor	Production Tools	Credit
PovertyIndex	0.165 (0.26)	1.046 (1.93)	-0.0131 (-0.03)	-0.291 (-0.40)
Income (sum)	-0.0849 (-1.90)	-0.0914* (-2.26)	-0.0835** (-2.88)	-0.0603 (-1.11)
Income (difference)	-0.0940 (-1.46)	-0.0878 (-1.28)	0.0455 (0.69)	0.183 (1.74)
Household size (difference)	0.0766 (1.07)	0.144 (1.71)	-0.0640 (-1.18)	0.198* (2.29)
Household size (sum)	0.324*** (5.65)	0.418*** (7.83)	0.371*** (8.95)	0.445*** (7.18)
Nworkers (difference)	-0.0732 (-0.31)	-0.283 (-1.43)	-0.362* (-2.16)	-0.439 (-1.91)
Nworkers (sum)	0.521** (3.16)	0.334* (2.29)	0.487*** (3.64)	0.0858 (0.51)
Land (difference)	0.192*** (4.84)	-0.00196 (-0.18)	-0.0172 (-1.95)	-0.0242* (-2.30)
Land (sum)	0.151*** (5.09)	0.0204** (2.64)	0.0217*** (3.62)	0.0140 (1.72)
Both muslim	-0.0110 (-0.04)	0.813** (2.67)	0.343 (1.36)	0.256 (1.08)
Both male hh head	0.178 (1.29)	0.464*** (3.45)	0.703*** (5.24)	0.456** (2.82)
Both female hh head	-0.294 (-0.58)	-1.086 (-1.51)	-0.0548 (-0.10)	0.479 (1.29)
Both have formal education	0.00818 (0.04)	-0.177 (-0.84)	-0.361* (-1.99)	-0.0589 (-0.29)
Age (difference)	0.0120 (0.10)	0.135 (1.12)	-0.0929 (-0.90)	-0.458** (-3.07)
Age (sum)	0.151 (0.95)	0.104 (1.09)	-0.0833 (-1.21)	-0.0645 (-0.66)
<i>N</i>	73573	73628	73573	73628

Results of logit regression for binary data, inference accounting for three way clustering

*t* statistics in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Tabella B.4: Oaxaca-Blinder decomposition for Internal links

	Land	Labor	Production Tools	Credit
<b>Differential</b>				
Prediction_1	-1.397*** (-7.82)	-1.335*** (-6.32)	-0.653*** (-3.94)	-1.516*** (-12.42)
Prediction_2	-1.569*** (-7.56)	-0.655*** (-3.50)	-0.879** (-2.63)	-1.487*** (-6.71)
Difference	0.172 (0.63)	-0.680* (-2.41)	0.226 (0.60)	-0.0293 (-0.12)
<b>Decomposition</b>				
Endowments	0.0516 (0.26)	0.252 (1.14)	0.498 (1.31)	0.476 (1.40)
Coefficients	0.307 (1.33)	-0.371 (-1.46)	0.228 (0.63)	-0.0142 (-0.05)
Interaction	-0.186 (-0.86)	-0.561 (-1.95)	-0.500 (-1.27)	-0.491 (-1.45)
<i>N</i>	2768	2776	2765	2771

*t* statistics in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Tabella B.5: Oaxaca-Blinder decomposition for External Links

	Land	Labor	Production Tools	Credit
<b>Differential</b>				
Prediction_1	-1.694*** (-8.40)	-2.689*** (-13.61)	-2.217*** (-13.98)	-1.653*** (-13.95)
Prediction_2	-3.798*** (-9.36)	-4.500*** (-7.33)	-3.045*** (-9.45)	-2.436*** (-14.11)
Difference	2.104*** (4.64)	1.812** (2.81)	0.828* (2.30)	0.783*** (3.74)
<b>Decomposition</b>				
Endowments	-0.722 (-1.50)	-0.366 (-0.64)	-0.132 (-0.32)	0.264 (1.05)
Coefficients	1.937*** (3.90)	1.400* (2.06)	0.605 (1.41)	0.598** (2.78)
Interaction	0.889 (1.67)	0.777 (1.25)	0.354 (0.72)	-0.0795 (-0.32)
<i>N</i>	2601	2573	2756	2776

*t* statistics in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



# Bibliografia

- J.L. Arcand and M. Fafchamps. Matching in community-based organizations. *Journal of Development Economics*, 98(2):203–219, 2012.
- O. Bandiera and I. Rasul. Social networks and technology adoption in northern mozambique. *The Economic Journal*, 116(514):869–902, 2006.
- A. Banerjee, E. Duflo, and N. Qian. On the road: Access to transportation infrastructure and economic growth in china. Technical report, National Bureau of Economic Research, 2012.
- Abul Barkat, SH Khan, M Rahman, and S Zaman. *Economic and social impact evaluation study of the rural electrification program in Bangladesh*. 2002.
- Kate Bird, A McKay, and Isaac Shinyekwa. *Isolation and poverty: the relationship between spatially differentiated access to goods and services and poverty*. 2011.
- P. Carneiro, K.T. Hansen, and J.J. Heckman. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college. Technical report, National Bureau of Economic Research, 2003.
- T.G. Conley. Gmm estimation with cross sectional dependence. *Journal of econometrics*, 92(1):1–45, 1999.
- T.G. Conley and C.R. Udry. Learning about a new technology: Pineapple in ghana. *The American Economic Review*, 100(1):35–69, 2010.
- Ogunlade Davidson and Stanford a. Mwakasonda. Electricity access for the poor: a study of South Africa and Zimbabwe. *Energy for Sustainable Development*, 8(4):26–40, December 2004.
- J. De Weerdt and S. Dercon. Risk-sharing networks and insurance against illness. *Journal of Development Economics*, 81(2):337–356, 2006.
- Stefan Dercon, Daniel O. Gilligan, John Hoddinott, and Tassew Woldehanna. The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages. *American Journal of Agricultural Economics*, 91(4):1007–1021, November 2009.



- D. Donaldson. Railroads of the raj: Estimating the impact of transportation infrastructure. Technical report, National Bureau of Economic Research, 2010.
- Dave Donaldson. Railroads of the Raj: Estimating the impact of transportation infrastructure. 2010.
- Steven N. Durlauf and Hisatoshi Tanaka. Understanding Regression Versus Variance Tests for Social Interactions. *Economic Inquiry*, 46(1):25–28, January 2008.
- J. Escobal and C. Ponce. The benefits of rural roads. enhancing income opportunities for the rural poor. Technical report, Grupo de Análisis para el Desarrollo (GRADE), 2003.
- M. Fafchamps and F. Gubert. The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326–350, 2007.
- RW Fairlie. An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement*, 2005.
- M. Gonzalez-Navarro and C. Quintana-Domeque. *Street Pavement: Results from an Infrastructure Experiment in Mexico*. Industrial Relations Section, Princeton University, 2010.
- CS Goodman. Do the poor pay more? *The Journal of Marketing*, (January):1–25, 1968.
- James J. Heckman. Econometric Causality. *International Statistical Review*, 76(1):1–27, April 2008.
- James J Heckman. Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of economic literature*, 48(2):356–398, June 2010.
- GW Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics*, 86(February):4–29, 2004.
- Hanan G. Jacoby and Bart Minten. On measuring the benefits of lower transport costs. *Journal of Development Economics*, 89(1):28–38, May 2009.
- H.G. Jacoby. Access to markets and the benefits of rural roads. *The Economic Journal*, 110(465):713–737, 2000.
- E. Miguel and M. Kremer. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217, 2003.
- Ren Mu and Dominique van de Walle. Rural Roads and Local Market Development in Vietnam. *Journal of Development Studies*, 47(5):709–734, May 2011.
- P.E. Todd and K.I. Wolpin. On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485):F3–F33, 2003.

P.E. Todd and K.I. Wolpin. Ex ante evaluation of social programs. *Annales d'Economie et de Statistique*, pages 263–291, 2008.

Dominique Van De Walle and Dorothyjean Cratty. Impact Evaluation of a Rural Road Rehabilitation Project. (January), 2002.

Peter Warr. Road Improvement and Poverty Reduction in Laos. *2007 Conference (51st), February 13-16, 2007.*, 2007.