ECONOMIC ANALYSIS OF SUBSISTENCE

FARMERS IN BOLIVIA:

THREE ESSAYS

By

NANEIDA REGINA LAZARTE ALCALÁ

Bachelor of Science in Economics
Universidad Católica Boliviana
La Paz, Bolivia
1996

Master of Arts in International Cooperation and
Management of Development Projects
Universidad Complutense de Madrid
Instituto Universitario Ortega y Gasset
Madrid, Spain
2002

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2010

ECONOMIC ANALYSIS OF SUBSISTENCE

FARMERS IN BOLIVIA:

THREE ESSAYS

Dissertation Approved:

Dr. Lee Adkins

Dissertation Chair

Dr. Bidisha Lahiri

Dissertation Adviser

Dr. Andreas Savvides

Committee Member

Dr. Federico Aime

Outside Committee Member

Dr. Mark E. Payton

Dean of the Graduate College

ACKNOWLEDGMENTS

This work is the result of more than two years of intensive and dynamic learning. For this endeavor I've had the support and inspiration of a number of exceptional people to whom I would like to express my sincere and deepest gratitude.

My profound gratitude goes to Dr. Lee Adkins, my mentor, guide, teacher and friend. He kept an eye on the progress of my work and was always available when I needed his advice. His thoughtful suggestions, critical comments, and helpful hints were always present in the critical path of this project. His dedication will always be appreciated and remembered.

As my research advisor, Dr. Bidisha Lahiri accompanied me in each stage of this work. Her comments and constructive feedback helped me significantly improve and polish the work at hand. I am very thankful with her.

A very special acknowledgment is due to Dr. Andreas Savvides, who accepted to remain part of the project despite being across the sea. His careful reading of my drafts and helpful comments and suggestions greatly improved the focus and quality of my work. His teachings, emphasis on research, and many of his publications on economic development, inspired me and led me to the selection of the area of my specialization.

A warm thank goes also to Wilma Wehr, Administrative Support Specialist at the Economics and Legal Studies Department, whose outstanding professionalism, timely support, and kind smile made it possible to concrete all the stages of my graduate studies.

Finally, none of this would be possible without the support, endurance, patience and love of my mother, my husband and my son. They were my motivation to work hard and my inspiration to move forward. The endless hours, days, weeks and months of work that prevented me to be with them are recorded in every word of this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I – PREAMBLE: THE AGRICULTURAL SECTOR IN BOLIVIA

1.1 Introduction

This chapter presents a brief review of the agricultural sector in Bolivia; it presents its three remarkably different geographic regions, the type of farming performed in each one, the crops they produced, and the animals they raise. It also explain the participation of the most important sub-sectors in the country's agriculture sector, in terms of their contribution to the sector's GDP, growth trends, and land holdings. Finally, it gives some information regarding land tenure and soil quality, and the functioning of relevant markets for the farmers.

1.2 Bolivia's geography and its regions

With an annual US$ 1,363 of GDP per capita[1], Bolivia is among the poorest countries in Latin America with one of the highest share of rural population (36%) compared to the average of its surrounding neighbors or to the average of Latin America (around 22% in both cases).[2]Poverty in the country is particularly prevalent in rural areas where almost 80% of the population is poor compared with 53% of the urban areas. Likewise, the distribution of

[1]The Gross Domestic Product (GDP) numbers have been taken from INE (2008); they are preliminary data for the years 2006- 2007.

[2] CEPAL (2006).

the national income is highly unequal, especially in the rural area, with a rural Gini coefficient of 0.63.[3] Based on these numbers, it is evident that the major structural and institutional reforms introduced during the past 23 years in the country, which promoted the return of economic stability and growth, did not succeed in terms of poverty, inequality and social exclusion. The accomplishment of a significant progress on such areas is not only legitimate, but also critical for the preservation of the democratic political system.[4]

Bolivia is a sparsely populated country (9.1 million people or 8.3 inhabitants per km square), with the *Altiplano* and *Valles* regions more heavily populated than the *Llanos* one. Although the country lies entirely within tropical latitudes, climate conditions vary widely from tropical in the lowlands to frigid in the highest parts of the Andes. The northern plains, or the *Llanos,* have a tropical wet climate with year-round high temperatures, high humidity, and heavy rainfall. The central valleys (*Valles*) are located at an altitude between 5,900 and 9,500 feet above sea level, and have a tropical wet and dry climate. Finally, the *Altiplano* or highland plateau, is located at an altitude of 9,800-13,000 feet above sea level, is swept by strong cold winds, and has an arid, chilly climate, with sharp differences in daily temperature and decreasing amounts of rainfall from north to south.

This geographic and climate diversity, coupled with other historic features, have given way to remarkable regional differences that are reflected in the regional levels of production; i.e. the GDP per capita of the department of Tarija which is geographically located in the Valleys of, is US$ 3,033, compared to US$ 1,972 for the department of Pando located in the Lowlands, or to the US$ 1,003 of the department of Potosi located in the Andean region.

[3] The country's urban Gini is 0.55.Data for poverty and inequality comes from CEPAL (2006).

[4] Many authors have extensively explained the subject. For a comprehensive work see for example Government of Bolivia (2001).

Next, we develop some of the highlights of the agricultural activity at the national level, and for each of the three geographic regions.

1.3 The agricultural sector

Farm production has a very important role in Bolivia's economy. Based on data collected from the ECLAC (Economic Commission for Latin America and the Caribbean)[5] the total agricultural gross domestic product (agricultural GDP) has represented for the country an average of 13% of the total gross domestic product for the last 15 years, with a slightly decreasing trend during the 2000s. The same relation shows a much lower participation average, 5%, for both, the Latin American Southern-Cone countries, and Latin American countries.

There are significant differences, however, among the various sub-sectors that make up the agricultural sector in the country. Next we briefly mention them.

1.3.1 Agricultural sub-sectors

Some are particularly dynamic, such as the industrialized farming, while others, i.e. traditional farming, have stagnated or even declined over time.[6] The agro-industrial GDP grew at an annual average of 11% between 1990 and 2004, while the average growth rate for the non-industrialized agriculture products was 2% for the same period.[7] According to CEPAL *et. al* (2001), the enhanced farm production during the 1980s, 1990s and first years of 2000's, in both agro-industrial crops and more traditional crops, has been a result of many factors. Among them are an increased external demand (especially for the Andean

[5] Access through the web page http://www.eclac.org/estadisticas/bases/

[6] The way the agriculture sector is structured in this paper, and most of the data, are based on the work of Malky (2004).

[7] We can understand agro-industry as a field in the industry that transforms the raw materials from agriculture, livestock, fishery and forest into added-value products.

market), the expansion of the amount of land devoted to cultivation, and the usage of improved seeds.

The industrialized farmers are mainly geographically located in the *Llanos,* and produce export products like cotton, sugarcane and soybeans. Its participation in the sector's GDP has been at an average of 13% during the period 1990-2004. This group owns comparatively larger amounts of land, are more specialized, modern and profitable, and more easily linked to foreign markets.

The non-industrialized agriculture farmers are mainly located in the *Altiplano,* and account in average for the 47% percent of the total agricultural GDP (1990-2004). This group is mainly comprised by farmers that own small plots of land, have little specialization or technology, are based on family work, and whose agriculture outputs are devoted to self subsistence or domestic markets. The production of this group consists mainly of roots and tubers, corn, vegetables and fruits. According to the Agrarian Question literature,[8] we can characterize these small farmers as a "poverty refuge" *minifundio,* economically incapable of significant accumulation, and yet resistant to total elimination, probably because the peasants themselves highly value their land as a way to assure their subsistence.

The numbers for the country's output per land-unit are below the region's average for the majority of the agriculture products that Bolivia produces, as UDAPE (2004) shows. Even though the study does not give any specific information of the productivity of the two already mentioned types of agriculture farming, it concludes that agricultural yields are especially low in the Andean region characterized by poor small farmers. This view is also

[8] The Agrarian Question debates the destiny of the small-scale farmers e.g. its survival or extinction over time; it includes three different hypotheses, namely: the *Chayanovian* view that explains the survival of a competitive and stable peasant sector, the *class differentiation* view, that argues the disappearance of the sector over time, and the "poverty refuge" view, that imputes a resistance and longevity of small farmers, with income levels that only ensure their survival. More details in Carter and Mesbah (1993).

shared by the analysis of the World Bank (1995) that explains the profile of rural producers in Bolivia in terms of the same two groups; one very poor and traditional, mainly located in the Andean region and Valleys, called *minifundios* (e.g., farms with too little land), and the other one, called the *modern subsector,* located in the tropical lowlands, and comprising medium and large holdings, ranging from 50 to tens of thousands of hectares.

Livestock production is the agriculture sub-sector that has the second highest percentage of participation in the total sector's GDP (27% in average), and has expanded at an annual average of 2.5% during the same period. The livestock sector produces poultry, lamb, beef, camelids and pork products, and also derived products such as wool, leather, milk and eggs. Based on climate conditions, soil quality and entrepreneur skills differences, and in terms of our geographic regions, cattle is mostly raised in the eastern part of the country, poultry in the central valleys, and camelids in the Altiplano.

Finally, the last important sub-sector is the production of coca leaves, that accounts for an average of 6% of the total agriculture GDP and a negative annual growth rate of 7.8% during 1990s and first part of 2000s. The relevance of the coca production sector is not only due to its economic impact, but also for its political connotation. During recent years, the production of coca leaves shows an increasing trend, which could be explained by the expansion of the amount of land devoted to its cultivation, and also by the decrease in the eradication efforts of coca plantations due to the active and combative unions and the *cocalero* movement. The vigorous efforts to eradicate coca and promote special programmes for alternative crops seem to have failed. The maximum number of hectares legally allowed for coca production is 12,000 hectares, but according to Laserna (2004), the actual plantations exceed 30,000 hectares. These volumes of coca production situate the country today globally in the third place, after Colombia and Peru.

The areas where coca plantations are concentrated are the tropical valleys of the departments of Cochabamba (Central Valleys) and La Paz (Altiplano). Specifically, coca plantations are cultivated in the Chapare[9] and Yungas regions of those Departments[10].

Aymara and Inca civilizations have historically cultivated coca leaves for centuries in the high subtropical Yungas valleys for traditional (i.e. legal) consumption (mainly for medicinal purposes), hence, this region was designated to be the primary legal cultivation zone under Law 1008 since 1988.[11] Since all other coca production, including that in the Chapare, was deemed illegal and subject to eradication, the results of the vigorous efforts to eradicate coca plantations in that region have helped to increase the coca production in and around the legal zone of the Yungas exceeding the legal amount of hectares allowed in that region.

According to much of economists in Bolivia, one of the most problematic and relevant issues in the country's economic history has been the land tenure and soil quality. Next, we will mention some of its most important features.

1.3.2 Farming land, and land tenure system

The most important factor of production and asset for rural farmers has traditionally been land. It contributes to wealth not just directly, as an input of production or as a way of accumulation, but it also might allow its owners the possibility of access to capital (credit) when it can be used as collateral (especially in an environment where production risk cannot be insured), or it can even entail some *status* among neighbors or relatives.

[9] When authors refer to the Chapare region, it includes not only the Chapare province, but also Carrasco and Tiraque, all of them in Cochabamba department. For more details see Laserna (2004).

[10] In this case the reference to the Yungas region includes the provinces Nor and Sud Yungas and Caranavi, and the "localidades", equivalent to a village, of La Asunta and Apolo. Ibid., p. 7.

[11] http://www.congreso.gov.bo/11leyes/

The land suitable for intensive farming in the country totals only 3% of its total area[12]; most of the arable land is located in the lowplains (72%), some in the central valleys have 19%, and the remaining 9% in the *Altiplano*.[13]

According to Barraclough and Domike (1966) during the 1950s and early 1960s, Latin America had one of the most polarized agrarian systems in the world; large land extensions or *latifundios*[14] accounted for 5% of the agricultural units and owned 80% of the total land; *minifundios* represented 80% of the agricultural units, but owned only 5% of the land. Land distribution in Bolivia has always been worse than the rest of Latin America, and its agrarian reform considered among the most complex and broader in the region.[15]

Conflicts over land, which have occurred recurrently in the country's history, and the paternalist and rent-seeking perspectives of the state prevalent since the colonial era, have had a decisive impact on the current distribution of land and other productive assets. Looking at the farm structure that prevailed in Bolivia during the last agricultural census, there were 314,600 farms that cover around 23 million Has (including livestock pastures). From those, the small farms cover only 1.5% of the total farm area. At the other extreme are the large holdings occupying 70% of the total farm area are representing only 1.5 % of the holdings (World Bank, 1995).

This concentration of land in a few wealthy hands has been frequent in the country since colonial times. The feudal agrarian structure established along the country in the second half of the 19th century, and still widespread during the first half of the 20th century, expropriated and distributed *indigenous* land among the Conquistadors and new

---

[12] Most of the land in Bolivia is suitable for extensive livestock (25%), for extensive farming (25%), and about a 20% is not suitable for any kind of farming at all; see Morales (2000).

[13] Ibid.: 12.

[14] Latifundio comes from the latin word lātifundium (lātus, "spacious" and fundus, "farm, estate"). The term is used in Bolivia to describe large extensions of land owned by one landlord; they were called *Haciendas* during the colonial times and were allowed to have forced labor (indigenous).

[15] Thorp (1998).

terratenientes[16], and put later indigenous people under peasant servitude. Even in those early years of the country's history, regional differences were already present; while in the *Altiplano* and *Valles* regions the process of expropriation and *feudal* system continued until 1953, the Lowlands saw the setting of the foundations of the agrarian capitalism.[17]

The social revolution that put in power a new government in 1952 soon saw the first fruits. In 1953 the Agrarian Reform Law was enacted and it established as its main objectives: a) the elimination of the large landholdings where no technological improvements have been applied to the production process, and where the owner did not work the land by himself (*latifundios*); b) the reversion of those lands to the state; c) the abolition of peasant servitude; and d) the distribution of the reclaimed land to landless peasants (through a *settlement process*).[18]

Even when a large number of rural workers and peasants were the principal beneficiaries of the Agrarian Reform, the bureaucratic procedures established for implementing the reform, the pressure from powerful groups, and the lack of peasants' resources to acquire agriculture machinery and adopt new technologies of production, facilitated the return of much of the land to the large owners and promoted their transformation and modernization from the old *haciendas* into the new capitalist farms.[19]

During the decades that followed the Agrarian Reform, the state has implemented three different processes of rural land acquisition.[20] The first one took place in the *traditional* regions of the reform, that is, the highlands and semi-arid valleys, and was absolutely redistributive; large estates or *latifundios* were reversed in favor of the state, and later

---

[16] Landowners of large holdings enjoyed economic and political power, and worked the land with wage workers and slaves.
[17] Antezana (1969).
[18] Agrarian Reform No. 3464 of 1953.
[19] Thorp (1998).
[20] This section is based on Muñoz (1999).

redistributed to the peasants occupying them. The total land acquired by this group is around 20 to 25 million hectares, but only about 700,000 are actually cultivated due to the low quality of land. The most common land tenure system is the family farm with a small landholding.

A second process of land occupation was a result of "settlement programs" (*programas de colonización*) sponsored by the state in the humid *Valles* and part of the *Llanos*.[21] Under this form of access to land, the distributions reached between 3 and 5 millions of hectares, with an actual cultivated land of around 300,000 hectares per year. The resulting land tenure system has been the gradual transformation from *colonos* or settlers into commercial farmers, with different rates of transformation among the settlement areas[22]

Finally, there has been a "consolidation process" (programas de consolidación), through which private property rights over the rural land have been allocated; the beneficiaries have been large landowners in the eastern plains of the country (*Llanos*), who occupied the land when the Agrarian Law was enacted. However, this demand-based allocation of thousands of hectares of public land has continued benefiting well-off individuals, and hence promoting the formation of large agriculture properties for speculative purposes also. Here agriculture and livestock farms coexist.

Since the Agrarian Reform did not improve the unequal distribution of rural lands and did not vindicate the historical rights of indigenous people, and given the existence of high insecurity of tenure, in 1996 the INRA Law was enacted with the following primary objectives: i) to create a new tenure regime and a immediate titling for the communal lands

---

[21] With the main objective of increasing productive efficiency of the agricultural sector, the government of Bolivia launched in the decade of the 1960's a process of *colonization* whose main purpose was the settlement of farmers from the Altiplano and inter-Andean valleys, in the tropical lowlands of the departments of La Paz, Cochabamba and Santa Cruz. See more details in OEA (1975).

[22] *Colono* is a beneficiary of the settlement program.

of origin (TCO's for its acronym in Spanish) for indigenous groups, ii) to define a new distribution system for public lands, handing over land to landless or small peasants, and supplying land to the rest of potential buyers at market prices, iii) to establish a period of regularization of property rights, and iv) to consolidate land rights and redefine property rights.[23] However, it remained established in the new Law that small farmers could not use their land as collateral, i.e. ownership of small farms cannot be seized.[24]

Despite the reforms introduced in the legal framework and in the institutions governing land administration, the results have again fallen short of expectations in both land regularization and land distribution, which is clear from the data provided regarding the severely skewed land distribution existent in the country, and from the more than 30% of the landholdings that are still in process of being regularized (processes of reclaiming and titling). [25]

The results after decades of the Reform and after its modification in the framework of the Law 1715 in 1996[26], have been different depending on the region. However, the main tendency shows the re-concentration of land: more than 60% of the agricultural units do not surpass 4 hectares of extension, and they are predominantly located in the central valleys and in the Altiplano.

1.3.3 The functioning of relevant markets for agriculture

Regarding the functioning of the relevant rural markets in Bolivia, there are, as expected, multiple imperfections that constrain households in general, but small farmers in

---

[23] Munoz (1999) pp. 20-21.

[24] It is actually possible that a small farmer can use its land as collateral *if* he decides to be reclassified as a *medium farmer* that must pay taxes, and whose property (e.g. land holdings) can be seized. This seems to be rather a non-interesting option for the majority of the small farmers.

[25] See in detail also Urioste (2003).

[26] Ley de Modificacion de la Reforma Agraria (INRA)

particularly, limiting their ability to generate income from their asset endowments. Let us briefly review below some of the characteristic features of the credit and land markets in the country.

Bolivian capital markets, as expected, do not function well. There are two options to access formal credit, through commercial banks or through microfinance institutions, both of them regulated by the state. However, the two of them discriminate against small landholders, rationing them out completely due to legal barriers as is the case of the banks, or offering them credit at rather high interest rates as is the case of the microfinance suppliers (Valdivia 2004)[27]. Informal sources of credit such as relatives, friends, money lenders, or Non-Governmental Organizations (NGO) are available for some rural households; the topic however is beyond the scope of this paper. Nevertheless, Carter (1989) explains that informal *lenders* translate informational problems and the weak collateral position of small farms into high interest rates. This would make informal credit even more conservative than borrowings from formal markets, which would leave the small farmers with a credit market from which they cannot borrow, or they can do it at very unfavorable conditions, which may end up discouraging any attempt of borrowing.[28] For the large farmers, mainly geographically located in the north-eastern part of the country (*Llanos*), there is plenty of access to the capital-rationed credit market through commercial banks, where the available capital is a function of their land holdings which are used as collateral.

The country's rural land market, in turn, has not been found to be as active as in the rest of Latin America. Based on empirical work, it has been established that the sale of rural land in

[27] Microfinance institutions do offer access to credit to small borrowers (in both urban and rural areas) without asking for collateral, but the loans are short-term, their size is rather small, and the interest rates quite high, reflecting the high operational costs of these institutions for collecting, and the risk of the loans.

[28] According to the World Bank (1996), the average loan size from informal lenders varies between US$ 50 and US$ 500, with average maturity fluctuating between one and four months, and annual interest rate as high as 50% on small unsecured loans.

the region is active but with just an average of 2.8% of annual rotation of landholdings (Echeverria et. al. 1998);[29] unfortunately there is no such information in the case of Bolivia. Muñoz (1999), nonetheless, explains that the Bolivian land market is highly restricted due to lack of up-to-date titles, and by community rules, especially in the Andean region, and is highly segmented by size and social status. A great number of land transactions are not properly registered due to high costs, and to the lack of a valid proof of ownership (title). In general, land sales can be characterized into two groups.[30] The first takes place in the *Altiplano*, where the land sales occur between peasants belonging to the same community, or belonging to different communities but related by marriage or other close links, or between one member of the community and an outsider, with prior consent of the community.[31] The second group comprises the lowplains region where, in general terms, there exists better access to the land market, even for small settlers, with the main problem being the consolidation of the land rights.[32]

1.3.4 Income diversification approaches in rural Bolivia

The strategies that rural households use to overcome poverty are varied. As elsewhere, in Bolivia the more important ones are diversification out of farm activities and migration, both of which most of the time complement the income coming from farm work (UDAPE, 2004)[33]. Nevertheless, farm income remains the most important means of earning a living,

---

[29] The average has been calculated based on the data provided by Echeverria *et al.* (1998) for land markets in Colombia, Ecuador, Honduras and Venezuela.

[30] None of the studies provide comprehensive information regarding the land market functioning in the Valleys region; for most of it the information is definitely not available, except for some parts of the humid Valleys and *colonization* zones, where a dynamic market exists. The migrants' lack of ties to the foreign lands (*colonizadores*), and the higher quality of the land could explain the dynamism.

[31] World Bank (1996).

[32] Unfortunately, the information regarding the land tenancy market is not available; based on what is implied in the work on which we have relied for our previous explanation we can assume that it mainly follows the same functioning as the land sales market.

[33] UDAPE is the Government Unit of Economic Policy Analysis.

accounting for 52% of the income of rural households.[34] Even though the importance of farm income is unquestionable for all regions of the country, regional differences arise across the three agroclimatic regions. The percentage of the rural household's income coming from farm work is around 50% in the Altiplano and Central Valleys, and 64% for the Lowlands households.[35]

In general, nonfarm activities, and the income that rural households obtain from them, have been used to compensate low returns of farming, and/or to decrease the risk related with farm activities. However, small farmers frequently diversify out from the farm given that their land holdings contribute little to their income and to their own consumption. Also, market seasonality of farm operations and labor demand condemn them to seasonal unemployment, and hence to high levels of underemployment which lowers even more their labor productivity (World Bank 1996).

Turning now to migration and remittances, it has been shown that Bolivia is among the lowest urbanized countries in South America, i.e. the country's proportion of people living in rural areas is still high (36%)[36] despite its continuous decline since the middle 1980's, and it has been projected that it will continue to fall during the following decades.[37] Migration movements originated in the rural area represents an annual 30% of the total migration flows in the country[38], and among those, rural-urban migrants represent an

---

[34] Farm income includes agriculture, livestock and derived products.

[35] Jimenez and Lizarraga (2003).

[36] Behind Bolivia are Paraguay with a 59% and Guyana with a 28%.

[37] Since 1985 the decline in Bolivian rural population as a percentage of the total population has exceed the average percentage of the rest of countries in South America by more than 1%, and it is projected to remain at that level for the next four decades (UN 2008).

[38] The average includes migration flows in 1997 and 2002 from Tannuri-Pianto et. al. (2004). The other 70% is originated in urban and metropolitan areas.

average of 59%,[39] or approximately 26,700 households per year, a number that is expected to keep increasing as already explained.

Based on data provided by Andersen (2002) in his study on rural-urban migration in Bolivia, each of the three geographic regions in the country has a capital that attracts rural-urban migrants; La Paz (and El Alto) in the *Altiplano*, Cochabamba in the *Valles* (with Tarija growing fast due to its natural gas reserves), and Santa Cruz in the *Llanos* that has been growing faster than the rest of the departments during the last 50 years. As the author explains, the presence of different urban magnets for rural immigrants, imply that in Bolivia, as opposed to other developing countries, no city so far has reached mega dimensions. Based on Tam (1994), on average 48% of the population of the major cities destination for rural-urban migrants, are immigrants.

Hence, given the country's history of land reform, the prevalent distribution of landholdings, and the suitability of each region's agroclimatic conditions, the distribution of poor rural households across farm and nonfarm activities would most likely be affected by those elements, properly accounted for in the subsequent estimations and results explanation.

[39] The remaining 41.4% are rural-rural migrants. Ibid. pp. 5.

CHAPTER II - REMITTANCES AS AN INCOME DIVERSIFICATION STRATEGY FOR BOLIVIAN FARMERS

2.1 Introduction

According to the United Nations (UN 2008), in 2008, for the first time in world history, the world's total urban population would have reached the historic threshold of half of the global population, and it is expected to continue to increase. The proportion of the population living in urban areas in the less developed regions will reach this historic landmark around 2020 , and is likewise expected to increase. In more developed countries, the urban population is expected to grow more modestly relative to rural populations during the next decades. Hence, we can expect that the future urban population increase in the world will be primarily driven by the increasing percentage of people living in urban areas in the developing regions.

Latin America and the Caribbean shows an unusually high level of urbanization for its level of development (78% in 2007), considering that the average for the two other developing areas, Asia and Africa, is 40%. Among Latin America's three sub-regions, i.e. Caribbean, Central America and South America, the latter is the main contributor to that percentage, with a 79.5% of its population already living in the urban areas in 2007. However, South America embraces striking differences in patterns of urbanization. For example, in Argentina and Venezuela more than 90% of the population already lived in urban areas in 2005. In other countries such as Bolivia, however, the proportion of urban population is relatively low (64%) ; this suggests that the future increase in urbanization rates in the region will mainly come from these countries.

Migration from rural areas is among the most important determinants of the urban population growth especially in the developing countries (UN 2004). In Latin America, retention capacity of the population in rural areas remains low and its contribution to urban growth high. There are three sources of urbanization, e.g. natural increase, reclassification of rural settlements into cities and towns, and net rural-urban migration, the latter contributes, more than 40%, on average, to the urbanization of its countries in the region (CEPAL 2000).

In Bolivia the proportion of people living in rural areas is still high despite its continuous decline since the mid 1980's, and despite the downturn trend shown in the rest of South America; however, it has been projected that it will continue to fall during the following decades.[40] Migration movements originating in the rural area represent an annual 30% of the total migration flows in the country[41], and among those, rural-urban migrants represent an average of 59%,[42] or approximately 26,700 households per year. This number is expected to increase as already explained.

Todaro (1995) suggests several important questions in the study of migration that need to be addressed, especially when the focus is on rural-urban migration. First, why do people migrate from their home villages and what variables determine such decision? Second, how does migration affect the social and economic development of the source and the destination regions? This paper concentrates on the first of these questions.

---

[40] Since 1985 the decline in Bolivian rural population as a percentage of the total population has exceed the average percentage of the rest of countries in South America by more than 1%, and it is projected to remain at that level for the next four decades (UN 2008).

[41] The average includes migration flows in 1997 and 2002 from Tannuri-Pianto et. al. (2004). The other 70% is originated in urban and metropolitan areas.

[42] The remaining 41.4% are rural-rural migrants. Ibid. pp. 5.

It is widely accepted that economic considerations provide strong motivation to migrate.[43] The absence of crop insurance and shortage of liquidity are among the most important constraints that push rural families to look for diversification across alternative sources of income, which will secure a source of income not just for the migrants, but also for the family that stays behind in the village (Lucas 1997).

According to Taylor (1999), remittances represent the largest direct positive impact of migration on incomes and production of the rural families, in particular, and on migrant sending areas in general. Regmi and Tisdell (2002) explain that remittances are often the reason for the migration decision and its most important consequence. The impact of remittances on migration varies across countries, and may vary across regions within a country for several reasons. Impact depends on household characteristics; the functioning of the market in which migration and remittances decisions are taken; constraints faced by households; and, the tradition of migration/remittances-reception of the surrounding environment of the households.

Despite the fact that remittances might not be the dominant source of income among rural households[44], their impact on income on those that receive them is considerable; based on case-studies in some countries of the developing world, they represent up to 16% of the total household's income[45]. Unfortunately, no such information is available for Bolivia.

Among the large existing literature related to migration and remittances, two areas have been extensively studied. The first relates to the effects of remittances on poverty and inequality; the literature shows that remittances help reduce the incidence, depth and severity of poverty in

[43] Todaro (1995).
[44] In Burkina Faso, only 6.3% of the rural households in the sample received internal remittances (Wourtese, 2008), while in Guatemala the number was 14.6 % (Adams, 2004).
[45] In Burkina Faso, remittances as a share of total households income represent a 10.4% (Wourtese, 2008), Guatemala on the same indicator is 15.78% (Adams, 2004), and Egypt 15% (Adams 1991).

developing countries[46]; the evidence regarding the impact of remittances on inequality, however, has not been as conclusive[47]. The second area relates to the motives of migrants for sending remittances; they range from *pure altruism* (where the sole reason for the migrant for remitting is to support family consumption back in the hometown), to *pure self-interest* (where remittances are made for the aspiration to inherit or to invest in the rural town), or some combination of the two. Both areas have been applied to migration across countries, and to migration within countries. However, in both cases the analysis focused on the remitters' characteristics and their contextual settings. [48]

Adams *et. al.* (2008), using data from Ghana, takes a different approach. This paper analyzes remittances in a framework where the focus was precisely on the origin of the income flows rather than on the existence or not of migration assets in the rural household.

Finally, two are the main reasons that have been established to explain the presence of remittances among the income sources. Taylor (1999) and Lucas (1997) suggest that remittances represent an income diversification strategy. The World Bank (2006) posits that remittances ease working capital constraints, and hence represent a source of liquidity for the farmer. This has yet to be tested econometrically.

It is essential to understand all the roles that migration remittances play aside from increasing income and expenditure, or smoothing consumption. The contribution of remittances in terms of liquidity can represent additional benefits for poor farmers and their possibilities of overcoming market failures and increasing their risk-spreading strategies, strategies that have proved to help reduce poverty.

---

[46] See for example Adams *et.al* (2008) for Ghana, Wourtese (2008) for Burkina Faso, Taylor *et. al* (2005) for Mexico, and Adams (2004) for Guatemala.

[47] The World Bank (2006) explains that the effect depends on who receives the remittances (the better-off or the less well-off), if we are considering the effects in the short or long term, and other variables that affect their distribution.

[48] See for example Gopal and Tisdell (2002) for Nepal, Brown (1997) for the Pacific Island, Hoddinott (1992) for Kenya, and Lucas et. al. (1985) for Botswana.

This paper uses empirical analysis to formally test the econometric significance of migration remittances in further income diversification strategies, using a nationally representative dataset from Bolivia and its rural-urban migrants. If migration remittances are an effort by rural households to overcome market failures for credit and insurance, and hence represent a *means* to diversify their income sources into nonfarm activities, a positive effect is expected. The country is an interesting and relevant case study since its characteristics in terms of market failures, low urbanization rates, and increasing importance of nonfarm income strategies resemble those of many countries in each of the seven developing regions of the world. Valid inferences can be drawn from the study.

Also, this work includes a regional analysis in terms of the proposed hypothesis. The use of interaction effects will control for the likely existence of agroclimatic and idiosyncratic regional differences, and confirm or reject the proposed role of migration remittances at the regional level.

Unlike more traditional approaches to study migration and remittances, where the focus is on the migrants and their characteristics, this research work concentrates on the rural households from which the family members have been sent as migrants, their characteristics, and the contextual settings in which they make decisions. We follow in this respect the novel approach proposed by Adams *et al*. 2008.

Finally, among the variables that explain the existence of remittances we carefully incorporate the existence of migration networks in the social group to which each household belongs to. This has been found to have a significant effect in receiving remittances and has not always been considered in similar studies.[49]

[49] Lucas (1997) establishes that one important group of factors that influence migration and remittances decisions are the contextual setting, or general characteristics of the sending community.

The rest of the paper is organized as follows. In Section 2.2 the hypothesis regarding migration remittances and their role for nonfarm income diversification strategies is proposed. Section 2.3 presents the econometric model used to estimate the relationship between remittances and nonfarm income diversification. Section 2.4 outlines the database used for the modeling, and the variables involved in the estimation process. Section 2.5 presents the estimations' results and relevant marginal probability effects. Section 2.6 concludes.

2.2 The role of migration remittances in the farmers household's budget

Remittances can affect households' budget and wealth in different ways. Remittances directly increase the income of the rural household that receives them; hence they could help poor farmers to escape poverty. Remittances also contribute to smooth household consumption. They ease capital constraints, faced primarily by poor rural farmers. Finally, remittances increase household expenditures (World Bank, 2006). In this work, we focus our analysis in the role of migration remittances of providing working capital.

Rural farmers choose between different types of income, which can be earned singly (no diversification) or in various combinations (diversification). The alternative for rural residents are farm income, farm income and remittances, farm income and income from other nonfarm activities, or finally from all three sources of income simultaneously.

Before explaining the way in which we hypothesize how rural farmers decide on income diversification through remittances, let us briefly make two notes in regards to migration decisions and the receipt of remittances. First, not all migrants should be expected to migrate for reasons related to remittances. Even though an income diversification strategy through remittances may begin with the decision to send family members away as migrants (e.g. rural-urban migration), the decision of sending migrants may be motivated by other factors and necessities. Based on data provided by Andersen (2002) in his study on rural-urban migration in Bolivia, at most 18% of the migrants had remittances as the primary reason behind migration;[50] but of course, not everyone migrant remits. Therefore, from the standpoint of income diversification, the mere existence of remittances in the rural household is what matters, since this is consistent with the desire to diversify income.

Second, following Niimi and Özden (2008), we assume that if remittances are observed as part of the rural household's income, it is because the household sent at least one family member away as migrant. If it had not, no remittance would be observed. This rules out the possibility that the rural household might receive remittances even if no family members have migrated; we treat this option as unlikely or negligible.

We can now proceed with our proposition regarding the motivations and contextual conditionings involved in the process of income diversification through migration remittances.

A good characterization of the limiting conditions that rural farmers in developing countries face is the one developed by Lucas (1997), where the author states that agriculture is a high risk activity. Farmers face the prospects of floods, droughts, pests and cattle disease for which insurance rarely exists, or when it does exist, as Stark (1988) explains, the transaction costs may be prohibitive, especially for poor small farmers. Hence, insurance may be impossible to obtain. If so, rural farmers must look for a method of self-insurance, such as diversification across different sources of income. The object is to develop sources of income that are not positively correlated with farm income, e.g. sending family members as migrants to obtain remittances and/or diversify into other nonfarm activities.

[50] Among rural-urban migrants, 50% stated that family reunion was the reason for migration, 26% said education, 4% due to job moved, and 2% for health; the remaining 18% mentioned job search as the reason for migrating to the urban area.

Risks and lack of insurance are not, of course, the only limitation faced by rural farmers; lack of capital and imperfect or inexistent credit markets might also constrain them (Taylor 1999) in their desire of making farm and/or nonfarm investments (Taylor and Wyatt 1996).

Under such characterization of the limitations faced by rural farmers, and the self-insurance strategies that these agents follow to overcome them, a rural family may decide to diversify income[51] using an income source that counterbalances the lack of working capital due to the imperfect or inexistence credit markets.[52]Therefore, we hypothesize that initially rural farmers diversify (e.g. through nonfarm activities) as a self-insurance strategy against agricultural risks. However, if the farmer liquidity constrained and faces an imperfect or inexistent credit market, he/she first need to loosen this constraint in order to undertake any business venture off the farm. Therefore, remittances end up being a *means* for income diversification, through the provision of working capital, to those households that lack access to credit markets, or provide "cheaper" capital for those that can access the market but the costs are extremely high.

In concordance with our hypothesis, we explore the effect of remittances on the propensity of rural households to diversify income through nonfarm work, conditional on the characteristics of the households and on the environment in which they make their decisions. The estimation method used for this purpose is explained in the next section.

2.3 Econometric model and estimation

Given our hypothesis regarding the role of remittances in nonfarm income diversification strategies, the goal is to model a situation in which the decision to remit affects the decision of a

[51] Decisions on income diversification are now widely understood as a family strategy rather than a individualistic process, see for example Lucas (1997) and Taylor and Wyatt (1996).

[52] The thesis that remittances are a diversification strategy has been widely analyzed and explained; see for example Taylor (1999) and Lucas (1997).

household to diversify income through the nonfarm work. It is likely that unobserved factors affect both decisions for a typical farmer and that the decisions share many of these.

The bivariate probit model, an extension of the probit model, allows the presence of two binary (0,1) dependent variables correlated through their errors. Following Greene (2008), the outcomes of the two discrete choices can be viewed as a result of an underlying regression or index function, which captures the economic benefit calculation that leads to the decision of taking an action ($Y = 1$) or not ($Y = 0$).

Denote each equation with subscript i, and let x_i be vectors of observed *exogenous* variables that affect the utility of the decision-maker (e.g. household), ε_i be the error terms for the two possible discrete outcomes, which are jointly normally distributed with means zero, variances equal to one, and, ρ be the coefficient of correlation between the errors of the model. The bivariate probit model for each observation can be written as:

$$\begin{gathered} y_1^* = x_1'\beta_1 + \varepsilon_1, \qquad y_1 = 1 \text{ if } y_1^* > 0, 0 \text{ o.w.}, \\ y_2^* = x_2'\beta_2 + \varepsilon_2, \qquad y_2 = 1 \text{ if } y_2^* > 0, 0 \text{ o.w.}, \\ E[\varepsilon_1|x_1, x_2] = E[\varepsilon_2|x_1, x_2] = 0, \\ \text{Var}[\varepsilon_1|x_1, x_2] = \text{Var}[\varepsilon_2|x_1, x_2] = 1 \\ \text{Cov}[\varepsilon_1,\varepsilon_2|x_1, x_2] = \rho \end{gathered} \tag{2.1}$$

where y_1 represents the presence of nonfarm income for each household and y_2 the presence of remittances in the same household. For those rural farmers that we observe have diversified income through nonfarm activities, $y_1 = 1$, otherwise $y_1 = 0$, and for the ones that receive remittances $y_2 = 1$, otherwise $y_2 = 0$.

The model additionally assumes that the errors are uncorrelated across observations, i.e. ε_1 and ε_2 for each observation $t = 1,2, \dots, T$ are *iid* as pairs across the rural households, which means that the errors of the two types of outcomes for each household, i.e. diversification into nonfarm activities and reception of remittances, are the ones that are assumed to be correlated. This implies that diversification into nonfarm activities and reception of remittances, are correlated through their unobserved characteristics.

According to our proposition in Section 2.2, remittances (y_2) determine the presence of nonfarm income (y_1), making remittances endogenous in the nonfarm income equation; this means that the observed outcome for receiving remittances is correlated with the error term of the diversification outcome (ε_1). A model that accounts for this possibility for each observation in the sample is:

$$\begin{aligned} y_1^* &= \gamma y_2 + x_1'\beta_1 + \varepsilon_1, \qquad && y_1 = 1 \text{ if } y_1^* > 0, 0 \text{ o.w.,} \\ y_2^* &= x_2'\beta_2 + \varepsilon_2, && y_2 = 1 \text{ if } y_2^* > 0, 0 \text{ o.w.,} \end{aligned} \tag{2.2}$$

The system is fully recursive, as long as the conditional mean of y_2 does not depend on y_1, and consistent estimation of the parameters can be made using full information maximum likelihood (Greene, 1998). The statistical assumptions regarding the errors terms in (2.1) still hold.

The four possible mutually exclusive outcomes, for each point in the sample, from our two equations can be denoted by y_{10} (when $y_1 = 1$ and $y_2 = 0$), y_{01}, y_{11}, and y_{00}, and the probability for each of them equal to[53]:

$$\begin{aligned} P_{11} = \text{Prob}[y_1 = 1, y_2 = 1 | x_1, x_2] &= \text{Prob}[y_1{=}1 | y_2{=}1] \times \text{Prob}[y_2{=}1] \\ &= \{\Phi_2(y_1, y_2 = 1)/\text{Prob}[y_2 = 1]\} \times \text{Prob}[y_2 = 1] \\ P_{11} &= \Phi_2(\gamma + x_1'\beta_1, x_2'\beta_2, \rho) \\ P_{10} &= \Phi_2(\, x_1'\beta_1, -x_2'\beta_2, -\rho) \end{aligned} \tag{2.3}$$

[53] The bivariate normal probabilities for the four possible outcomes have been derived by Greene (2008), pp. 823.

$$P_{01} = \Phi_2(-\gamma - x_1'\beta_1, x_2'\beta_2, -\rho)$$

$$P_{00} = \Phi_2(-x_2'\beta_2, -x_1'\beta_1, \rho)$$

where, by convention, Φ_2 stands for the CDF of the bivariate normal distribution. The four expressions in Equation (2.7) are precisely the four terms that enter the log-likelihood function for the bivariate probit model whose set up is in Equation (2.1). Hence, in our recursive simultaneous-equation model we can proceed as if no endogeneity would occur.[54]

If we denote θ_1 as a vector of the right-hand side variables of equation y_1, and θ_2 the corresponding vector of right-hand side variables in equation y_2, the contribution made by each observation in the sample (i.e. each farmer) is the logarithm of the probability that the two dependent variables y_i $(i = 1,2)$ take on their observed value. The log-likelihood function of our recursive simultaneous-equation model, including all the observations in sample, is the sum of the individual contributions:

$$lnL_t = \sum_{i,j=0,1} d_{ijt} \ln P_{ijt} \, ({\theta_1}^*, {\theta_2}^*, \rho^*) \tag{2.4}$$

where d_{ij} is an indicator variable ($d_{ij} = I(y_i = i, y_j = j)$), equals one when its argument is true, and zero otherwise, with y_i and y_j representing the actual choices of individual t. In (6), equations y_1 and y_2 describe a system of equations for which the parameters are to be estimated simultaneously. The Full Information Maximum Likelihood (FIML) estimator is consistent and fully efficient for all the parameters in the model (Greene, 1998).

Finally, we need to ensure the identification of the model. According to Maddala (1983), the identification of the diversification equation requires that at least one variable from the remittances equation be excluded from the diversification equation. Wilde (2000), however, establishes that in the case of a bivariate model, the parameters are identified even if such

[54] As Greene op. explains, this conclusion comes from the fact that the marginal probability for y_2 in Equation (2.6) is equal to the univariate probit model for y_2.

exclusion restrictions do not exist. He explains that identification is simply feasible based on the presence of varying exogenous regressors. However, since identification based on exclusion restrictions is being found to be more robust (Yörük, 2009) we make sure to include those restrictions in the model.

2.4 Data

The data used in the present work comes from the database of the Program for the Improvement of Surveys and the Measurement of Living Conditions in Latin America and the Caribbean (MECOVI for its acronym in Spanish)[55], which is conducted by the INE (Bolivian Bureau of Census) in Bolivia, and provides access to it through its web page www.ine.gov.bo.

The MECOVI's have been conducted annually since 1999, and on-line information is available from the surveys conducted during 1999 through 2002. Since each survey does not track the same households, and some key questions are not asked in each survey, we perform a cross-section analysis for the survey of the year 2000. The stratified sampling procedure in the MECOVI 2000 is designed to eliminate sampling *bias* due to the household is included in the survey (Rivero and Mollinedo 2000).

The annual surveys collect data on such diverse topics such as income, expenditures, education, health, employment, food consumption, assets holdings and migration. It needs to be emphasized, though, that since the MECOVI surveys target variables related to the living conditions of the population and not migration and remittances specifically, they contain limited information on these topics. With respect to migration, the information about migrants is available at the households of destination, and not at the households of origin. This makes

[55] The Program is executed by the World Bank (IBRD), the Inter-American Development Bank (IDB) and the United Nations Economic Commission for Latin America and the Caribbean (ECLAC), as well as specialized institutions or agencies in countries participating in the Program. Subsequently other donors, such as Canada, Denmark, Germany, Japan, Norway, Sweden, UNDP, USA, and the Soros Foundation, have supported the Program.

impossible to know if the migration assets are held by the rural farmers, our target group.[56] What we have is information regarding remittances, domestic and international,[57] and whether the households receive them or not. If they do, the exact amount is provided. As Adams (2004) also establishes, it would be desirable to have information regarding the migrants of the rural household. However, having detailed information about the characteristics and the environment in which the farmers make decisions makes it possible to explore the role of remittances as a diversification income strategy.

The MECOVI survey for 2000 covers sample units from urban and rural households; for the present work we concentrate on the rural sample that comprises 2,108 households, or 9,092 persons, over 166 localities in the nine Departments of the country.[58] However, among the total rural households only 1,960 are in our final sample since our goal is to estimate the degree of income diversification of the rural *farmers* thorough remittances and/or other nonfarm activities.[59] Finally, the classification among the 1,960 farmers according to income source is shown in Table 1.[60]

[56] In relation to migration, the survey reports whether some member of a household is an immigrant, and if it is so, it provides information regarding the general location of the geographical area of origin (e.g. if the person has migrated from within the country they are asked the Department, Province and Municipality, and if they are from abroad, the country and city names). Unfortunately, this information is not enough to track the rural household from where the person would have migrated when the migration originated from another part of the country. The survey unfortunately does not ask if the households (rural or urban) have some members that have migrated and so are not present during the interview. Therefore, it is impossible to include information regarding the migrants that rural households might have sent, and then among those, to differentiate the ones that receive remittances.

[57] The term *domestic* applies to remittances that are sent from anywhere within the country, and *international* for those remittances that are sent from abroad.

[58] From the total 2,108 households, 825 are located in the *Altiplano* region, 756 in the *Valles* and, 527 in the *Llanos*. The survey also classifies the households according to their location into two groups: rural populated centers (212 households) and rural dispersed areas (1,896 households).

[59] Rural households, whose primary and/or secondary occupation is not related with farm work, have been dropped from the sample.

[60] We later drop 257 observations, and end up estimating the model with 1,703 households that have complete information.

Table 1. Rural households classification according to income source

Only farmers	1,086
Farmers receiving remittances	219
Farmers with nonfarm income	559
Farmers receiving both, remittances and income from nonfarm work	96

In the next sub-section, we specify the variables involved in the estimation of the bivariate probit model.

2.4.1 The independent variables

The variables included as explanatory are individual-specific and can broadly be classified within four groups. These include household characteristics (such as the age of the head and the number of adult and children members), household human capital assets (proxied by the head's education attainment and total number of years of education in the household), physical assets of the household (proxied by its landholdings), and the contextual characteristics of the surrounding environment where the rural household resides (proxied by some characteristics of the social networks of the household, and distance to the nearest capital of Department).

The logic behind those explanatory variables lies, in general, on the standard available literature on migration/remittances (Adams *et. al.* 2008) and diversification into nonfarm activities. In the case of the characteristics of the household, we can expect that if the altruistic motive is behind remittances (Lucas and Stark, 1985), households with older heads, fewer male members, and more children are more likely to receive remittances. Migrants can be thought to remit more if among those left behind are elderly, with little labor force, and many dependents[61]. With respect to human capital assets, and following Hoddinot (1992), we can expect that the higher the education attainment the better access to the formal sector of the

[61] The composition of the household variables considers all the members left behind, after some have migrated.

rural labor market, and hence the rural household is less likely to be liquidity constrained. Hence we expect a negative impact on the remittances outcome. We proxy human capital with two variables: the years of schooling of the head of the household, and the total years of schooling of the entire household. A squared specification is included to permit nonlinear returns to schooling.

Asset holdings, which are proxied by landholdings, have been extensively used to approximate the wealth of the household in the study of migration/remittances (Wourtese 2008), however based on contradicting arguments and empirical results, we cannot establish a priori its final effect on the remittances decision. For instance, a larger quantity of land per capita increases the marginal productivity from land and may reduce the pursuit of remittance income. However if the aspiration to inherit is an important reason to remit, as hypothesized (Lucas and Stark, 1985), the larger the potential of inheritance, the higher the probability of the rural household to receive remittances. We use landholdings per capita in the remittances equation to capture these effects and the direction of this relationship will be determined empirically.

It has also been widely explained and empirically demonstrated that social networks greatly affect the decisions made by entities such a household (see for example Taylor et. al. 2005). For sampling reasons, the INE divides the country geographically into UPM's (Primary Sample Units)[62], and each observation in the sample belongs to one of them. Hence, we use the UPM as the group of reference (*social networks*) for each rural farmer. In terms of the determinants of the propensity to receive remittances two variables are included in the model at the group level. First, we control for the number of households receiving remittances in the group of reference, and secondly, a group measure of wealth is included to capture indirect effects on migration remittances as suggested by Adams *et al.* (2008). These control variables are included as they

[62] The INE divides Bolivia into approximately 21,000 UPM, each bringing together an average of 50 housings. The master sample used for the survey in 2000 contains 2,500 UPM, and the final sample for the MECOVI 2000 encompasses 150 UPM. More details in http://www.eclac.cl/deype/mecovi/taller9.htm.

might influence the availability of obtaining information, the costs of transportation, and the costs of sending money. The variable used to control for the group's wealth is the squared income of the group; this specification will eventually reduce multicollinearity problems with the land asset variable.

Finally, there are also contextual characteristics that influence the decisions of the farmers and need to be incorporated as explanatory variables. As such, we control for the distance to the nearest department's capital, since it is the nearest economic and political center for the household where rural farmers send family members away in order to pursue remittances.

In terms of our nonfarm income equation, we include the remittances outcome variable, over which we want to determine whether it has the hypothesized positive and significant impact on the diversification decision. Additionally, we also include in this equation the regressors that control for the characteristics and composition of the household, its human and physical capital assets, and the characteristics of its surrounding environment. In terms of the first group of variables, we can expect that the younger the head, the higher the number of male adults, and the higher the number of dependents in the rural household, the higher the propensity to diversify into nonfarm activities. Regarding human capital, evidence shows that households with higher education levels engage more frequently in nonfarm activities, and that human capital has an important effect on the level of nonfarm income. Hence, the years of schooling of the head of the household, and the total years of schooling of the entire household, are both included in the diversification specification, along with a squared term to permit nonlinear returns to schooling. As we mentioned before, the land assets variable has usually been used to proxy wealth. However, among all the types of assets that the rural household owns, land is among the less liquid and hence might not help diversification in terms of providing the means to do it. As a consequence, we expect a negative impact on the diversification decision. Next, we also expect that social networks influence household's choices regarding diversification decisions. The higher the income of the group of reference the greater the access to resources from the network to start a business or other entrepreneurial activities. Finally, we also believe that the contextual characteristics of the environment where the rural household resides influence the decisions of the farmers regarding undertaking nonfarm activities. The distance to the nearest department's capital is included here as well, since the department capital is the nearest economic and political center for the rural household, where they might find nonfarm work.

Finally, dummy variables are conveniently included to capture the existence of possible regional effects. Two of them are used to discriminate among the three agroclimatic regions (i.e. *Altiplano*, *Valles*, and *Llanos*), and a third one to differentiate the regions where coca leaf production exits.

2.5 Estimation results

Table 2 includes some descriptive statistics of the variables used in the nonfarm income and remittances equations of our bivariate probit model. Recall that nonfarm income refers to whether a household diversifies farm income from nonfarm activities, and remittances in turn, to whether the household receives remittances. Since we want to contrast among households coming from the three different geographic regions, we present the data in this and the following tables, in regional terms.

With respect to human capital, Table 2 shows that households in the *Llanos* have more human capital than households living in the other two regions. It also shows that while households living in the *Llanos* have the highest mean land value, the asset is highly dispersed in all three regions. In terms of skewness, the asset is positively skewed in all three sub-samples, with the *Valles* region having a relative longer tail.[63]

In terms of our two binary outcomes, Table 3 provides a cross-tabulation statistics for each region. According to it, a smaller percentage of households in the *Llanos* receive remittances compared to the other two regions; however, it is precisely in that region where there is a higher portion of the households having income coming from nonfarm work, while the smallest percentage of households with nonfarm income among regions is in the *Altiplano*.

Table 2. Descriptive statistics, by region [(a)]

	Altiplano				*Valles*				*Llanos*			
Variable	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Remittances (1=yes; 0=no)	0.18	0.38	0	1	0.18	0.38	0	1	0.11	0.31	0	1
Nonfarm income (1=yes; 0=no)	0.25	0.43	0	1	0.38	0.49	0	1	0.40	0.49	0	1
Number of male adults	1.0	0.7	0	5	1.2	0.7	0	5	1.4	0.9	0	6
Number of children	1.0	1.3	0	6	1.0	1.2	0	5	1.3	1.3	0	6
Age of the household head	48.6	16.9	13	98	46.2	15.7	11	90	44.6	14.8	15	90
Household total years of schooling	11.6	10.5	0	64	12.7	10.8	0	68	16.3	12.1	0	76
Head's years of schooling	4.1	3.9	0	17	3.8	3.8	0	17	4.4	3.5	0	16
Land assets [(b)]	826.0	2,151	0	25,000	2,343	8,079	0	175,000	5,337	29,570	0	448,000
Land assets per capita	309.7	825.1	0	12,500	939.1	3,931	0	87,500	1,567	7,275	0	89,600
Number of households receiving remittances in the group of reference	2.6	2.2	0	11	2.5	2.4	0	9	1.6	1.8	0	9
Wealth of the group of reference [(c)]	4,706	3,915	546	20,926	7,657	5,847	586	30,385	12,285	7,514	3,093	37,123
Distance to the nearest's capital of Department [(d)]	69	37	16	193	76	44	7	197	125	76	13	470
N	798				714				448			

[(a)] *Source*: Own calculations based on MECOVI Survey for year 2000, Bolivian National Statistic Institute (INE).
[(b)(c)] Measured in US$.
[(d)] Measured in miles.

[63] The overall skewness value for land is 20.11. In regional terms, the measure is 7.27 for the *Altiplano*, 16.29 for the *Valles*, and 11.13 for the *Llanos*.

Table 3. Cross-tabulation of Nonfarm income Vs. Remittances, by region[a]

	Altiplano		
	NONFARM INCOME =0	NONFARM INCOME =1	Total
REMITTANCES =0	61%	21%	82%
REMITTANCES =1	14%	3%	18%
Total	75%	25%	100%
	Valles		
	NONFARM INCOME =0	NONFARM INCOME =1	Total
REMITTANCES =0	51%	32%	82%
REMITTANCES =1	11%	6%	18%
Total	62%	38%	100%
	Llanos		
	NONFARM INCOME =0	NONFARM INCOME =1	Total
REMITTANCES =0	54%	35%	89%
REMITTANCES =1	6%	5%	11%
Total	60%	40%	100%

[a] *Source*: Own calculations based on MECOVI Survey for year 2000, Bolivian National Statistic Institute (INE).

Households in the *Altiplano* and the *Valles* are more likely to receive remittances than households in the *Llanos*. Since these two regions are poorer compared to the third one, this could be a sign that altruism is a stronger motive for receiving remittance in those two regions, and that their constraints in terms of liquidity are more severe. Also, *Altiplano* farmers diversify their income less frequently compared to ones in the other two regions. This would suggest in turn a relatively lower willingness to do it, or lower means to do it. The results on the impact of remittances on the diversification decision will provide insights regarding these hypotheses. On the other hand, the statistics also show that income diversification is more widespread in the *Llanos*. This could be driven by a stronger willingness to undertake risk-spreading strategies in the region, and that households in that region have more successfully found ways to reduce liquidity constraints. The *Valles* region is more of a mixed story, with higher number of households receiving remittances, but a relatively high proportion of them involve in nonfarm employment. Here we also want to investigate and bring some insights regarding the role of remittances in the diversification strategies, in these high prone nonfarm employment regions.

The full information maximum likelihood estimates from the bivariate probit model are given in Table 4. As mentioned already, the estimator is consistent and fully efficient for all the parameters in the model. Each set of results, marked with a number at the top, show the estimates of the remittances outcome equation in the second column, and the estimates for the presence of nonfarm income in the first column. Column (1) includes, in the nonfarm income equation, the remittances discrete variable, all the household characteristics and human capital variables, along with a social network variable (*wealth of the group of reference*), and a contextual variable (*distance to the nearest capital of Department*). The specification of land in the diversification equation is the total value of the asset, and it acts as the exclusion restriction that helps identify the presence of the remittances equation. The remittances equation on the other hand, also includes the same household characteristics and human capital variables, as well as the same social network and contextual variables. Additionally, we also add the household's land assets per capita, and the number of households receiving remittances in the group of reference. These two variables are the exclusion restrictions that identify the nonfarm income equation.

Table 4. Bivariate probit model estimates

	(1)				(2)				(3)			
	Nonfarm Income		Remittances		Nonfarm Income		Remittances		Nonfarm Income		Remittances	
	Coefficien	*t*-ratios	Coefficien	*t*-ratios	Coefficien	*t*-ratios	Coefficien	*t*-ratios	Coefficien	*t*-ratios	Coefficien	*t*-ratios
Remittances (1=yes; 0=no)	0.46 **	2.03			0.45 **	1.96						
Age of the household head	-2.90E-03	-1.06	0.01 ***	3.91	-1.30E-03	-0.47	0.01 ***	3.96	-1.50E-03	-0.54	0.01 ***	3.95
Number of male adults	-0.04	-0.67	-0.33 ***	-4.60	-0.06	-1.09	-0.33 ***	-4.62	-0.06	-1.01	-0.34 ***	-4.63
Number of children	0.02	0.68	-0.02	-1.47	0.02	0.63	-0.06	-1.50	0.02	0.64	-0.06	-1.50
Head's years of schooling	0.04 ***	3.03	-0.05 ***	-3.02	0.05 ***	3.65	-0.05 ***	-2.86	0.05 ***	3.65	-0.05 ***	-2.86
Household total years of schooling	-0.03 ***	-3.14	-0.012	-1.02	-0.03 ***	-3.15	-0.132	-1.13	-0.03 ***	-3.21	-0.132	-1.13
Household total years of schooling squared	0.01 ***	5.48	0.002	1.49	0.01 ***	5.42	0.002	1.54	0.01 ***	5.46	0.002	1.53
Land assets	-2.68E-07 ***	-2.76			-3.00E-07 ***	-2.76			-6.00E-07 ***	-2.82		
Land assets per capita			1.60E-06	0.76			1.49E-06	0.69			1.49E-06	0.69
Number of households receiving remittances in the social group of reference			0.25 ***	12.95			0.26 ***	12.67			0.25 ***	12.63
Wealth of the social group of reference	3.50E-05 ***	6.77	6.21E-06	0.82	-1.10E-03 ***	4.55	3.46E-05	0.38	2.72E-06 ***	4.61	3.46E-05	0.39
Distance to the nearest's capital of Department	-4.40E-04	-0.73	5.24E-04	0.57	-0.001 *	-1.69	2.00E-04	0.27	-0.001	-1.60	2.00E-04	0.25
Coca production region (1=yes; 0=no)					0.01	0.08	-1.10E-03	-0.01	3.00E-03	0.03	-0.03	-0.01
Valles (1=yes; 0=no)					0.378 ***	4.43	0.05	0.49	-0.67 ***	-3.48	0.11	0.49
Altiplano (1=yes;0=no)									-1.04 ***	-5.53		
Llanos (1=yes; 0=no)					0.39 ***	3.65	0.14	1.01	-0.70 ***	-3.28	0.18	0.89
Remittances * *Valles*									0.44 *	1.86		
Remittances * *Llanos*									0.82 **	2.56		
Remittances * *Altiplano*									0.43	1.62		
Constant	-0.87 ***	-4.79	-1.88 ***	-7.79	-1.05 ***	-5.58	-1.89 ***	-7.69			-1.89 ***	-7.65
N	1703				1703				1703			
Log-likelihood	-1504.03				-1490.12				-1497.78			
LR chi2	584.8				612.6				695.0			
rho (ε_1, ε_2)	-0.33				-0.34				-0.37			
	(0.1273)				(0.13)				(0.13)			
Likelihood-ratio test of rho (ε_1, ε_2)=0	6.01				5.80				6.82			

Significance levels: *10%, **5%, ***1%

As we can see from this first set of results, remittances are highly and positively related to diversification. The head's level of schooling, household total years of schooling, and income of the reference group each have positive and significant impacts on the probability of income diversification. The results on the total years of education for the household are quite interesting. Given the signs of the coefficients, there exists a minimum education level (i.e. 1.5 years) above which the impact of education on the probability to diversify income is not only positive, but increasingly positive. Below such threshold, the impact is negative which can be interpreted as for those extremely low education levels, the returns to education from farm work and from nonfarm work are similar. The coefficient of the land assets variable is significant and with the expected sign, but since its magnitude is quite small, the impact of the variable is almost negligible.

In turn, the remittances decision is, as expected, positively affected by the age of the household's head and adversely impacted by the number of male adults. In terms of human capital, the head's number of years of education reduces the probability of receiving remittances. The more educated the head of the household, the less likely to receive remittances due to a loosening liquidity constrain. The number of dependents, the income of the reference group, the value of land assets per capita, and distance to the nearest capital of Department, have no effect on the propensity of remittances. Finally, the results show that migration networks are an important consideration for remittances. The higher the number of households in the social network who receive remittances, the higher the probability for each of them to receive remittances as well.

In terms of the correlation coefficient between the two structural disturbances, rho, it has an estimated value of -0.33. Given the standard error for the parameter (0.13), the Wald statistic for the hypothesis that $\rho = 0$ is 6.06 which is greater than the critical value for ${\chi^2}_{[1]}$. This result suggests the likely existence of unobservable characteristics of the households that influence both outcomes. An asymptotic similar test to the Wald statistic was also obtained automatically after the estimation of the model of Column (1).The value of a likelihood-ratio test of $\rho(\varepsilon_1, \varepsilon_2) = 0$ again confirms our earlier results. With a test statistic of 5.54, the null hypothesis is rejected at the 5% level of significance.[64] Therefore, the simultaneous estimation of both equations appears to be justified relative to the estimation of independent probit models. The bivariate probit is thus consistent and provides fully efficient estimates for our model (Greene 1998). The negative sign of the correlation coefficient implies that unobserved and/or unmeasured factors that increase the probability of receiving remittances also decrease the nonfarm income diversification propensity. Finally, as Greene and Seaks (1998) show, the results of the likelihood-ratio test can also be used, asymptotically, as a Hausman test for the exogeneity of the remittances discrete outcome in the diversification equation. This means that the correlation results also suggest that receiving remittances is, as expected, an endogenous variable in the model. Its positive and highly significant coefficient confirms the role of remittances in relaxing capital constraints that frequently prevent rural farmers from diversifying its sources of income.

The positive and significant effect of remittances on the nonfarm income diversification equation remains robust to different specifications of the model as it is shown in Table 3. Column (2) adds two dummy variables to the model to explore the possible existence of intercept regional differences. The first one includes three categories to distinguish among the three geographic regions, and the other aims to differentiate the regions where coca leaves are produced. The regional dummies for the *Valles* and *Llanos* regions are found to be significant for

[64] An extensive analysis of the different methods available to test the hypothesis $\rho = 0$, in simultaneous equations models, and involving limited dependent variables, can be found in Monfardini et. al. (2008).

the nonfarm income equation, (although not for remittances), but the one for the coca leaves production region is not significant for neither of the equations. [65]

Independently of the individual significance of the dummy regional variables, we also need to test whether they altogether significantly improve the prediction of our outcomes. This was done using the likelihood-ratio test of the null hypothesis
H_0: all regional dummy coefficients equal to zero. Since the test statistic value of 27.84 (p-value 0.000) was significantly larger compared to the relevant critical value, there was sufficient evidence of the joint significance of the regional dummy variable.[66]

Based on those results and on the rejection of the equality of the model parameters across regions for the income diversification equation, but not for the remittances equation,[67] the last set of results in Table 4, Column (3), includes interaction terms for the three regional dummy variables with the remittances discrete outcome in the nonfarm income equation. This last specification aims to provide more insights regarding the existence of regional effects in terms of the impact of remittances on the nonfarm income diversification decision. The resulting coefficients show that two out of three regional interaction coefficients are positive and significant at conventional levels. The coefficient for the *Altiplano* region is positive and nearly significant at 10%. In terms of magnitudes, remittances in the *Llanos* have a larger effect on diversification than they do in the *Valles* region.

---

[65] The *Valles* and *Llanos* regional dummies coefficients represent deviations with respect to the reference category *Altiplano*, which has been dropped as customary to avoid multicollinearity.

[66] We also run the test including a dummy variable for coca leaf production regions (1 if the household is located in a region where it has been established there is coca production and 0 otherwise); the results confirmed the appropriateness of the inclusion of such regional dummy variables in the model.

[67] We formally tested if the estimated coefficients for the three agroclimatic regions were equal to each other in terms of each of the two equations in the model. We used the SUEST test, a variation of the Hausman specification test that ensures that the test will be well defined (i.e. the estimator of the variance of the difference between the two estimators is guaranteed to be positive semidefinite). The test statistic has a Chi-square distribution with k degrees of freedom (White, 1982). The detailed results for the SUEST test are found in Appendix A.

Our battery of specification tests includes two final steps. First, we need to make sure that our interaction terms for the three regions are jointly significant, and hence improve the specification of the model. The Hausman test of the constraint that those coefficients are all zero has a chi-squared distribution with 3 degrees of freedom. The resulting statistic of 6.75 with a p-value of 0.07 allow us to reject the null hypothesis at the 10% level. Secondly, we want to know if our regional interaction terms' coefficients are all statistically equal to each other. Again for this purpose we use a Hausman test which in this case has a chi-squared distribution with 2 degrees of freedom. The test statistic of 2.47 with a p- value of 0.292, tells us that there is no enough evidence to reject the null hypothesis of equality. This of course could have been anticipated based on the almost identical coefficients of the *Altiplano* and *Valles* regions.

It is worth to note that among all the different specifications in Table 4, the control variables have always maintained their sign and significance level.

Before examining the magnitudes of the marginal effects, one more check of the specification is performed. A consistent estimation of the parameters depends on the exogeneity of diversification in the remittances equation. If it is endogenous then the system is not fully recursive and the estimates will not be meaningful. Although this proposition is not directly testable, the possibility that the causality should be reversed can be explored. To that effect the model is estimated treating the diversification as being determined by all of its factors, save remittances, and treating diversification as an endogenous determinant of remittances. This basically flips the model around; if diversification is insignificant in the remittances equation then we have indirect evidence that the two are not jointly determined.

Thus, another bivariate probit model is estimated, but this time with the nonfarm income variable as endogenous variable in the remittances equation. The estimated coefficient is -0.36

with a standard error of 1.03. These results are consistent with the original specification that remittances affect diversification and not vice versa.

As it is widely known, only the sign and statistical significance can be directly interpreted from the estimated coefficients of a binary choice model; as Winkelmann et.al. (2006) explain, the parameter estimates β_1, β_2 and γ in our model do not directly measure the marginal effect of each x_i on the binary response variables. The marginal probability effect is then what we need to estimate. Such effect of remittances on nonfarm income diversification was computed as the change in the diversification probability when the remittances dummy variable changes from zero to one, and all other variables are fixed at their means. The marginal effect for the *Valles* is 0.19, and for the *Llanos* 0.29, both coefficients highly significant; that is, the nonfarm income propensity of rural households that receive remittances is 19 and 29 percentage points, respectively, higher than that of non-receiving remittances households. The marginal effect for the *Altiplano* is relatively smaller, but statistically insignificant at conventional levels.

Table 5. Estimated Marginal Effects

	Nonfarm income equation		
Variable	Total effect	*Z*	Type of variable
Age of the household head	0.001	1.860	Continous
Number of households receiving remittances	0.017	11.430	Continous
Land assets	0.000	0.570	Continous
Number of male adults	-0.035	-2.840	Continous
Head's years of schooling	0.011	2.310	Continous
Number of children	-0.006	-1.090	Continous
Household total years of schooling	0.003	2.430	Continous
Wealth of the group of reference	0.000	2.730	Continous
Land assets per capita	0.000	-1.990	Continous
Distance to the nearest's capital of	0.000	-0.620	Continous
Coca production region (1=yes; 0=no)	0.348	1.420	Binary
Valles (1=yes; 0=no)	0.007	1.080	Binary
Llanos (1=yes; 0=no)	0.011	1.150	Binary
Remittances * *Valles*	0.191	3.560	Endogenous
Remittances * *Llanos*	0.291	2.970	Endogenous
Remittances * *Altiplano*	0.088	1.560	Endogenous

The estimated marginal effects of the variables in the nonfarm income diversification equation are given in Table 5. The derivation of the formulas for such effects of a binary variable, and that of a continuous variable, on a bivariate probit model are shown in detail in Greene (1998). Note here that the marginal effect of a variable in the nonfarm income diversification equation may be a sum of two terms. In the case in which a variable appears in both of our equations, one will be the direct effect of such variable on Equation (2.1), and the other one will be the indirect effect of the underlying variable on equation (2.1), through the second equation (i.e. age of the head of the household or number of kids in the household). A second case occurs when a variable appears just in the nonfarm income diversification equation, in which case the effect of the explanatory variable is direct (i.e. the endogenous variable remittances interacting with the regions, and land assets). We yet have a third possibility for other variables, such as land assets per capita, that only appears in the remittances equation, and which marginal effect on nonfarm income diversification decision will be indirect.

2.6 Conclusions

This paper uses a nationally-representative household survey to study the role of migration remittances in nonfarm income diversification strategies in Bolivia. There are three main findings.

First, and according to the literature, migration remittances represent a complementary income source for rural households since they provide them with liquidity. The calculation of the marginal probability effects show that, at the national level, nonfarm income propensity of rural households that receive remittances is 13 percentage points higher than that of non-receiving remittances households.

Second, the results also suggest that the variable remittances is, in fact, endogenous and with a significant effect on nonfarm income diversification; not taking this effect into account could result in biased and inconsistent estimates. The significance of the correlation coefficient justifies the use of the biprobit estimation model.

Third, accounting for the existence of significant regional differences, the paper finds that remittances has a substantial positive effect on nonfarm income diversification in the *Valles* (with a marginal effect of 0.19) and *Llanos* (with a marginal effect of 0.29) regions, but not in the *Altiplano*. These results are consistent with the profile outlined for the existence of two different types of rural farmers in the country. The small poor farmers, mainly located in the Andean region, are viewed as practicing subsistent farming, where the reception of remittances may have the sole objective of supporting consumption. In the other two regions, however, the existence of capitalist farming, oriented to the domestic as well as foreign markets, could help explain the stronger willingness among those farmers to undertake risk-spreading strategies. The imperfect and sometimes inexistent insurance and credit markets make them search for diversified income sources off the farm. Hence, remittances are used in those two regions as a source of liquidity that could help compensate the imperfect functioning of the capital and credit markets in the country.

This paper uncovers the significance of an additional positive role of migration remittances for poor farmers. Their presence help reduce liquidity constraints allowing rural households to increase their propensity to diversify income out from the farm using those resources. The regional differences found in terms of this role and their relation with different profiles of farmers, make this research work quite relevant for most developing regions.

Finally, our findings regarding the additional positive effect that migration remittances represent for small poor farmers clearly reinforce the pertinence of policies that facilitate these transfers in terms of transaction costs and the development and/or strengthening of a financial structure. It has been shown that major improvements can be done in these regards in most of the developing world.

CHAPTER III – LAND SIZE, DISTORTIONS, AND MARKET IMPERFECTIONS IN EXPLAINING THE EXISTENCE OF SUBSISTENCE FARMERS IN BOLIVIA

3.1 Introduction

With an annual US$ 1,363 of GDP per capita[68], Bolivia is among the poorest countries in Latin America, and *the* poorest compared with its proximate neighbors[69]. Poverty is particularly prevalent in rural areas where 80% of the population is poor compared with 53% of the urban areas. The distribution of the national income is highly unequal, especially in the rural area, with a rural Gini coefficient of 0.63.[70] There are also striking differences at the regional level, a GDP per capita of US$ 3,033 in the department of Tarija which is geographically located in the Valleys, compared with the US$ 1,972 for the department of Pando located in the Lowlands, or with the US$ 1,003 of the department of Potosi located in the Andean region.

Bolivia's share of the rural population in total population is still high (36%) compared to the average of its surrounding neighbors or to the average of Latin America (around 22% in both cases)[71]. However, and despite the fact that the country had one of the highest percentages of rural population at its Independence, the rate of decrease in its rural

---

[68] The Gross Domestic Product (GDP) numbers have been taken from INE (2008); they are preliminary data for the years 2006- 2007.

[69] Latin America includes 18 economies: Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, México, Nicaragua, Paraguay, Perú, República Bolivariana de Venezuela, and Uruguay. Bolivia is bordered by Argentina, Brazil, Chile, Peru and Paraguay.

[70] The country's urban Gini is 0.55.Data for poverty and inequality comes from CEPAL (2006).

[71] Ibid. CEPAL 2006.

population has been greater than the decrease in the rest of its neighboring countries, closing the gap between its rural participation ratio and the average of the region.

Among the population in the rural sector in 2006, 76% was considered poor according to the INE (Bolivian Bureau of Census), compared with 60% for the whole nation[72]. Despite the high incidence of poverty in the country, however, it has been decreasing slowly during the last decades, especially in the rural area.

The strategies that rural households have used to overcome poverty are varied. As elsewhere, in Bolivia the more important ones are diversification out of farm activities and migration, both of which most of the time complement the income coming from farm work (UDAPE, 2004)[73]. Nevertheless, farm income remains the most important means of earning a living, accounting for 52% of the income of rural households.[74] Even though the importance of farm income is unquestionable for all regions of the country, regional differences arise across the three agroclimatic regions. The percentage of the rural household's income coming from farm work is around 50% in the Altiplano and Central Valleys, and 64% for the Lowlands households.[75]

The most important factor of production and asset for rural farmers has traditionally been land. It contributes to wealth not just directly as an input of production or because it is a way of accumulation, but it also might allow its owners the possibility of access to capital (credit) when it can be used as collateral (especially in an environment where production risk cannot be insured), or it can entail some *status* among neighbors or relatives.

---

[72] Data source: INE's web page www.ine.gov.bo, downloaded October 2008.
[73] UDAPE is the State Unit of Economic Policy Analysis.
[74] Farm income includes agriculture, livestock and derived products.
[75] Jimenez and Lizarraga (2003).

The fact that land *is* a source of income for rural households has been widely shown in empirical studies across the developing world, where, e.g., per capita income was found to be positively correlated to the area of land owned, controlling for other possible determinants of income.[76] Therefore, land ownership is potentially a good instrument for reducing rural poverty, an assertion verified by different analysts.[77]

After more than two decades of intense theoretical and empirical analysis of the production relationships in agrarian economies such as those of Latin America, where land and market imperfections may explain land-size and productivity relationship, Bolivia presents a case study that is characteristic of those studies. The highly unequal bimodal land ownership structure present in Latin America is even more striking in Bolivia. According to the last Bolivian agricultural census,[78] the size of about 70 percent of the farms does not exceed five hectares. The small farms cover only 3.7% of the total farm area. On the other end of the spectrum, about 3.7% of the holdings occupy about 90% of the farm area.[79]

The country faces rural market failures and legal barriers that prevent small poor farmers from accessing credit. Using a nationally representative survey database conducted by the INE, and using nonparametric techniques, this work shows that, under certain conditions, small farmers have a low and negative marginal income value of land compared to large farmers. We model two types of agents who face different constraints and hence face different optimum conditions.

The rest of this essay is organized as follows. Section 3.2 provides a review of the relevant literature. Section 3.3 presents the theoretical framework from which the marginal income effects of land are derived for small and large farmers. Section 3.4 briefly explains the

[76] See for example Scott (2000) for Chile, or Gunning et al. (2000) for Zimbabwe.
[77] See for example Warriner (1969) and Dorner (1992).
[78] The last agricultural census in Bolivia was conducted in 1984, and there is no other source of up-to-date data.
[79] World Bank (1995).

econometric model used to estimate the non-linear relationship between land endowments and productivity. Section 3.5 describes the dataset and the specification of the model. Section 3.6 presents and discusses the results. Finally Section 3.7 concludes.

3.2 Literature Review

This section provides a review of the relevant theoretical and empirical literature in order to position the current work in the context of the existing research.

Feder (1985) explains that, based on the rural production patterns in the less developed countries, the relationship between farm size and its productivity can be direct, inverse, or do not vary systematically at all. The author assesses that the existence of distortions in the relevant rural markets i.e. (land, insurance, credit and labor) potentially reconcile the conflicting results.

Binswanger *et at.* (1995) demonstrate that distortions or *imperfections* in a single rural market, however, are not sufficient conditions for the existence of a systematic relationship between farm size and its productivity to appear.[80] The existence of distortions in at least two markets can introduce systematic relationship between the two variables. Thus, the presence of multiple market distortions could explain a variety of farm size distribution and productivity structures.

As it is widely known, the production relationships in agrarian economies of the developing world are characterized by the absence of certain markets, or by their imperfect existence (Stiglitz, 1986). Bardhan (1989) uses the economic theory of information to explain that capital, labor and land markets are all intrinsically imperfect in agrarian economies.

[80] The authors explain, for example, that if credit is rationed according to farm size, but the rest of the markets are not imperfect, land and labor market transactions will produce a farm structure that equalizes yields across farms of different size.

Asymmetric information, covariance risk and incentive problems account for the absence of the insurance market, as well as for the inability of the credit market to be a feasible substitute for it.[81] Labor markets, on the other hand, face imperfections due to asymmetric information and incentive problems, as well as due to market seasonality of farm operations and labor demand. These characteristics not only increase labor costs, but can also cause unemployment and underemployment. Finally, covariate risk, imperfect credit markets, and policy distortions can affect, in turn, the perfect functioning of land sale markets as well.[82]

Carter and Kalfayan (1989) analyze that, given the existing market imperfections in the developing world, the relationship between farm size and productivity can be represented by a U-shaped curve. Based on their analysis of a Latin American-type agrarian structure, they explain that the relationship between land size and productivity is based on the existence of a dualistic model of agriculture.[83] There is a small farm sector, also called *traditional*, that relies on family labor, faces unemployment and/or underemployment, but demands fewer resources in terms of supervision. The other sector (also called *modern*), is comprised by large farms that despite their advantages (e.g. lumpiness of management skills and machines, better access to credit, and other risk-diffusion measures) rely on hired labor that needs to be supervised to be efficient (i.e. it demands resources). Concisely, farm productivity initially declines with farm size mainly due to the constraints in access to credit. After certain land-endowment threshold the inverse relationship reverses just when credit becomes available for larger landholders.

---

[81] Binswanger and Rosenzweig (1986) explain that asymmetric information (e.g. between workers and employers, or between borrowers and lenders) and incentive problems (i.e. moral hazard, adverse selection and screening effects), in turn, are consequences of the risks faced by rural households, the costs of acquiring information, and of individuals behavioral characteristics.
[82] For more details see Binswanger and Rosenzweig (1986).
[83] The underlying assumption is an inactive land market. For more details see Carter and Wiebe (1990) p. 9.

Empirical evidence for this U-shaped relationship can be found for example in Carter (1984) for India and Carter and Wiebe (1990) for Kenya. More recent work again confirms the limited or negative effects of larger land endowments for small farmers. For example, using data from Pakistan McCulloch and Baulch (2000) find that the amount of rain-fed land has no effect at all on the real income of the rural households. Imai and Gaiha (2002) use data from South India to show that the proportion of poor households actually increases between landless and small farmers, and it only falls as it moves from middle to large farmers.[84] Gamba and Mghenyi (2005) for Kenya show that income per capita increases for three, out of the four landholding size quantiles, but it decreases for the second quantile compared to the smallest one. The rate of increase of income rises as one moves toward larger landholdings.

There is some empirical evidence, however, that acknowledge that the only relative advantage for small farmers is related to the productivity of land; the marginal contribution of the asset is higher for small farmers than for larger ones. López and Valdés (2000) and their co-authors present evidence from six case studies in Latin America,[85]but warn that the impact is limited in terms of the household's income.[86] These studies, or any other of which we are aware of, have not included Bolivia in its analysis.

The results found by López and Valdés, regarding the higher marginal value product of land for small farmers, have been tested by de Janvry *et. al*. (2005) with data from Mexico to show that, given the existence of market imperfections (*market failures*), a small amount of land can create large income gains for the poorest rural households: an additional hectare of

---

[84] Those poor for 6-7 years.

[85] The case studies were for Chile, Colombia, El Salvador, Honduras, Paraguay, and Peru.

[86] As the authors explain, the elasticities of land they find suggest that a 10% increase in land would raise income per capita by less than 1.5%. With those estimates, for the income contribution of land to be substantial (rise poor households income above the poverty line) the increase of the landholdings of the poorest should be more than sixfold in some cases.

land for the smallest farmers (land holdings less than one hectare) has been shown to have the largest marginal contribution to income among all landholdings, increasing their average income by more than two times.[87]The assumption behind those results is the increasing credit access as the land holdings of small farmers, increases.

Bolivia faces rural market imperfections and market failures as any other developing country. Small farmers face unemployment, high levels of underemployment, and a rural environment where capital markets are far from function properly, or more commonly do not exist. Small poor farmers bear not only highly unequal distributions of income and land ownership, but also confront legal barriers[88] that, coupled with the market barriers (i.e. incredible high interest rates)[89], ration then out completely from accessing credit. The large farmers, on the contrary, have plenty of access to the capital-rationed credit market, with the available capital being a function of their land holdings that are used as collateral. Under such conditions, we expect for Bolivian farmers a marginal wealth effect of land that has a "U" shape. As Carter and May (1999) explain, poverty results not only from having little land, as is the case for small farmers, but also from being constrained in the ability to use it effectively and gain a return from it.

Therefore, based on a nationally representative survey database conducted by the Bolivian Bureau of Statistics (INE), we propose to test that, under certain conditions, small farmers may not have positives returns from larger land endowments unlike large farmers may do.

---

[87] The different income gains among households according to their landholdings means that the marginal income effect of land (that measures the impact of a "unit" change in land over income) is different, favoring the smallest. In terms of elasticities the impact would be the same just in percentage units.

[88] According to the INRA Law (1996), still in force at the time of the survey, small farmers could not use their land as collateral, i.e. ownership of small farms cannot be seized. If they did it, they were going to be reclassified as a *medium farmer* that must pay taxes, and whose property (e.g. land holdings) can be seized.

[89] Microfinance institutions do offer access to credit to small borrowers without asking for any collateral. However, since micro loans are short-term are rather small the interest rates that are charged are quite high, which reflects the high operational costs of these institutions for collecting, and the riskiness of the loans. For more details see (Valdivia 2004). According to the World Bank (1996), the average loan size from informal lenders varies between US$ 50 and US$ 500, with average maturity fluctuating between one and four months, and annual interest rate as high as 50% on small unsecured loans.

In this research work we derive the theoretical land productivity expressions to support this hypothesis for the two types of farmers (i.e. small and large) conditioned on their respective constraints. We obtained the equations based on the General Equilibrium Model develop by Carter and Kalfayan (1989), and the comparative statics approach followed by de Janvry *et al*. (2005). We make use of the *slack* variables approach to solve the Kuhn-Tucker conditions involved.

In terms of the empirical method, this work uses nonparametric techniques to model the relationship between land holdings and its productivity, which have been shown to more accurately capture the nonlinearities of such relationship (de Janvry *et al.,* 2005). Using data from Mexico, these authors propose a semiparametric partial linear model (PLR), following Robinson's (1998) method, using LOWESS estimation, and a bandwidth of 0.8. Our paper takes several steps forward in the following directions. First, in terms of the estimation of the semiparametric PLR model, it acknowledges the existence of both continuous and discrete explanatory variables among the regressors, which urges the application of nonparametric estimation techniques that differentiate them while smoothing. Li and Racine (2007) propose an approach that uses specific kernel density estimators for each type of variable, providing gains in terms of efficiency compared to the widely used nonparametric frequency approaches. Second, our work refines the selection of the smoothing parameter, using least squares cross-validation (CV) proposed by Li and Racine (2004). As it is widely known, smoothing selection using data-driven methods minimize the risks of biasedness and oversmoothing, usually present when the smoothing parameter is chosen arbitrarily. It has been shown that the CV method is asymptotically equivalent with the also popular Kullback-Leibler cross-validation method (Hurvich *et al*., 1998). Third, the specification of the model is formally tested against the parametric alternatives to ensure its correctness. We use the Hiao *et al*. (2007) model specification method since it concedes the existence of a mix of continuous and categorical variables among the regressors, which provides power gains while testing. Four, the hypothesized nonparametric regressor in the PLR model is conveniently tested in terms of its significance using a test statistic develop by Racine (1997). Additionally, with the purpose of confirming our results in terms of their robustness, the paper additionally uses the Difference-based method for estimating the semiparametric PLR model. Yatchew (1997) proposes such approach.

Finally, a regional-type analysis is performed to inquire the possibility of regional differences in terms of the explanatory variables of wealth, and more importantly, in terms of the nonlinear wealth effect of land proposed.

3.3 Market imperfections and productivity: Theoretical framework

The existence of legal restrictions regarding land use as collateral, the imperfect and rationed capital market, and the existing labor market distortions (supervision costs and underemployment), are all characteristics of the environment in which farmers take decisions and eventually produce. They can be used to explain differences in productivity between the two broad types of rural producers that define the Bolivia's agriculture sector (small versus large farmers). We use the widely known theoretical framework of Carter and Kalfayan (1989): General Equilibrium Exploration of the Agrarian Question to build our theoretical model.

3.3.1 The small farmers

A household that cannot access the imperfect credit market (capital-constrained), needs to supply labor to the market in order to meet subsistence needs and obtain liquidity[90], and purchases inputs for farm production, maximizes an income equation that is given by: [91]

$$Y = Q\left(L_f, I, T; z\right) - p_I I + w\,\Omega(L_s) \tag{3.1}$$

Where $Q(\cdot)$ represents the production function, L_f is the household's time allocation to farm work, I are non-labor purchased inputs, T is land endowment, and z denotes household characteristics and factors external to the household that help explain income; p_I and w are the market non-labor inputs price and wage per unit of labor time, respectively.

Under seasonality of labor demand and unemployment, the employment specification is given by $\Omega(L_s)$, which is the number of days employed in nonfarm activities as a function of L_s, that represents the days supplied to the labor market, with $\Omega' > 0$ and $\Omega'' \leq 0$. In this context, $w\Omega'$ represents the expected earnings from a marginal unit of labor supplied to the market. Note that unemployment may push an individual to underemployment.[92] Additionally, we have that households have a time constraint given by $L_s + L_f = \bar{L}$.

The information above can be used to express working capital needs and income sources as follows:[93]

90 The model assumes that a household that supplies off-farm labor will not hire in labor.

91 Price of the famer's final product is set equal to 1.

92 According to Dooley and Prause (2004) underemployment or "inadequate employment" may result from the existence of unemployment which makes workers, with needs and responsibilities, take almost any job available in the market. Agricultural workers, i.e., currently underproductive due to highly seasonal agriculture work, may end up being wage employed or self-employed with economic inferior conditions compared to the peak-season jobs.

93 Recall that production is roundabout and small farmers must finance its subsistence costs and the costs of purchased inputs with wage earnings.

$$p_I I + S_0 \leq w\ \Omega(L_s) \tag{3.2}$$

where S_0 represents the family subsistence needs less savings. Therefore, Equation (3.2) makes it clear that small farmers' production costs (working capital) can be solely financed with labor market earnings.

To solve the optimization problem of these households we let the Lagrangean equation be specified by $\mathcal{L}$, which maximizes household's income subject to her/his constraints, and the relevant Lagrange multiplier denoted with the lowercase Greek letter μ, which represents the capital constraint multiplier. Appendix A shows the complete set up of the small famers, where the general framework is developed using the Kuhn-Tucker conditions, and the final expressions of the first-order conditions are derived. Note here that our model assumes that the small farmers face capital constrains, which joined with the interior solution assumption, implies that these agents face a binding capital constraint or $\mu > 0$.

Hence, the quantities of labor and non-labor inputs will be chosen to fulfill the next first order conditions that maximize the constrained household's income:

$$\begin{gathered} \frac{\partial \mathcal{L}}{\partial L_f} = q_l - w\Omega' + \mu w\Omega' = 0 \\ \frac{\partial \mathcal{L}}{\partial I} = q_I - p_I + p_I \mu = 0 \end{gathered} \tag{3.3}$$

According to Equation (3.3), inputs are used by these farmers until marginal factor productivities are equated to total economic costs. This means that family labor will be used until its marginal productivity (q_l) is equal to the expected earnings from a marginal unit of labor supplied to the market ($w\Omega'$), minus a margin given by the shadow price of the working capital ($\mu w\Omega'$). The latter reflects the scarcity of working capital faced by small farmers, which lowers their productivity.

The *value function* of our income equation (e.g. the objective function at the optimum choice variables) can be ultimately written as a function of prices, household's endowment of productive assets, and other household characteristics present in the production function that determine the return of the assets, that is:[94]

$$Y^* = Y(w, p_I, \bar{L}, S_0, T; z) \qquad (3.4)$$

If we differentiate Equation (3.4) to see the effect of an increase in landholdings on income, we get:[95]

$$\frac{\partial Y^*}{\partial T} = q_T + \left(q_{l_f} - w\Omega'\right)\frac{\partial L_f^*}{\partial T} + (q_I - p_I)\frac{\partial I^*}{\partial T} \qquad (3.5)$$

where L_f^* and I^* denote the optimum values of our choice variables L_f and I. De Janvry *et al.* (2005) explain that in the presence of market imperfections (which for the case of the capital market means $\mu > 0$), and when the landholdings of small farmers can be used as collateral for credit, the terms in parenthesis are positive, and are interpreted as an additional positive income effect of an increase in landholdings, due to a hypothetical reallocation of production inputs and the existence of labor unemployment and underemployment.[96]

Under the market imperfections and legal provisions framework, explained in Section 3.2, we do not expect such positive additional effect of land for Bolivian small famers, and rather, we propose that any increase in their land endowments is expected to have the usual positive direct effect on income, but also a negative effect, which is evident when we compare the first expression in Equation (3.3) and the first term in parenthesis in Equation

---

[94] For the purpose of the evaluation of the income impact of landholdings, we borrow the comparative statics analysis from de Janvry *et al.* (2005).

[95] The details of the differentiation procedure and how we obtain Equation (3.5) are in Appendix C.

[96] The authors explain that $\mu w\Omega'$ is the positive difference between the first two terms of the first first-order condition in Equation (3.3), aspect that would end up adding a positive additional effect on income when land assets increase, and μ is also assumed to be positive. The same logic applies to μp_i.

(3.5). Concisely, the total income effect of one unit increase in land assets for the small farmers is given by:

$$\frac{\partial Y^*}{\partial T} = q_T - \mu w \Omega' \frac{\partial L_f^*}{\partial T} - \mu p_I \frac{\partial I^*}{\partial T} \qquad (3.6)$$

The reasoning behind Equation (3.6) is as follows. First, as Kevane (1996) explains, small farmers with liquid assets that are not sufficient to meet subsistence needs (argument highly strengthen by the fact that these agents have no access to the imperfect capital market at all), will *have* to allocate some labor to earn wages, as already explained. But, as land endowments increase, it is expected that higher amounts of inputs are going to be needed; i.e. the farmer need to devote more family labor hours to the farm, which were initially supplied to nonfarm activities. As a result, part of the resources, e.g. working capital, provided by nonfarm activities would be lost. Hence, is not possible to say a priori if the increase in land assets will ultimately report any increase in income; we would expect that to happen if q_T is greater than $\mu w \Omega' \frac{\partial L_f^*}{\partial T} + \mu p_I \frac{\partial I^*}{\partial T}$.

Secondly, even when a marginal positive effect on income is realized, any increase in landholdings will rather tighten the working capital constraint since the demand for higher levels of inputs is not accompanied by a greater availability of credit or of any other additional source of income.

The marginal value of land will only start to increase, through the reallocation of production inputs, at rather larger landholdings, where the existing legal barriers do not prevent larger farmers from obtaining credit using their land endowments. As it was explained in Section 3.2, it is not just that a small farmer cannot use his/her land endowments as collateral, but also according to the Bolivian homestead and exempt property provisions[97], property cannot be seized in case of default up to fifty hectares. Considering that around 90% of the famers own less than fifty hectares, we can expect that any increase in the marginal value of land comes to landholdings sizes much higher than the land endowments of the poor farmers.

To sum up the argument, the proposed increase in the marginal product of land due to an increase in landholdings is not feasible for Bolivian poor farmers; rather, we expect a falling marginal income, since the hypothetical larger land endowments could not be used efficiently in production, given that they are not accompanied by increases in working capital, and hence no reallocation of family labor-time is possible.

3.3.2 The large farmers

The large capital-rationed farmers are hypothesized not to supply labor to market but rather hire labor, and choose labor and non-labor purchased inputs to maximize their income, can be characterized by the following income equation:

$$Q(L, I, T; z) - p_I I - wL_d - r\beta(T) \tag{3.7}$$

where, aside from the terms already explained, L_d represents hired labor, and r is the interest rate of the amount of credit borrowed, which is rationed by the size of the land assets of the farmer ($\beta(T)$). These large farmers are in turn also exposed to informational problems that give rise to incentive problems, e.g. the hired labor needs to be supervised to be incentivized to work hard (Binswanger and Rosenzweig, 1986). Based on what Carter and Kalfayan (1989) proposed for such an environment, we assume a formal hierarchical labor supervision technology according to which the total units of labor (L) needed in

[97] See World Bank (1996).

production come from household labor L_f, which is self-supervising and hence one unit of the input produces one unit of effective labor, and from hired labor L_d, that requires supervision. Concisely:

$$L = L_f - \delta(L_d) + \gamma_0 L_d \tag{3.8}$$

Under this mode of supervision, we assume a constant hired labor productivity given by $\gamma_0 < 1$, allowed by a dedicated supervisory effort given by $\delta(L_d)$, with $\delta' > 0$ and $\delta'' \leq 0$. Additionally, $\delta(0) > 0$ indicates the existence of fixed costs, which represent a cost advantage for large farmers as the average cost of labor decreases with the amount of labor hired. Finally, the large farmers are modeled to devote all his endowment of time ($\bar{L}$) to farm work, that is $L_f = \bar{L}$.

The most important assumption of the large famers' model lies on the fact that their landholdings allow them to access capital, being the amount of the credit a function of the size of the land they own. Therefore, working capital needs, e.g. labor and non-labor inputs costs, can be financed through credit.

The capital needs faced by these agents can formally be written as:

$$p_I I + wL_d \leq \beta(T) \tag{3.9}$$

where $\beta(T)$ is the amount of quantity-rationed working capital available to a household with land endowment of size T, which is used to finance inputs. Since these households are also hypothesized to be capital-constrained, which reminds us the existence of an imperfect capital market, we once again assume a binding capital constraint, which implies that is $\mu > 0$.

As we did in section 0, $\mathcal{L}$ also represents the Lagrangean equation that maximizes the household's total income subject to her/his constraints, and μ represents the Lagrange multiplier for the capital constraint, or its shadow price, theorized to be positive according to the assumptions of a binding capital constraint and a interior solution. As in the case of the small farmers, the complete set up of the maximization problem of these farmers is developed in Appendix C.

The maximization problem for these agents with labor and non-labor inputs as choice variables leaves us the following first first-order conditions:

$$\frac{\partial \mathcal{L}}{\partial L_d} = q_l(\gamma_0 - \delta') - w(1 + r + \mu) = 0$$
$$\frac{\partial \mathcal{L}}{\partial I} = q_I - p_I(1 + r + \mu) = 0 \quad (3.10)$$

The interpretation of the equations is as follows: inputs are used until the marginal factor productivities are equated to their real economic costs, which contains two parts. For the labor input, for example, we have the direct cost that appears as an efficiency wage, or the market wage paid per unit of effective labor input, plus the borrowing price r, minus the shadow price of working capital ($\mu > 0$).[98]

The value function for the income can ultimately be written as a function of prices and household's endowment of productive assets, and various household characteristics that affect the return of the assets, that is:

$$Y^* = Y(w, p_I, r, \overline{L}, T; z) \quad (3.11)$$

[98] The efficiency wage is the market wage marked up by the difference between the constant hired labor productivity and the increase in the supervisory cost per a unit increase in the hired labor. Note that the greater the difference between them, the smallest the increase of the efficiency wage.

Since our aim is to find the income impact of an increase in landholdings, we need to differentiate Equation (3.11) with respect to land; the procedure leads us to the following marginal income effect of a one unit increase in land endowments:[99]

$$\frac{\mathrm{d}Y^*}{\mathrm{d}T} = q_T + \left(q_l(\gamma_0 - \delta') - w(1+r)\right)\frac{\partial L_d^*}{\partial T} + (q_I - p_I(1+r))\frac{\partial I^*}{\partial T} \tag{3.12}$$

We have already said that in the presence of perfect markets, the Lagrange multiplier μ would be equal to zero, which together with our FOC in Equation (3.10), would lead us to set the second and third terms in Equation (3.12) equal to zero (e.g. the expressions within the parenthesis); the marginal contribution of land to income, therefore, would be exactly equal to its marginal productivity.[100] However, since the markets are not perfect, supervision costs and access to rationing working capital would give to larger farmers the opportunity to obtain not just a direct positive effect on income, but also indirect benefits from the hypothetical increase in their landholdings. These results can be visualized in Equation (3.13), where we compare the expressions in Equation (3.10) and the terms within the parenthesis in Equation (3.13), obtaining the following results:

$$\frac{\mathrm{d}Y^*}{\mathrm{d}T} = q_T + \mu w\frac{\mathrm{d}L_d^*}{\mathrm{d}T} + \mu p_I\frac{\mathrm{d}I^*}{\mathrm{d}T} \tag{3.13}$$

These results make it clear that the access to working capital allows larger farmers to increase their operations and the usage of variables inputs, which has an indirect additional positive effect on income beyond the direct marginal product effect of land (q_T). As land endowment increases these agents are able to access more credit and thus increase their working capital, which will enable them to efficiently use their larger landholdings by increasing the usage of inputs of production. This increase will, however, lead to an increase

[99] The details of the differentiation procedure and how we obtain Equation (3.11) are in Appendix C.
[100] See that under perfect conditions, hired labor would be as good as family labor in terms of productivity.

in costs, but since these agents enjoy cost advantages due to fixed supervision costs, both effects could counter balance each other. Therefore, the final effect would give these agents the opportunity of increasing operations and enjoy increased levels of productivity as land endowments increase.

3.4 Econometric model

Based on the characterization of the two agents and the conditioning environment in which they make decisions, we need to estimate a model in which the relationship between land and income might be nonlinear, i.e. the impact of a unit change in the land variable is hypothesized to vary over the size of the holdings owned by rural farmers. Therefore, instead of assuming a functional form of the income regression, which depends on variables such as the household's assets, its demographics, and on the contextual characteristics of its proximate environment, this study will estimate a flexible form that permits marginal land effects to vary. To do this, *local* estimates or nonparametric regression models are applied, as they will allow us to gain a more representative estimate of the regression function (Keele, 2008).[101]

Nonparametric regression techniques basically use the data to estimate the value of the regression function at a given point using neighboring observations. Their flexibility in regards of functional form, though, is greatly offset by the decrease in precision and usefulness when several explanatory variables are included nonparametrically in the regression (Härdle *et al.* 2004). As Keele (2008) explains, there are two reasons why these models are less useful in such situations. The first is interpretation, plots of more than two explanatory variables are not very revealing. The second is the *curse of dimensionality*

[101] Global estimates are those models in which the analysis assumes a functional form for the model. Local estimation, or nonparametric regression models, describes the statistical dependency between two variables not with a single parameter but with a series of local estimates, (Keele, 2008).

problem, where data becomes too sparse, and hence the span must be increased to a point where the local estimates turn to be highly variable and less local.[102]

Therefore, the nature of the main relationship considered here suggests that one needs to *estimate*, rather to impose, the functional form of the relationship between land holdings and income. However, there are other explanatory variables that affect income but have been found to have a linear effect, and hence can enter the model in a parametric fashion.

Hence, the model should include both types of explanatory variables, which is precisely what semiparametric estimation models do. If it is additionally assumed that the nonparametric and parametric terms can be separately added up, the income equation for each household can be expressed as:

$$y_i = f(t_i) + x_{1i}\beta_1 + \cdots + x_{ki}\beta_k + \varepsilon_i \quad i = 1, \ldots, n \tag{3.14}$$

where y denotes a measure for wealth, t stands for the nonparametric variable in the model, i.e. land assets, and f is a smooth function whose density needs to be estimated from the data using a smoother. $\beta_1, \ldots, \beta_k$ are the model parameters, and ε_i are the error terms assumed to be independent of the explanatory variables $x_1, \ldots, x_k$ and t, with mean zero and variance σ^2.[103] Note that the unique assumption regarding f is that it is a smooth function.

Adding parametric components to the nonparametric as in Equation (3.14), creates the *Semiparametric Partially Linear (PLR)* model, whose estimation is outlined in the following paragraphs.

---

[102] For nonparametric regression estimation the analysis needs to construct a series of intervals or bins, which widths need to be determined, and within each bin the dependent variable y is estimated. The local estimate is the predicted value of y at a focal point x_0. Joining the predicted values produces the nonparametric estimates. When using nonparametric regression models, the choice of the bandwidth is the most critical modeling choice to take, and when specified as the proportion of the observations that will be included in the window (number of observations used in the local estimation), it is called the span. *Ibid.* pp: 28-29, (Keele, 2008).

[103] Davidson and Mackinnon (2004).

The partially linear model proposed by Robinson's (1988) requires five steps to obtain $\hat{f}$. First estimate the linear coefficients $\hat{\beta}_1, \dots, \hat{\beta}_k$. The conditional expectation of Y on the explanatory variable t for each household i in Equation (3.14) is:[104]

$$\mathrm{E}(y_i|t_i) = f(t_i) + \mathrm{E}(x_{1i}|t_i)\beta_1 + \cdots + \mathrm{E}(x_{ki}|t_i)\beta_k \tag{3.15}$$

Second, subtract Equation (3.15) from Equation (3.14) gives:

$$y_i - \mathrm{E}(y_i|t_i) = [x_{1i} - \mathrm{E}(x_{1i}|t_i)]\beta_1 + \cdots + [x_{ki} - \mathrm{E}(x_{ki}|t_i)]\beta_k + \varepsilon_i \tag{3.16}$$

Note here that the set of explanatory variables that enters the model in an ordinary linear fashion cannot contain a constant term. The subtraction removed the nonparametric element from Equation (3.15), leaving a simpler task to tackle. Third, once the estimates for the conditional expectations $\mathrm{E}(y_i|t_i)$ and $\mathrm{E}(x_{1i}|t_i), \dots, \mathrm{E}(x_{ki}|t_i)$ are obtained through nonparametric techniques, Equation (3.16) becomes:[105]

$$y_i - \hat{y}_i = (x_{1i} - \hat{x}_{1i})\beta_1 + \cdots + (x_{ki} - \hat{x}_{ki})\beta_k + \varepsilon_i \tag{3.17}$$

which is just a linear regression model with parameters $\beta_1, \dots, \beta_k$ that can be consistently estimated by ordinary least squares (OLS). Fourth, the unknowns in Equation (3.14) can be replaced by $\hat{\beta}_1, \dots, \hat{\beta}_k$, and obtain:[106]

$$y_i - x_{1i}\hat{\beta}_1 - \cdots - x_{ki}\hat{\beta}_k = \hat{y}_i = f(t_i) + \varepsilon_i \tag{3.18}$$

Finally, estimate Equation (3.18) using nonparametric regression to obtain $\hat{f}(t_i)$ using a suitable smoothing procedure.

---

[104] It is assume that $E(\varepsilon_i|t_i) = 0$.

[105] With a suitable choice of bandwidth h, $\hat{y}_i$ and $\hat{x}_{1i}, \dots, \hat{x}_{ki}$ are consistent estimators such that $y_i - \hat{y}_i$ is asymptotically equal to $y_i - \mathrm{E}(y_i|t_i)$ and $x_{1i} - \hat{x}_{1i}, \dots, x_{ki} - \hat{x}_{ki}$ are asymptotically equal to $x_{1i} - \mathrm{E}(x_{1i}|t_i), \dots, x_{ki} - \mathrm{E}(x_{ki}|t_i)$. See details in Davidson and Mackinnon (2004), pp: 689-690.

[106] It has been shown that $\hat{\beta}_1, \dots, \hat{\beta}_k$ are $\sqrt{n}$-consistent for $\beta_1, \dots, \beta_k$ and asymptotically normal, and that there exists a consistent estimator of its limiting covariance matrix. For further details see Robinson (1988).

The estimation procedure just explained requires the application of nonparametric techniques twice, and a simple and widely used approach to obtain them is Kernel regression. In terms of Equation (3.18), $\hat{f}(t)$ is obtained by regressing $\hat{y}_i$ on t_i, where:[107]

$$\hat{f}(t) \equiv E(\hat{y}_i|t_i) = \frac{\sum_{j=1}^{n} y_i k_{ij}(\frac{t_i - t_j}{h})}{\sum_{j=1}^{n} k_{ij}(\frac{t_i - t_j}{h})} \tag{3.19}$$

which is the well-known Nadaraya-Watson estimator.[108] k (*kernel function*) is the weighting function for the continuous variable t, which provides weights for all the observations included in the local estimation; it typically integrates to one and is symmetric around zero.[109] The parameter h in Equation (3.19) is the *bandwidth* that determines the number of the neighboring observations to use in the local estimation. As it controls the locality of the estimation, it is highly important in the estimation process, and hence crucial to choose it using methods that minimize the risks of biasedness and oversmoothing. One of such data-based methods is the popular least squares cross-validation (CV) proposed by Li and Racine (2004). According to it, the optimum bandwidth value to use in Equation (3.19) is the value of h that minimizes:

$$CV(h) = n^{-1} \sum_{i=1}^{n} [y_i - \hat{f}_{-i}(t_i)]^2 \tag{3.20}$$

where $\hat{f}_{-i}(t_i)$ is the *leave one out* nonparametric estimator of $f(t_i)$, that is, it excludes observation i when the prediction for it is being compute.[110]

[107] The same procedure applies for $\mathrm{E}(y_i|t_i)$ and $\mathrm{E}(x_{1i}|t_i), \dots, \mathrm{E}(x_{ki}|t_i)$, except for the type of kernel function, which we explain shortly.

[108] Nadaraya (1965), and Watson (1964).

[109] The kernel function used for the continuous land variable is the Second-order Gaussian.

[110] According to J. S. Racine (2008), the CV criterion performs reasonably well and provides the best fit in a variety of settings.

Robinson's method outlined above also demands the estimation of the conditional expectations $E(x_{1i}|t_i), \dots, E(x_{ki}|t_i)$. According to Racine and Liu (2007), the variability in the estimates of those expectations will be reduced if appropriates kernel functions are applied depending on the type of variable involved. For instance, the authors use Aitchison and Aitken's (1976) kernel function for unordered discrete variables and Wang and Van Ryzin's (1981) kernel for ordered discrete variables.

3.5 Data

The data used in the present work come from the database of the Program for the Improvement of Surveys and the Measurement of Living Conditions in Latin America and the Caribbean (MECOVI for its acronym in Spanish)[111], which is conducted by the INE (Bolivian Bureau of Census), and provides online access. It collects data on such diverse topics as income, expenditures, education, health, employment, food consumption, assets holdings and migration.

The MECOVI surveys have been conducted annually since 1999, and online information is available from the surveys conducted during 1999 through 2002. Since each survey does not track the same households, and some key questions are not asked in each survey, we perform a cross-section analysis for the survey of the year 2000.

The MECOVI survey covers sample units from urban and rural households. In this chapter the focus is on the rural sample that comprises 2,108 households, or 9,092 persons, over

[111] The Program is executed by the World Bank (IBRD), the Inter-American Development Bank (IDB) and the United Nations Economic Commission for Latin America and the Caribbean (ECLAC), as well as specialized institutions or agencies in countries participating in the Program. Subsequently other donors, such as Canada, Denmark, Germany, Japan, Norway, Sweden, UNDP, USA, and the Soros Foundation, have supported the Program

166 localities in the nine Departments of the country. [112] Among them, we are interested on those that are farmers and for which we can measure the marginal effect of land. In terms of the geographic regions there are 588 households from the *Altiplano*, 484 from the *Valles*, and 302 from the *Llanos*, all with complete information, making a final sample size of 1,374.

The sampling procedure in the MECOVI 2000 combines stratification by agglomerate population size and a two-stage sampling. By using these sampling methods, the survey is designed to reduce *bias* that can be introduced in regards of what household is included in the survey. Additionally, the stratification of the population into groups ensures that the sample is representative (INE 2000).

Next, we specify the variables involved in the estimation of the income equation outlined in Section 3.3.

3.5.1 The dependent variable

The variable chosen to be the dependent in Equation (3.14) is consumption expenditure. This is a broad measure of the household wellbeing, over which we want to study the impact of larger land asset among rural farmers.

The reasons for preferring consumption over income as a proxy for well-being are widely explained elsewhere (e.g. Ravallion 2002, and Coudouel *et al*. 2002). The most relevant are: 1) it is more closely related to the concept of having enough to meet current basic needs, 2) it is believed to be better measured than income considering that in poor agrarian economies income flows are cyclical and erratic, which may hinder the provision of income

[112] From the total 2,108 households, 825 are located in the Andean region, 756 in the Valleys and, 527 in the Lowlands. The survey also classifies the households according to their location into two groups: rural populated centers (212 households) and rural dispersed areas (1896 households).

information during a household survey, and 3) consumption better reflects the ability of households to meet basic needs through savings and credit.

The database use for this work provides detailed information regarding consumption expenditure, with the basic unit of observation being precisely the household.

3.5.2 The independent variables

The explanatory variables incorporated in the welfare equation specification can be grouped into four categories: 1) human capital, specified as the education level of the household's head and the total number of years of education in the household[113], 2) household characteristics such as the number of adults and the number of children, 3) household physical assets such as land and livestock, and 4) a variable that controls for the contextual characteristics of the village where the household resides, e.g. distance to the nearest capital of Department.

Regarding land assets, which marginal welfare effect is believed to be nonlinear, the 2000 survey includes the value of the asset owned by the household. This is the semiparametric variable included in the specification of Equation (3.14).

Given the striking regional differences in the country, a set of regional dummy variables is constructed, with three categories to differentiate among the three agroclimatic regions. Also, given its relevance in the agricultural GDP, another dummy variable is included allow differences between regions where coca leaves are produced and those where they are not.[114]

---

[113] The possibility of correlation between these two variables is formally tested during the estimation process.

[114] The areas where coca plantations are concentrated are the tropical valleys of the Departments of Cochabamba (Central Valleys) and La Paz (Altiplano). Specifically, coca plantations are cultivated in the Chapare and Yungas regions.

3.6 Estimation Results

Table 6 lists the variables we use in the estimation of Equation (3.14), along with their means, standard deviations, maximum and minimum values, for the entire or national sample, as well as for the three contrasting geographic regions of the country. As it shows, consumption expenditure, the variable used as the dependent in the semiparametric partial linear model (semiparametric PLR), varies considerably among regions in a range that goes from US$. 65.2 to US$. 155.3; e.g. the average value for the Highlands is 40 percent and 60 percent smaller compared to the average consumption expenditure in the Valley and Lowplains regions, respectively. Note that the standard deviation for these two regions is significantly larger as well.

Concerning human capital variables, there are no significant regional differences in terms of the average of the household's head years of education, which its overall level is four years. However, for the second measure of human capital, i.e. total number of years of schooling in the household, a typical family in the *Llanos* has between three and four years more schooling in average than their peers in the other two regions, *Valles* and *Altiplano*.

Table 6. Descriptive statistics, national and by region[(a)]

	National				Highland				Valley				Lowplains			
Variable	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Consumption expenditure[(b)]	101.4	95.8	3.7	1380.1	65.2	52.2	3.7	335.6	108.0	93.9	6.0	1303.4	155.3	126.3	11.3	1380.1
Head's years of schooling	4.1	3.8	0	17	4.1	3.9	0	17	3.8	3.8	0	17	4.4	3.5	0	16
Household's total years of schooling	13.1	11.2	0	76	11.6	10.5	0	64	12.7	10.8	0	68	16.3	12.1	0	76
Number of adults	3.3	1.8	1	9	3.0	1.7	1	9	3.3	1.7	1	9	3.7	1.9	1	9
Number of children	1.1	1.3	0	6	1.0	1.3	0	6	1.0	1.2	0	5	1.3	1.3	0	6
Land assets[(c)]	2,459.4	15,414.7	0.0	448,000	826.0	2,151.2	0.0	25,000	2,343.4	8,078.5	0.0	175,000	5,337.3	29,570.2	0.0	448,000
Livestock assets[(d)]	40.0	54.9	0.0	1,020	38.3	43.6	0.0	352	37.7	66.4	0.0	1,020	47.1	54.2	0.0	328
Distance to the nearest's capital of Department[(e)]	84.0	55.1	7.5	469.7	68.7	37.0	16.3	192.8	76.5	43.7	7.5	196.8	125.4	76.3	12.7	469.7
N	1,373				592				484				298			

[(a)] *Source*: Own calculations based on MECOVI Survey for year 2000, Bolivian National Statistic Institute (INE).
[(b)(c)] Measured in US$.
[(d)] Total units of farm animals.
[(e)] Measured in miles.

With respect to land assets, i.e. the variable that enters the model in Equation (3.14) in a nonparametric fashion, the statistics show that rural households living in the *Llanos* have the highest mean value, which is 2.3 times higher than in the *Valles*, and 6.5 times higher compared to the ones living in the *Altiplano* region. The standard deviation of this variable in the three regions is high, especially in the *Llanos*.

Finally, with regard to the livestock variable, the *Llanos* region again presents the highest average value, with the dispersion of the variable being higher in the *Valles*. There are no significant regional differences in terms of the average number of adults or number of children living in the rural households.

The analysis of the relationship between land and consumption begins with an exploratory plotting of a nonparametric estimate of the regression function $y = f(t) + \varepsilon$, as well as the estimates of a pure parametric quadratic specification for land.[115] As Figure 1 shows, the kernel estimate greatly diverges from the quadratic estimate at the left tail of the graph, where land assets are approximately smaller than US$ 2,000. Thereafter, both estimates behave similarly and in parallel. Whereas more than 80 percent of the sample has the value of its land assets equal to or less than US$2000, these preliminary results are quite relevant and show that the effect of land on consumption expenditure is evidently nonlinear. The p-value from a test of significance of the continuous regressor *land* indicates that this variable is highly significant at all the conventional levels (< 0.001).[116]

[115] The nonparametric routines are kernel-based, and the method used to select the bandwidths is the least-squares cross-validation (Hayfield and Racine, 2008).

[116] The approach used by Racine (2008) is to find out whether a regressor is irrelevant or redundant. The computation procedure for the statistic is different depending on the type of the variable involved, i.e. categorical or continuous. The null hypotheses for a continuous variable is $E(y|x,z) = E(Y|z)$, where y is the dependent variable, x is the continuous regressor that might be irrelevant (i.e. land), and Z is a vector of the remaining explanatory variables. The test statistic denoted with I, tends to zero in probability under H_0, and bootstrap procedures are used to obtain its distribution.

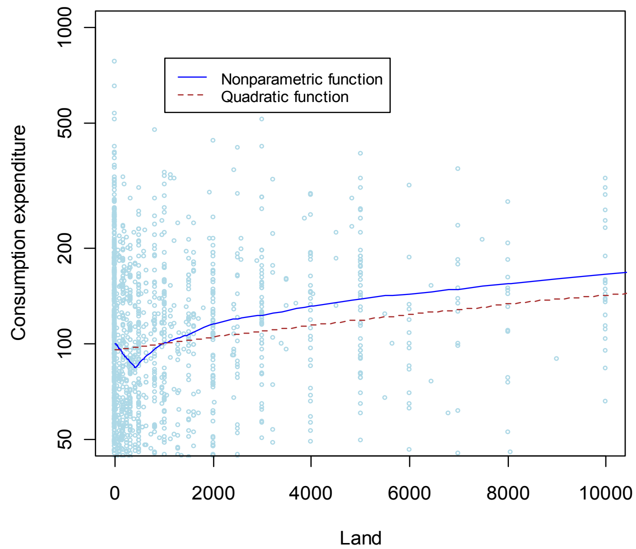

Figure 1: Effect of land on consumption expenditure

Before moving into the estimation of our semiparametric PLR model, we also want to investigate whether the functional form in Equation (3.14) is appropriate for our cross-section sample. Two specification tests are run for this purpose. First, based on a test proposed by Hsiao et al. (2007), we test the null hypothesis that the consumption model has a parametric form such as $y = t\gamma + X\beta + \varepsilon$, against the proposed alternative that it takes the semiparametric following form $y = f(t) + X\beta + \varepsilon$. Recall that in Equation (3.14) there were k regressors that entered the income equation parametrically. The nonparametric kernel-based model specification test differentiates discrete from continuous variables while smoothing and it is asymptotically normal distributed under H_0 . The resulting statistic from Hsiao and co-authors' test is 8.17 (with a p-value < 0.01), result that rejects the linear model in favor if the semiparametric one at the 0.1% level of significance.[117] In terms of the Goodness-of-Fit, the linear model has an R^2 of 50.47%.

Next, using the same test for model specification, we want to test a parametric quadratic model for land, with all the other regressors in Equation (3.14) also included, against the semiparametric formulation (the alternative). With an estimate test statistic of 4.02, and the corresponding p-value < 0.01, once again there is enough evidence to reject the full parametric null hypothesis in favor of the semiparametric PLR model. The quadratic specification for land improves the fit slightly (R^2 of 51.25%).[118]

After obtaining the specification results on the consumption equation, the paper now proceeds with the estimation of the semiparametric partial linear model. To do so, it implements Robinson's (1988) method outlined in Section 3.4 and using *The np Package*, which as mentioned before, uses specific kernel functions for quantitative and qualitative variables, and it applies by default least-squares cross-validation bandwidth selection (Hayfield and Racine, 2008).

---

[117] The distribution of the test statistic "J_n" under the null hypothesis is obtained by bootstrapping methods to avoid any finite-sample issues, based on which the bootstrap p-values are obtained (Hayfield and Racine, 2008).

[118] According to Hayfield and Racine (2008), it is possible to use a mixed residual option to perform the underlying specification tests, which leads to gains in terms of efficiency. The resulting statistics and p-values from this alternative method confirmed and showed sufficient evidence to reject the parametric specifications in favor of the semiparametric formulation.

Table 7. Semiparametric PLR model estimates

Variable	Coefficient		SE
Head's years of schooling	1.99	***	0.62
Household total years of education	2.43	***	0.23
Number of adults	1.48		2.08
Number of children	1.04		1.87
Livestock assest	0.17	***	0.03
Distance to the nearest's capital of Department	-0.12	***	0.04
Valley (1=yes; 0=no)	35.30	***	3.83
Lowplains (1=yes; 0=no)	59.65	***	4.94
N	1374		
R^2	0.61		
Residual standard error	3472.34		

Significance levels: *10%, **5%, ***1%

Table 7 contains the coefficients estimates for the parametric component of the consumption function (recall that consumption is used as a proxy for the household wellbeing), which includes household human and physical capital variables, household characteristics, a contextual variable, and the geographic-specific dummy variables.[119] At a first glance, a comparison of this model with the parametric ones indicates that the fit outperforms them with a R^2 of 61.16%.

The results in terms of the two human capital variables included in the model are as expected. They have a positive and significant impact on the household consumption expenditure. Another key asset affecting wellbeing is the ownership of livestock assets. They nourish family, provide cash when sold, and serve as inputs in farming. The variable is highly significant and it is positively associated with the household wellbeing. In terms of the household composition variables, none is statistically significant. The contextual variable distance to the closest capital of Department has a negative impact on the level of

[119] A dummy variable for the regions where coca leaf production exits was initially included, but later dropped since the coefficient estimate turned out not to be significant at conventional levels, for any of the model specifications estimated for the consumption function.

the wellbeing. Being located further away from the market tends to reduce the probability of nonfarm income opportunities, and the size of the market where the farm products can be traded. After we control for differential asset positions and contextual characteristics, regional dummies are highly significant. Households in the *Valles* and *Llanos* regions have a considerable higher level of wellbeing compared to those in the *Altiplano*. The positive coefficients might reflect the better agroclimatic conditions, soil quality, type of farming practiced, and other idiosyncratic specificities of those regions that make them enjoy higher economic wellbeing levels.

The effects of the parametric linear variables on the farmers' wellbeing can be visually confirmed in Figure 2 through the partial regression plots along with their bootstrapped error bounds. While each 2D plot of the outcome consumption expenditure versus each covariate $x_1, \dots, x_k$ is displayed, all the other covariates are held constant at their respective medians.

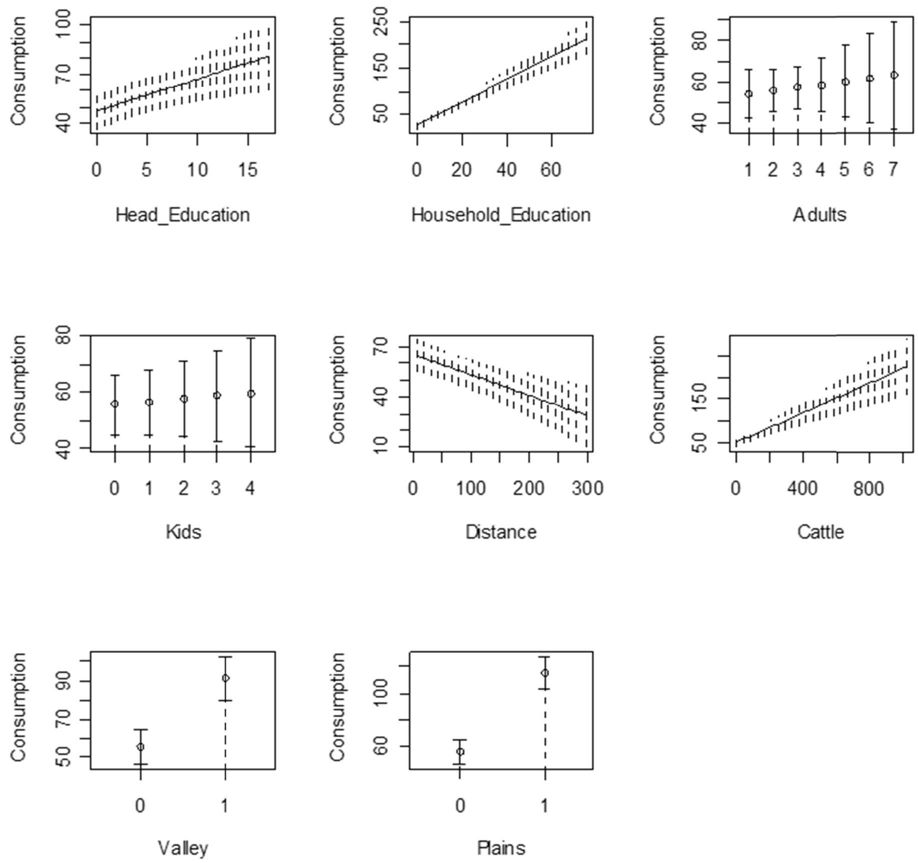

Figure 2: Partial local linear nonparametric regression plots with bootstrapped error bounds

Figure 3 displays the semiparametric estimates of f, that is, the nonlinear effect of land on consumption, where the horizontal axis shows the value of land assets and the vertical axis shows $\hat{f}(t)$. As proposed, the estimated relationship between both variables has a "U" shape, i.e. the marginal wealth effect of land is negative for small land asset values, generally positive thereafter, an increasingly positive after some point.

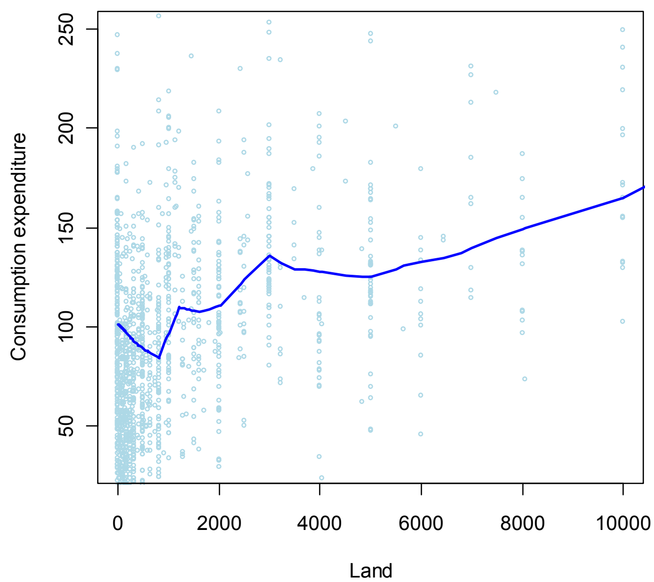

Figure 3 - Semiparametric PLR model: Effect of land on wealth

In terms of the economic impact, Table 8 presents the average percentage change in wealth as the value of the land assets held by rural households' increases. The fitted values obtained indicate, for example, that for those rural farmers whose land assets' value is at most US$ 100, which accounts for 11% of the observations in the sample, the average consumption expenditure decreases in 45% compared to those that have no land. Thereafter, the marginal wealth effect is decreasingly negative until land values that are greater than US$ 900. This suggests that up to that threshold, larger land endowments for small farmers cannot be used efficiently in production partially because they are not accompanied by increases in working capital.

The estimated results also show that the positive effects of an increase in the value of land assets on wealth are effective for farmers with larger land assets, which would correspond more likely with industrialized and/or modern agricultural farmers. The increase in the wellbeing that those farmers enjoy, as the value of the land assets increase, ranges from

11% up to 169% on average. A total of 36% of the observations in the sample have this positive marginal return of land.

The results from a significance test on land establish that this variable, with a highly nonlinear impact on the wellbeing of the rural farmers in the sample, is significant at all conventional levels. With a normally distributed test statistic value of 9.099, and a p-value less than 0.001, the restricted model in which $f(t)$ is a constant against the nonparametric specification (the alternative) is by far rejected. Yatchew (2003) proposed a test using a differencing semiparametric PLR estimation procedure that is relatively less efficient compared to other smoothing methods such as the Robinson's (1998) estimator. Yatchew's differencing estimator method and the full set of results from it are both found in Appendix D. The overall results are highly similar to the semiparametric PLR estimates in Table 7 in terms of the sign of the coefficients, but there are evidently some gains in terms of significance due to the smaller standard error estimates.

Table 8. Marginal wealth effect of land[a]

Land value (US$)	Average wealth change[b]	Percentage of farmers affected[c]
0 - 100	-45%	11%
101 - 200	-40%	10%
201 - 300	-37%	6%
301 - 400	-30%	7%
401 - 500	-18%	9%
501 - 600	-15%	1%
601 - 700	-32%	2%
701 - 800	-15%	1%
801 - 900	-12%	4%
901 - 1,000	11%	5%
1,001 - 2,000	10%	11%
2,001 - 5,000	28%	12%
5,001 - 10,000	42%	4%
> 10,000	169%	3%

[a] Based on the fitted values from the semiparametric PLR model

[b] Percentage change in the average consumption expenditure compared to the average level when land assets are zero.

[c] Percentage rate compared to the whole sample

The statistically significant regional dummy estimates reported in Table 7, suggest that individual semiparametric estimations should be performed for each region. Before we do so, however, we want to explore the (dis)similarity of the parametric effects across regions. In order to do so we perform a test proposed by Yatchew (2003), which is based on the estimates of the parametric variables and the corresponding residual variances, both obtained for each regional subpopulation.[120] Under the null hypothesis, the statistic converges to the Chi-square distribution with K degrees of freedom. The null of similarity between each pair of regional parametric estimates were plainly rejected, with computed values of the χ^2 test statistic equal to 31.057, 16.925 and 26.22, for the pairs *Altiplano-Valles*, *Altiplano-llanos*, and *Valles-Llanos*, respectively. In all three cases the corresponding p-values were smaller than 1%.

[120] The statistic for any two regions is calculated as: $\left(\hat{\beta}_{reg1} - \hat{\beta}_{reg2}\right)' \left(\hat{\Sigma}_{\hat{\beta}_{reg1}} + \hat{\Sigma}_{\hat{\beta}_{reg2}}\right)^{-1} \left(\hat{\beta}_{reg1} - \hat{\beta}_{reg2}\right) \rightarrow \chi^2_{dim\,(\beta)}$.

Based on the preceding test results, Table 9 provides the parametric estimates for each of the regional subpopulations. As it is evident, the three sets of estimates are similar in terms of the direction of the marginal effects on wealth for those explanatory variables that are statistically significant. The coefficients of the household's assets, human capital and livestock, confirm their expected positive impact on wealth in all three regions. However, there are some variations in terms of statistical significance levels and the size of the coefficients across regions, as could be expected based on the rejection of the parametric-equality tests performed. The negative wealth impact of the distance to the nearest economic and political center is significant for two out of the three regions. In terms of the household's composition variables, they are only significant for one region at a time.

Table 9. Regional Semiparametric PLR model estimates

	Highland			Valleys			Lowplains		
Variable	Coefficient		SE	Coefficient		SE	Coefficient		SE
Head's years of schooling	1.42	***	0.61	0.72		1.50	4.09	***	1.40
Household total years of education	2.15	***	0.25	4.33	***	0.55	1.32	***	0.44
Number of adults	-3.05		2.32	-2.77		4.53	12.07	***	4.13
Number of children	3.90	***	1.85	-1.10		4.41	3.76		3.87
Livestock assest	0.13	***	0.04	0.14	***	0.05	0.30	***	0.06
Distance to the nearest's capital of Department	-0.20	***	0.04	0.02		0.08	-0.28	***	0.07
N	592			484			298		
R^2	0.39			0.28			0.79		
Residual standard error	1467.7			6175.8			3478.6		

Significance levels: *10%, **5%, ***1%

Having estimated the parametric effects separately for each sub-population, there is one last question that needs to be answered regarding the land variable and its significance in the consumption expenditure specification. Do land assets significantly affect wealth at the regional level? Based on the same test used for the overall significant effect of the variable, where in the restricted model $f(t)$ is a constant and under the alternative it has the nonparametric specification, we found that the asset is significant in all three agroclimatic regions. The normally distributed coefficients of 1.33, 3.50, and 27.64 allowed us to reject the null hypotheses at 10% level in the *Altiplano*, and at 1% level in the other two regions.[121]

Figure 4 graphically presents the regional semiparametric curves of the wealth effect of the value of land assets. The horizontal axes show the US$ value of land owned by rural households and the vertical axes the regional kernel estimates of $f(t)$. These results confirm our earlier semiparametric estimates when the overall effect of land was obtained using the whole sample. There is an initial negative marginal effect of land at low values of the asset in all three regions. This might be the result of increases in landholdings which are not accompanied by a greater availability of credit or of any other additional source of income. The farmer faces a tighter working capital constraint.

The specifics of the negative effect in terms of wealth, the number of rural households affected, and the inflection point at which the marginal value of land changes to be positive, widely varies across regions. In the case of the *Altiplano* region, for land assets with a value no greater than US$ 50, the average impact on wealth is negative, and the percentage of the households affected negatively is 8%. From there on, the highest average increase in wealth is 194%. For rural households living in the *Valles*, the negative wealth impact is effective for land values that are smaller than US$ 600, and affects 20% of the regional sample. Thereafter, the wealth of the households with larger land assets increases 1.56 times on average. Finally, in the case of the *Llanos* region, the highest negative marginal effect of land is a 19% decrease in wealth, and the percentage of the households facing it 14%. In this case, the land value at which the marginal wealth effect of the asset turns to be positive is approximately US$ 4,000. The wealth of the households that enjoy this positive returns increases 1.1 times on average.

---

[121] The corresponding p-values were 0.09, 0.00, and 0.00.

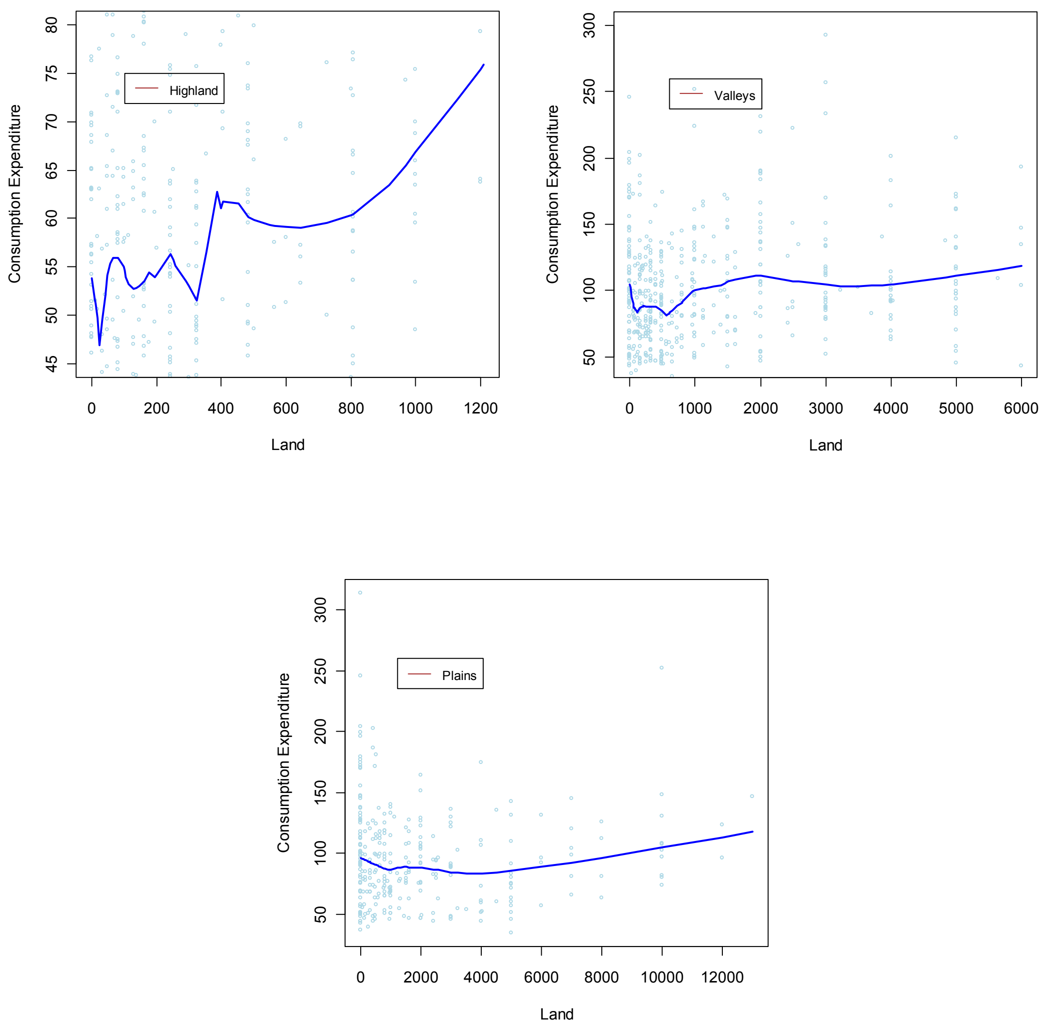

Figure 4 - Semiparametric PLR model: Regional effects of land on wealth

3.7 Conclusions

The production relationships in agrarian economies of the developing world have been widely recognized to be characterized by the absence of certain markets (or their imperfect existence). Additionally, economic theory of information suggests that both capital and labor markets are likely to be intrinsically imperfect in agrarian economies. These are all characteristics of the environment in which farmers make decisions and eventually produce. This environment, and the access to the relevant rural markets, however, may be different for different rural farmers, depending on the scale of production, i.e. the size of their land endowments. For instance, small farmers that do not access the imperfect credit market (and hence are capital-constrained) face the necessity to supply labor to the market in order to meet subsistence needs and obtain liquidity. Increases in land endowments in such circumstances may not return large (or even positive) marginal wealth effects, since they might not be able to use them efficiently in production. These farmers face a trade-off between access to working capital through off-farm labor, and availability of labor force to work in their own farm. On the other hand, the paper also models large farmers that have access to credit markets through their land endowments, which allows them to efficiently use larger landholdings by increasing the usage of inputs of production, which leads to cost advantages due to the existence of fixed supervision costs.

Based on those considerations, the theoretical model derived for each type of farmer showed us that the marginal return of land for the small farmers might be smaller compared to the larger ones, or even negative depending on the size of the terms involved in the derived income value function. For the larger farmers the model showed that increasingly positive returns to land can be expected.

Using data from a nationally-representative household survey in Bolivia, three are the main conclusions that emerge from the estimation results using a Semiparametric Partial Linear Regression model. First, the relationship between land and wealth is significantly nonlinear. Second, for farmers with very low values of land endowments (e.g. less than US$50), results showed that wealth decreases with higher land values. For land values between US$50 and US$ 900, the effect is decreasingly negative. Beyond land values of US$ 900, the effect is on average positive and rising. These results suggest that only farmers with land holdings with value that is above that threshold can effectively access the credit market and use land endowments as collateral. Third, accounting for regional effects, we found support for the nonlinear effect of land and the initial negative marginal wealth effect in all three agroclimatic regions. However, the impact in terms of wealth reduction, the number of adversely affected farmers, the value of the land at which the marginal return of the land starts to be positive, and the average increase in the wealth that farmers experience thereafter, widely varies in regional terms.

Therefore, our findings refine and contribute to the conclusions drawn by other studies where it is shown that, in some developing countries, small farmers would enjoy higher marginal returns on land compared to large farmers. This paper establishes that positive marginal returns on the asset are effective if the demand of additional working capital, as land endowments increase, is accompanied by greater access to resources through the credit market.

The regional analysis performed confirms the results, and reveals that regardless of regional agroclimatic and idiosyncratic differences, the constraints that an imperfect credit market imposes persist.

The relevance of policies that improve the access of small farmers of the developing countries to a more competitive financial market becomes clearer. Additional policies that accompany those that are simply related to increased access to land by the poor are also highly advisable.

CHAPTER IV – NONFARM LABOR, POVERTY AND INEQUALITY IN RURAL BOLIVIA

4.1 Introduction

There is a consensus regarding the increasing importance of nonfarm activities and the income streams coming from them, for countries in the developing world. These diversification activities that have been used to compensate the low and risky returns from farm related activities.

The mixed empirical evidence regarding the effect of nonfarm employment on the income of the poor in the rural area and on the overall income distribution, however, has motivated vast theoretical and empirical analyses. The differentiation of the nonfarm activities in terms of their return, the level and value of the assets of the poor, and the overall conditions of the rural and national economies in terms of availability of land, labor and wealth distribution, can be used to explain the variety of the results.

Bolivia presents numbers for poverty and inequality that are among the highest in Latin America. At the same time, as it is the case for other developing countries, the participation of nonfarm income has been constantly increasing over the past years. These facts have raised questions regarding the effect of the presence of the nonfarm income on the income of the households.

Despite the results from some empirical works for Bolivia that explain the potential unequalizing effect of nonfarm income, this work proposes that based on the characteristics of the country in terms of access to land, supply of labor, and market barriers mainly faced by poor farmers, it might have a positive impact (i.e. equalizing). In short, we want to find out if nonfarm income is sufficient to help poor farmers escape from poverty, rather than just help them survive, and if it has an equalizing effect in the overall income distribution. Evidence in these matters will fill the void of the current available literature in terms of a rigorous empirical analysis on the country.

Using a nationally representative dataset for Bolivia, we follow Chinn (1979) and de Janvry et al. (2005) to address these questions by simulating a “hypothetical” state in which nonfarm activities do not exit. The results show that the depth and severity of poverty in Bolivia would have been higher if diversifying farmers would have not engaged into nonfarm activities. The results show, however, that the increase in income was not sufficient to abate the number of poor. We found no evidence that nonfarm income affect negatively nor positively the overall income distribution.

The study proceeds in four further sections. Section 4.2 presents a brief review of the relevant literature. Section 4.3 explains the econometric model and estimation procedures. Section 4.4 presents the data and the specification of equations of the model. Section 4.5 shows and discusses the results. Section 4.6 concludes.

4.2 Literature review

Farm income is still the main source of income for rural dwellers in developing countries (e.g. in Reardon *et. al,* 1998b, and Elbers and Lanjouw, 2001). Income coming from nonfarm activities, however, is not just an extraordinary event anymore, but a regular means of living. As it has been widely shown, nonfarm work, and the income that rural households obtain from them, is an important strategy among countries of the developing world since it has been used to compensate the low returns in farm related activities (Reardon *et al.*1998a). Comparing regions, Latin America was found to have an average share of 40% of nonfarm income in total rural household income, just 2 points behind Africa, and 8 above Asia.

Diversification into nonfarm activities usually responds to market failures and the insufficiency of farm income. Its ultimate goal, however, is to achieve poverty alleviation (Barret et al., 2001). Rural participation in nonfarm activities potentially improves rural income as a whole. The ultimate effect of nonfarm sources of income on poverty, however, will depend on the nonfarm activity sector to which the household have access: one with low returns that will only ensure a survival strategy, or one with rather high returns that can be an option to escape poverty (Davis et al. 2007). De Janvry et al. (2005), for example, show that nonfarm income sources in China effectively reduce the incidence, depth and severity of poverty among rural households with small landholdings.

Income diversification activities can be classified in different ways; Barret *et al.* (2001), for example, use three criteria: sectoral, functional and spatial; Elbers and Lanjouw (2001) distinguish according to productivity; Escobal (2001) classifies first between self and wage employment, and then according to the level of the skills (entry barriers and rates of return). Based on the classification suggested by de Janvry *et al.*(2005), the income of the rural farmers is classified into three groups: *farm income*, which comprises income coming exclusively from activities related to agriculture such as forestry, livestock and fishery; *nonfarm income,* which includes agricultural wages, formal and informal nonagricultural wages, and the proceeds from self-employment activities; and income from *other non-*

productive sources such as pension transfers, grants/subsidies, rents, and financial income. We follow this classification to group the sample into two groups: farmers that do not diversify income (i.e. households with farm income only), and farmers that do diversify into nonfarm work (i.e. households with farm and nonfarm income).

In terms of the distribution of income, de Janvry and Sadoulet (1993) explain that the impact of nonfarm activities on the distribution of income in the rural areas greatly depends on the distribution of vital agricultural assets such as land. This helps to understand the varied results across regions within a country, and across countries. Concisely, Adams (2001) states that in land-scarce and labor rich countries, the inadequate access to land "push" poorer farmers into the nonfarm sector, which provides them with a better way of increasing income and with employment opportunities. Note that nonfarm income might still be an unequally distributed income source, but not as bad as the overall income distribution. Empirical evidence for the equalizing effect of income sources generated in nonfarm activities is abundant.[122]On the other hand, it is expected that nonfarm income will increase overall income inequality in land-rich and labor-scare countries where access to land keep the majority of the farmers involved solely in agricultural work. Only the richest are "pull" into the nonfarm sector. In this case nonfarm income is more unequally distributed than farm income. Empirical evidence for the unequalizing effect of nonfarm income can be found from country studies in each developing region.[123] In terms of geographical regions, the rough pattern seems to be that in Africa, nonfarm employment tends to increase inequality, the opposite is true for Latin America, and Asia is definitely a mixed story.

[122] See for example De Janvry, et al (2005) for China, Adams (2002) for Egypt, Adams (1995) for Pakistan, Lanjouw (1999) for Ecuador, and Chinn (1979) for Taiwan.

[123] Davis et al. (2007) studies 15 developing countries and concludes that nonfarm income is increases income inequality. Other studies with similar results include: Elbers and Lanjouw (2001) for Ecuador, Adams (2001) for Jordan, Escobal (2001) for Peru, Reardon and Taylor (1996) in Burkina Faso, and Collier et al (1986) in Tanzania.

Despite the controversy concerning the effect of inequality on economic development, based on cross-country data analysis, Easterly (2007) has shown the significant barrier that inequality is to development, not only lowering per capita income, but also negatively affecting other aspects of it, such as the quality of institutions and schooling.

According to the CEPAL (2006), Bolivia has one of the highest percentages of poor people in the rural area of Latin America. Even though during the last 10 years the poverty incidence has decreased in the country by more than three percentage points[124], for which the major contributor is the rural decrease in poverty, it has not been enough, and it won't be, to reduce poverty significantly and achieve the Millennium Development Goal for the country.

In terms of income inequality, the high levels existing in the country are not decreasing over time and, on the contrary, they are increasing. According to the National Bureau of Census (INE), the Gini coefficient increased from 0.58 to 0.59 between 1999 and 2006[125]. The index was 0.42 in 1990 according to the Income Inequality Database of the World Bank[126].

As it is the case for other developing countries, traditionally rural Bolivia has been seen as driven almost uniquely by agriculture; policies and resources have been focused on the farm sector as the way to alleviate poverty. Nevertheless, according to the INE the percentage of workers involved in the agricultural sector in 2005 was 37% and the percentage for the rural area was 79%. Although these numbers seem high, they are not unusual for a developing country, and have been decreasing over time: in 1999 they were 39% for the entire country, and 82% for the rural area. Hence, the importance of nonfarm income has been increasing over the past years.

---

[124] According to the INE (Bureau of Census), from its web page www.ine.gov.bo, the incidence of poverty at the national level dropped from 63.47% to 59.92% between 1999 and 2006, the urban numbers dropped from 51.36% to 50.27% during the same period, and in the rural area from 84.0% to 76.47%.

[125] Access through the web page www.ine.gov.bo.

[126] Access through its webpage: http://econ.worldbank.org

Jimenez and Lizarraga (2003) examine the structure of the household's income in the rural area of Bolivia, and show that almost 40% of the income comes from activities not related to agriculture. This confirms that agricultural activities are still the main source of income for rural dwellers. Nonfarm activities, however, are not an extraordinary event anymore, but a regular means of living that increases its participation in total income as time goes by.

Studies for rural Bolivia have suggested that the rise in inequality in the country could be due to the increase in nonfarm income participation in the income sources of the rural households. Sanchez (2006) and Jimenez and Lizarraga (2003) have found empirical evidence for the potential positive contribution of nonfarm activities to the overall income inequality. The available empirical evidence, however, is based on limited data that do not differentiate the effect of nonfarm income across the different agroclimatic regions. The works mainly focused on the determinants of participation and income related to nonfarm activities, and only reports the potential contribution of each source of income to inequality.

Contrary to those studies, this work proposes that given that Bolivia is a land-unequal distributed and labor-rich country, the prevalent inadequate access to land might "push" poorer rural households into diversification, and hence nonfarm income can have an equalizing effect on overall income inequality. In terms of poverty, we want to find out if nonfarm income sources are enough to help poor farmers to escape poverty.

We follow the methodology first developed by Chinn, D. (1979) for evidence from Taiwan, and later by de Janvry, *et. al,* 2005 for China, to assess the impact of nonfarm income in rural Bolivia. It simulates the importance of nonfarm income under a "hypothetical" situation where the nonfarm opportunities are eliminated to see how the rural household incomes, rural poverty, and rural inequality would have been in such a case. The analysis takes special care of the specific characteristics that the households have, such characteristics that make some of them diversify, while makes others remain as pure farmers.

In terms of the estimation process, a selection model is used to obtain the fitted incomes of pure farmers that will later be used to predict the hypothetical incomes for the diversifying ones. According to Winkelmann and Boes (2006) the best method to deal with sample selectivity is Maximum Likelihood Estimation (MLE), with the resulting parameter estimates being consistent, asymptotically efficient, and asymptotically normal. Identification of the model under sample selectivity is addressed by the inclusion of specific exclusion restrictions.

4.3 Econometrics of the model

4.3.1 The log-income estimation procedure

Following de Janvry *et. al* (2005) procedure to predict the income of the rural households as if all were only farmers, it first needs to be estimated the income equation for the actual pure farmers. Since the data contain diversifying and non-diversifying households, estimates of the farmer's income, based just on the farmers subsample, could give rise to sample selectivity problems. It is necessary to account for the diversification decision to obtain unbiased estimates.

Hence, the initial dependent variable is binary and can be denoted by z_i for household i, which equals one when the household is a pure farmer, and equals zero when the household diversifies out from farm work. Therefore, z_i can be assumed to have a Bernoulli probability function of the form:[127]

[127] For details see Winkelmann, R. and Boes, S. (2006)

$$f(z_i|X_i) = \pi_i^{z_i}(1-\pi_i)^{1-z_i}, \qquad z_i = 0,1, \qquad i = 1,2,\dots,N \tag{4.1}$$

where X_i is a vector of observable characteristics, and π_i is equal to $G(X_i'\beta)$, whit this function observing the condition that: $0 \leq G(X_i'\beta) \leq 1$.

According to Crown (1998), using ordinary least squares estimation for such binary dependent variable would present several concerns regarding their econometric validity.[128] More efficient procedures for modeling binary outcomes are the Logit and Probit models. It can be shown that Logit and Probit MLE are highly similar in terms of the sign and magnitude of the slope coefficients, though the numerical values may depend on the normalization chosen. For simplicity, the paper assumes the model errors are normally distributed and hence Probit model is chosen.

In terms of our variables, the Probit model has the following form:

$$\pi_i = \Phi(X_i{'}\beta) = \int_{-\infty}^{x_i'\beta} \frac{1}{\sqrt{2\pi}} exp\left[-(\frac{t^2}{2})\right] dt \tag{4.2}$$

where $\Phi(\cdot)$ is the cumulative density function of the *standard normal distribution.*[129]

The use of a latent variable model is a helpful tool to derive conditional probabilities within the framework of sample selectivity and Probit models.[130] The diversifying decision and the income of the pure farmers are the two equations of the model, where the dependent variables y_i° and z_i° are latent or unobserved, and are generated by a bivariate process:[131] [132]

---

[128] Among them: OLS estimates would invalidate the interpretation of the predicted values as probabilities, they could cause heteroskedasticity for some observations, and they would assume a constant marginal effect on the dependent variable when a one-unit change occurs in any of the explanatory variables, which is clearly not true.

[129] The underlying assumption is that, in the latent variable equation, the error terms are independently and normally distributed, with mean 0 and variance σ^2. Besides that, the required normalization procedure is applied.

[130] See for example Davidson and Mackinnon (2004), pp. 453-454.

[131] The section follows Davidson and Mackinnon (2004).

[132] Here it is worthwhile to stress the difference between the superscript "o" and the subscript "0", and the different contexts within which each one is used; the former is used conventionally to denote latent variables

$$\begin{bmatrix} y_i^\circ \\ z_i^\circ \end{bmatrix} = \begin{bmatrix} X_i\beta \\ W_i\gamma \end{bmatrix} + \begin{bmatrix} u_i \\ v_i \end{bmatrix}, \begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim NID\left(0, \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}\right) \tag{4.3}$$

where X_i is, as explained, a vector of exogenous variables explaining the income of rural households that remain as pure farmers, and β the vector of parameters that need to be estimated. W_i, on the other hand, is the vector of observations on explanatory variables for a dependent variable of which only the sign is observed. σ is the standard deviation of u_i, and ρ the correlation between u_i and v_i. As in any other typical model that involves latent variables, these are not directly observed, but rather inferred from the decisions taken by the households. The relationship between the observed and the latent variables can be summarized as follows:

$$\begin{gathered} y_i = y_i^\circ \; if \; z_i^\circ < 0; unobserved \; o.w. \\ z_i = 1 \; if \; z_i^\circ < 0 \; ; \; z_i = 0 \; o.w. \end{gathered} \tag{4.4}$$

In terms of Equations (4.3) and (4.4), there are two types of observations: a) households who decide not to diversify out from farm activities ($z_i^\circ < 0, z_i = 1$), and whose farm income y_i is observed (and equal to y_i°) along with both W_i and X_i; and b) households who decide to diversify their income out from farm activities ($z_i^\circ > 0, z_i = 0$), with such decision observed in the value of z_i, along with W_i.

Following Winkelmann and Boes (2006), a model such as the one contained in Equations (4.3) and (4.4) can be efficiently estimated using the Maximum Likelihood Estimation (MLE) method, from which consistent estimates of the parameters and standard errors can be obtained. Therefore, the likelihood function of the full sample, as reflected in Equation 5, contains two elements: the contribution of households that do not diversify, and that of the households that do diversify out from farm work. In other words:

---

whose principal characteristic is not being directly observed and are part of the set up of our econometric model; the latter, already introduced, denotes the situation in which all the households are assumed to be pure farmers, and so will be used in subsequent sections when we are predicting /simulating the corresponding household's income .

$$I(z_i = 1)\Pr(z_i = 1) f(y_i^{\circ}|z_i = 1), \tag{4.5}$$

and

$$I(z_i = 0)\Pr(z_i = 0)$$

where $I(\cdot)$ is an indicator function that takes the value 1, when its argument is true, and 0 otherwise.

Combining the two terms of Equation (4.5), taking logarithms, and adding them up for the N observations in the sample, the loglikelihood function of model is given by:

$$l = \sum_{z_i=1} \log\{\Pr(z_i = 1) f(y_i|z_i = 1)\} + \sum_{z_i=0} \log \Pr(z_i = 0) \quad i = 1,2, \dots, N \tag{4.6}$$

The second term in Equation (4.6) is the likelihood contribution of a household that diversifies into nonfarm activities, and equals $\Pr(z_i = 0) = \Phi(W_i\gamma)$[133]. The first term, which is the likelihood contribution of a household that does not diversify, can be written as[134]:

$$\sum_{z_i=1} \log\{\Pr(z_i = 1|y_i^{\circ}) f(y_i^{\circ})\}$$

where $f(y_i^{\circ})$ is just the density of y_i° conditional on exogenous variables, which is normally distributed with mean $X_i\beta$ and variance σ^2. Using the facts that u_i and v_i are bivariate normal, and since y_i is observed when $z_i = 1$, we have:

$$\Pr(z_i = 1|y_i^{\circ}) = \Phi(-\frac{W_i\gamma + \rho(y_i - X_i\beta)/\sigma}{\sqrt{(1 - \rho^2)}})$$

Thus, our final loglikelihood function for the full sample of rural households is given by:

---

133 We use the fact that v_i is normal distributed.

134 Since the two error terms are bivariate normal, $v_i = {}^{\rho u_i}/_{\sigma} + \varepsilon_i$, where ε_i is a normally distributed random variable $(0, 1 - \rho^2)$. For more details see Davidson and Mackinnon (2004), pp. 487.

$$l = \sum_{z_i=1} log\ \Phi\left(-\frac{W_i\gamma + \frac{\rho(y_i - X_i\beta)}{\sigma}}{\sqrt{(1-\rho^2)}}\right) + \sum_{z_i=1} \log\left(\frac{1}{\sigma}\emptyset\left(\left(\frac{y_i - X_i\beta}{\sigma}\right)\right) + \sum_{z_i=0} log\ \Phi(\mathrm{W_i}\gamma) \quad (4.7)$$

$$i = 1,2, \dots, N$$

which is maximized with respect to γ, ρ, β and σ using the sample for pure farmers.

Once the estimates from the above explained sample selectivity model are obtained, the prediction of the individual log-incomes for the households that are involved in nonfarm activities are obtained, as explained in the next section.

4.3.2 Predicted income for diversifying households[135]

Consider a log-linear model, where the income for each household i is given by its logarithm form as follows:[136]

$$logy_i = X_i\beta + u_i \qquad i = 1,2, \dots, N \quad (4.8)$$

where X_i is a vector of observations on explanatory variables, and u_i are the error terms that account for unobservable characteristics of the household. In terms of conditional expectations, Equation (4.8) can be written as:

$$E(log\ y_i\ |X_i) = E(X_i\beta/X_i) + \ E(u_i/X_i) = X_i\beta \qquad i = 1,2, \dots, N \quad (4.9)$$

since it is assumed that $E(u_i/X_i) = 0$.

Among the N rural households, as explained, there are n_1 individuals whose sole source of income is from farming, and $N - n_1$ individuals that diversify their income sources. Given that this work looks to predict the income for all of them as if all would be pure farmers, the subscript "0" will denote such a predicted state. Note that in the case of the actual pure farmers, this notation denotes their actual current state.

135 The procedure for predicting incomes and some notation used in this section follows de Janvry et al.(2005).

136 The use of income in its logarithm form, instead of the linear alternative, is based on Ermini and Hendry (2008), where the results of multiple mis-specification tests favor the use of income in such form.

Making use of such notation, and following Wooldridge (2002), Equation (4.8) can be rewritten in a decomposed way, and the log income for any household i, predicted to only participate in farm activities, is now given by:

$$log\, y_{0i} = E(log\, y_{0i}\, |X_i) + u_{0i} = X_i\beta_0 + u_{0i}, \qquad i = 1,2, \dots, N \qquad (4.10)$$

Given the existence of the two types of households, the prediction of the income for each household i, as if it would be a pure farmer, can be split into two groups: one for the households that are actually pure farmers and whose predicted income is the *observed* total income; and the other one for the households that actually *diversify* and whose predicted income would be the *predicted* income as if they were pure farmers.

From the estimation of the income model for the actual n_1 pure farmers, and the $\hat{\beta}_0$ parameters gotten, we can obtain the predicted income for $N - n_1$ diversifying households, as if they were pure farmers. Following the approach suggested by de Janvry *et al.* (2005), three variables are required to pursue this objective. First, the same X_i explanatory variables included in the estimation of the log income for the actual pure farmers, which are also observed for the diversifying farmers. Second, the estimated parameters $\hat{\beta}_0$, already obtained from the MLE. Finally, an estimator for the unobserved term u_{0i}; for this, a random variable e_{0i} is constructed based on the statistical properties of the pure farmers' errors u_{0i}. Therefore, $\hat{e}_{0i} = \hat{\sigma}_0 \Phi^{-1}(r)$, where $\hat{\sigma}_0$ is the estimated standard error of u_{0i} for the group of pure farmers, r is a random number between 0 and 1, and $\Phi^{-1}$is the inverse of the CDF of the standard normal distribution.[137]

With this information at hand, it is easy to predict the log-income for the $N - n_1$ group of diversifying households, as if they were only involved in farm activities, in the following manner:

137 This procedure makes sure that the unobserved terms for the predictions follow the same variation and probability distribution as the observed residuals for the pure farmers.

$$\widehat{log}y_{0i} = \hat{\beta}_0 X_i + \hat{e}_{0i} \qquad\qquad i = n_1 + 1, \dots, N \tag{4.11}$$

with the corresponding predicted income equal to:

$$\widehat{y_{0i}} = exp\left(\widehat{log}y_{0i}\right) = exp\ (\hat{\beta}_0 X_i + \hat{e}_{0i}) \qquad\qquad i = n_1 + 1, \dots, N \tag{4.12}$$

With the predicted incomes obtained, we can proceed now with the assessment of the impact of nonfarm income on poverty and inequality.

4.3.3 Impact of nonfarm income on rural inequality and poverty

This section briefly explains the methods used to measure poverty as well as the distribution of income, and further assesses the impact of nonfarm income on them. It explains the indicator used to measure household welfare and the chosen measures of poverty and inequality.

4.3.3.1 Poverty

In this chapter, poverty is understood as a "pronounced deprivation in well-being" as explained by the World Bank (2005). In its more basic approach, this definition assesses whether households or individuals have enough resources to meet their needs by comparing each observation's income or consumption expenditure to a predefined threshold. Since the approach used in this paper requires classifying the farmers according to income sources, the variable used to proxy for well-being is income.

As also explained by the World Bank (2005), the choice of a poverty measure is highly arbitrary. The measures included here aim to provide complementing estimates that account for the number of poor, the severity of poverty, the inequality among the poor, and the median income of the poor, or the median statistic. Concisely, the poverty measures included in the analysis are the three most widely used developed by Foster, Greer, and

Thorbecke (1984), also referred as P_α indices or FGT class of poverty measures, along with the median income of the poor. In terms of the P_α indices, all three can be derived from the following expression:

$$P_\alpha = \frac{1}{N}\sum_{i=1}^{n}(\frac{L-y_i}{L})^\alpha I_i m_i$$

where N is the weighted sample size, i subscripts households, y is the relevant measure of welfare in per capita terms (i.e. household income per capita)[138], L is the poverty line, I is an indicator equal to 1 if $y_i < L$, and 0 otherwise, m_i stands for the weight of each household in the measurement given by its number of members, and α is a parameter equal or greater than zero. When $\alpha = 0$, the index measures the proportion of total population that is below the poverty line, also known as the headcount index, which assesses the extent or incidence of poverty.[139] The headcount index is easy to calculate and interpret, but it does not account for the intensity of poverty. If in the above poverty expression we set $\alpha = 1$, the measure obtained is the Poverty Gap Ratio, or aggregate shortfall as a proportion of the poverty line, averaged over the whole sample. It counts the poor among the population, and how poor they are, i.e. it measures the depth of poverty.[140] Finally, if $\alpha = 2$, the expression results in the Squared Poverty Gap, which is a more difficult measure of poverty to interpret, but it has the advantage of putting more weight on the poorest among the poor. The remaining summary measure is the median income among the poor, and its inclusion intends to

---

[138] As already explained, the survey data used in the present work considers the household as the unit of observation. This requires assuming that all members of a given household enjoy the same level of well-being, whenever measures such as poverty are only meaningful in per capita terms (World Bank, 2005).

[139] It is insensitive to the degree of poverty and the distribution of income among the poor.

[140] It does account for the degree of poverty, but not for the distribution of income among the poor.

provide a very straightforward evaluation of poverty: half the observations in the sample have values smaller than it, and half have values greater than it.[141]

Once the poverty measures are chosen, and the estimation and prediction procedures explained in Sections 4.4.1 and 4.4.2 are accomplished, the evaluation of the impact of nonfarm income in terms of poverty can be performed.

Next, the procedure for evaluating the impact of diversification strategies on the distribution of income among farmers will be explained.

4.3.3.2 Inequality

Inequality measures, as compared to poverty, make a more extensive use of the data since they include all the observations in the calculations regardless their poverty status. However, and like measurements of poverty, there are a number of options from which it is possible to choose. One of the most common measures is the Gini coefficient, which possesses many desirable properties.[142] As a way to evaluate the robustness of the results, the paper also includes the estimation of the Theil index, which unlike the Gini, is decomposable.

The first inequality measure can be formally obtained from the following expression:

$$Gini = 1 + \frac{1}{N} - \frac{2}{N^2 \mu_m} \sum_{i=1}^{n} m_i \bar{r}_i y_i$$

where N is again the weighted sample size, i subscripts households, y stands for the household's income per capita sorted in descending order, and m_i is the weight assigned to each household according to its number of members. The term $\bar{r}_i$ is the average rank of all

[141] For a more comprehensive analysis of the numerous existing measures of poverty and inequality see for example World Bank (2005).
[142] They are mean and population size independence, symmetry and transfer sensitivity.

the members in the household, and μ_m is the weighted average value of y. As the index approaches 1 the distribution becomes more unequal.

The second index of income inequality can be expressed as:

$$Theil = (\frac{1}{N})\sum_{i=1}^{n} m_i(\frac{y_i}{\bar{y}})\,log\,\frac{y_i}{\bar{y}}$$

which also includes the average income in the sample's $\bar{y}$.

The above explained poverty and inequality measures are computed for the following two states: 1) the actual state in which there are farmers that remain as pure farmers and farmers that do diversify out from the farm, and 2) the predicted, or counterfactual, state in which the pure farmers maintain its non-diversifying condition, but the diversifying farmers are predicted as pure farmers. It is expected that if nonfarm income helps reduce poverty and equalize income distribution among Bolivian rural households, the measures would be smaller in state 1 compared to state 2.

4.4 Data

This section describes the basics of the household survey data including the sample frame, the size of the relevant subsample for the study, the specification of the equations of the model, and some descriptive statistics regarding the variables included in it.

4.4.1 The household survey

The data used come from the database of the Program for the Improvement of Surveys and the Measurement of Living Conditions in Latin America and the Caribbean (MECOVI for its acronym in Spanish)[143], which is conducted by the INE (Bolivian Bureau of Census). It collects data on such diverse topics as income, expenditures, education, health, employment, food consumption, assets holdings and migration. The MECOVIs have been conducted annually since 1999, and they are available online until the 2002. Since each survey does not track the same households, and some key questions are not asked in each survey, we perform a cross-section analysis for the survey of the year 2000.

All the surveys provide a sampling framework that ensures they are statistically representative of urban and rural Bolivia, as well as at the regional level: The Andean region (*Altiplano*), the lower Andean region (*Valleys*), and the lowland plains of the Amazon Basin (*Llanos*). According to Reardon *et al.*(2000) different agroclimatic conditions create different incentives to diversify income. Likewise, the three regions are explicitly separated during the estimation.

The survey for 2000 comprises 4,857 sample units, i.e. households, from the urban and rural areas. We concentrate on the rural subsample of 2,108 households, with a total of 9,092 individuals.[144] Among them 919 are located in the *Altiplano*, 782 in the *Valles*, and 407 in the *Llanos*.[145] Finally, the households are distributed in 239 Primary Sample Units (UPM for its acronyms in Spanish) across the 9 Departments of the country.[146]

---

[143] The Program is executed by the World Bank (IBRD), the Inter-American Development Bank (IDB) and the United Nations Economic Commission for Latin America and the Caribbean (ECLAC), as well as specialized institutions or agencies in countries participating in the Program. Subsequently other donors, such as Canada, Denmark, Germany, Japan, Norway, Sweden, UNDP, USA, and the Soros Foundation, have supported the Program

[144] Rural area is defined in this work as population concentrations of 2000 or less.

[145] The MECOVI also classifies the households according to their location into two groups: rural populated centers and rural dispersed areas; from the 2108 rural households, 121 are located in the first one and, 1898 in the second one.

[146] The PSU or *UPM* is defined as the unit of sampling used for this survey. INE divides Bolivia into UPMs that correspond to an average of 130 households in the populated centers, and approximately 50 in the dispersed area. Based on the number of households to be surveyed, the number of UPMs to be included in the survey is then decided.

The sampling procedure in the MECOVI 2000 combines stratification by agglomerate population size and a two-stage sampling. By using these sampling methods, the survey design reduces *bias* due to the households that are included. Additionally, the stratification of the population into groups ensures that the sample is representative. The two-stage sampling procedure involves in the first stage a Probability Proportional to Size (PPS) technique for the selection of the UPMs where the probability of selecting a sampling unit is proportional to the size of its population (number of housing).[147] According to Hansen, M. H. *et al.* (1943), the use of this technique decreases the sampling variance and eliminates the bias that arises when the primary units are drawn with equal probabilities. In the second stage of the selection, the housings or Secondary Sample Units (USM for its acronyms in Spanish) were selected with equal probability using systematic sampling.[148] All this information is taken into account when the econometric estimation is conducted.

The data from the MECOVI survey of 2000, allows the classification of income diversification activities suggested by de Janvry *et al.*(2005) since it includes monetary income and income in kind for each household, earned from different sources during the preceding 12 months to the survey[149]. This work only consider households that have income coming exclusively from self-employment in agriculture (farm income) and households that have both farm and nonfarm sources of income simultaneously, as shown below:

[147] According to the MECOVI 200 Methodological Document, the number of housing to be considered in each UPM is 8 for the extended area and 16 for the dispersed area. Ibid.: pp. 14.
[148] Given that the UPMs are selected with PPS, each point in the data needs to be weighted by the probability of the households falling in the sample. Therefore, the sampling weights are the reciprocal of the probability of its selection, and are already included in the MECOVI database for each household. See more in INE (2000).
[149] The MECOVI 2000 was conducted between November and December 2000.

Table 10. Rural households classification according to income source diversification [a]

Only farmers	1,305
Farmers with nonfarm activities	655

[a] *Source*: Own calculations based on MECOVI Survey for year 2000, Bolivian National Statistic Institute (INE).

4.4.2 The equations specification

Four sets of variables have traditionally being used in the determination of income diversification among rural households across the developing world: demographic characteristics, human capital endowments, physical capital endowments, and those related to the environment where the household is located. Table 11 explains the actual covariates that are included in the two equations of the sample selectivity model explained in Section 4.4.1. The logic for the inclusion of such variables generally lies on the standard available literature on income diversification into nonfarm activities.

With respect to the selection equation, it is expected that male-headed households have much higher probability to remain as pure farmers if there exists unequal access to inputs and other resources (i.e. extension advice) that could translate into a higher crop income per hectare compared to female-headed households (Minot et al., 2006). In terms of the household's male labor force, it can be anticipated that the higher the number of male members, the higher the labor force that is available for income-generating activities in general, and in particular, for farm-related activities (Evans and Ngau, 1991). Regarding the number of kids of the household, Minot et al. (2006) explain that more dependents results in more mouths to feed, and in the presence of agricultural risks, this, in turn, reduce the likelihood of the household to remain as a pure farmer. Finally, the age of the household head is also included as part of the variables that explain the household composition; this variable is believed to be positively related to the probability of being exclusively a farmer, given that older heads are more experienced famers that could enjoy higher levels of productivity.

Table 11. Variables included in the model

Equation	Dependent variable	Independent variables
Selection equation	a) For farmers that do no diversify out from farm activities: $z_i = 1$ and $z_i^* < 0$ b) For farmers who diversify their income through nonfarm activities: $z_i^* > 0$ and $z_i = 0$	Household characteristics: Number of kids under 7 years old, number of male adults, age of the household head, and a dummy variable for households where the head is male. Human capital: Years of education of the household members seven years old and older, and those of the head of the household.[(a)] Household assets: Household's land endowment per capita and agricultural equipment, as the main rural assets. Contextual variables: Income of the social group of reference, distance between the household's locality and the nearest department's capital, and peer's effects regarding nonfarm income diversification.
Income Equation	Log-income of pure farmers	Household characteristics: Number of kids under 7 years old, number of male adults, age of the household head, and a dummy variable for households where the head is male. Human capital: Years of education of the household members seven years old and older, and those of the head of the household.[(b)] Household assets: Household's land endowment and agricultural equipment, as the main rural assets. Access to infrastructure: Dummy variables for the access to electricity and pipe-borne water. Contextual variables: Distance between the household's locality and the nearest department's capital.

[(a)(b)] The employment section in the MECOVI questionnaire is applied to all household members seven years and older; this is the minimum age from which work is considered significant.

In terms of human capital, Reardon et al. (2000) explain that households with higher levels of education tend to have better chances for nonfarm employment, enjoy higher wage earnings, and succeed in business, as well as to be more productive farmers. Taylor and Yunez-Naude (2001), however, show that even when the returns from schooling are high for farm and nonfarm work, evidence shows that education induces farmers to diversify into nonfarm activities; a negative association between education and the dependent variable of the selection equation is hence expected. Additionally, given that much of the decisions of the farmers are taken at the household level, the education of the head as well as of the whole household are both included in the equation.[150]

In terms of agricultural assets, two variables are used as proxies and are conveniently included in the selection equation: the value of land per-capita and that of agricultural equipment such as tractors. The former enters as a measure of land availability for agricultural production and the latter as relevant inputs that greatly improve farm productivity per hectare.[151] [152] Consequently, it is expected that both affect positively the likelihood of being exclusively a pure farmer.

It has been also shown that rural infrastructure, such as electricity and pipe-borne water, has an important effect on nonfarm income activities facilitating the establishment of self-employed businesses (see for example Batunde, 2009, and Escobal, 2001). Thus, we expect a negative association between those variables and the probability of being a pure farmer.

In terms of the explanatory variables that account for the characteristics of the surrounding environment where the farmer resides, the existence of a wealthier social group of reference, and hence a greater access to resources for income diversification, would be negatively associated with the likelihood of remaining as a pure farmer. Also, proximity to urban areas (i.e. nearest Department capital) would in turn increase the probability of nonfarm income diversification, and be as well negatively associated with the dependent

---

[150] The possibility of correlation between these two variables is formally tested during the estimation process.

[151] Using land value of the landholdings as a measure of agricultural assets reflects also land quality, Peterson (1986).

[152] Official data of the land market, i.e. number of transactions, prices, etc., it is not readily available for the Bolivia. Considering, however, that more than 50% of the total distribution (and redistribution) of land in the country needs to go through a process of titling or re-titling (*saneamiento*) (INRA, 2002), and that there exist capital and cultural constraints that prevent a free access to the land market, we treat household's per capita land variable as exogenous, with no reallocations due to participation in nonfarm activities.

variable in the probit equation.[153] In terms of the peer's participation in nonfarm employment, it has been shown that it has a large impact on an individual's ability to engage in such type of employment (Araujo et. al, 2004). Therefore, we define peers as all neighbors surrounding each household, or people living in the same dwelling[154]. However, as de Janvry *et. al.*(2005) explain, it is valid to expect that the peers' decision to diversify could be determined by the same factors determining any household's decision. In order to deal with this issue, we follow their approach and include the average of the exogenous variables included in the selection equation as peers. We deliberately exclude the *distance to the nearest capital's department* variable, since we can expect that it has a high correlation with the *household's distance* variable.

According to the literature in the case of sample selectivity, individuals should usually take into account the variables affecting the outcome equation (i.e. log income for pure farmers), when making the decision (whether to remain as pure farmers or not). It is desirable to have variables that determine the diversification decision, but that do not affect income.[155] These variables, which are loosely referred to as "instruments" since they represent exclusion restrictions with respect to income, make estimates more precise.

Regarding the log income equation, most of the variables included in the selection equation are also believed to explain the income level of the pure farmers. In terms of household characteristics, and based on Taylor and Yunez-Naude (2001), it can be expected that as the number of small children increases in the household, home consumption increases as well, which will cause a decrease in farm income. On the other hand, as the number of adult males increases, farm production and hence income are expected to increase. The variable

[153] The variable is not included in the data set. Instead, the distance between the locality where the farmer is geographically located and the nearest capital of department was obtained from the Bolivian Military Geographic Institute (IGM for its acronym in Spanish).
[154] We can think of a dwelling (*localidad*) in Bolivia as a village or small town.
[155] For more details see for example Wooldridge (2002) and Winkelmann and Boes (2006).

age of the head should also have a positive impact on farm income if, as explained, older heads are more experienced famers subject to more specialization (Minot et al., 2006). Finally, if there exits gender biased in the access to productive inputs, a dummy for male-headed households should have a positive sign.

Regarding human capital assets, evidence shows that households with higher levels of education are likely to be more productive farmers and have higher returns to farming. Evidence for Mexico and Kenya, for example, clearly reveal the positive effect of different levels of education on income coming from the production of staples.[156]Likewise, it is also expected that the other two family assets have a positive effect on income given their positive impact on farm production.

The variables representing access to infrastructure are also believed to be positively associated with farm income, since it has been shown they increase farm productivity of activities that are already being undertaken by rural farmers, i.e. electric operation of machines or agro-processing (Kooijman-van Dijk and Clancy, 2010). Finally, the access to the market, proxied by the distance to the nearest capital of Department, is expected to be as well positively associated with the income of the farmer given that it increases the size of the market where he can sell his products, and also increases the availability of inputs.

Finally, the usual dummy variables are included in both equations to differentiate the estimation among the different agroclimatic regions, and the ones in which there are also coca leaf crops.

[156] See more details in Evans and Ngau (1991) for Kenya, and Taylor and Naude (2001) for the case of Mexico.

4.4.3 Descriptive analysis

Table 12 presents summary data for the variables used in the estimation of the sample selectivity model, with a total of 1,696 observations for which complete data is available. The statistics are shown for the entire sample, as well as for each geographic region of the country. In terms of the demographic characteristics of the households, the table reveals there are no sizable regional differences. With respect to human capital, it shows no significant regional differences in terms of the average of the household's head years of education, which has an overall average level of four years. In terms of total number of years of schooling in the household, however, a typical family in the *Llanos* has three and four years more schooling in average than their peers in the other two regions, *Valles* and *Altiplano* respectively.

Table 12. Descriptive statistics, national and by region[a]

Variable	National				Highland				Valley				Lowplains			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Household Characteristics																
Number of children under age 5	0.7	0.9	0	5	0.6	0.9	0	4	0.7	0.8	0	4	0.6	0.9	0	5
Number of male over age 15	1.2	0.8	0	6	1.0	0.7	0	5	1.2	0.7	0	5	1.4	0.9	0	6
Age of head	46.8	16.1	11	98	48.6	16.9	13	98	46.2	15.7	11	90	44.6	14.8	15	90
Dummy for gender of the household head (1=male; 0=female)	0.9	0.4	0	1	0.8	0.4	0	1	0.9	0.3	0	1	0.9	0.3	0	1
Human Capital																
Number of years of schooling of the household head	4.1	3.8	0	17	4.1	3.9	0	17	3.8	3.8	0	17	4.4	3.5	0	16
Total number of years of schooling of the household	13.1	11.2	0	76	11.6	10.5	0	64	12.7	10.8	0	68	16.3	12.1	0	76
Household assets																
Land assets[b]	2,459	15,415	0.0	448,000	826.0	2,151	0.0	25,000	2,343	8,079	0.0	175,000	5,337	29,570	0.0	448,000
Other productive assets [c][d]	227.5	3,939	0.0	120,000	17.4	364.7	0.0	10,000	95.0	1,537	0.0	35,000	810.0	7,958	0.0	120,000
Total household income[e]	87.6	205.7	0.0	4,162	53.4	133.1	0.0	2,288	92.6	229.5	0.0	3,562	140.4	254.7	0.0	4,162
Access to infraestructure																
Dummy for access to electricity (1=yes; 0=no)	0.2	0.4	0	1	0.2	0.4	0	1	0.2	0.4	0	1	0.3	0.4	0	1
Dummy for access to pipe-borne water (1=yes; 0=no)	0.3	0.5	0	1	0.3	0.5	0	1	0.4	0.5	0	1	0.2	0.4	0	1
Contextual variables																
Income of the social group of reference[f]	7,514	6,323	546.5	37,123	4,706	3,915	546.5	20,926	7,657	5,847	585.8	30,385	12,285	7,514	3,093	37,123
Distance to the nearest's capital of Department[g]	84.0	55.1	7.5	469.7	68.7	37.0	16.3	192.8	76.5	43.7	7.5	196.8	125.4	76.3	12.7	469.7
N	1,696				682				625				389			

[a] *Source* : Own calculations based on MECOVI Survey for year 2000, Bolivian National Statistic Institute (INE).
[b] [c] [e] [f] Measured in US$. Dollar calculations made with the exchange rate of 2000, US$. 1 = 6.19 Bs.
[d] Value of agricultural equipment such as tractors.
[f] Squared of the total income of the surrounding households.
[g] Measured in miles.

The data also demonstrate the existence of regional differences with regard to various household assets. Comparing the value of three such variables included in the table (i.e. land, agricultural equipment and income), rural households living in the *Llanos* have the highest mean, which in all the cases is more than two times higher than the value in the *Valles* and *Altiplano* regions. The biggest asset-value regional difference comes from comparing agricultural equipment, and the region with the lowest values, concerning all three variables, is the *Altiplano* region. Finally, the standard deviation for all three variables is high, especially in the *Llanos*.

The statistics of the access to infrastructure variables confirm the overall low coverage levels existing in the countryside area in the country. In relative terms, households residing in the *Llanos* have the highest percentage of rural farmers with access to electricity, but the lowest in terms of access to pipe-borne water.

Finally, with regards to the contextual variables, or variables that describe the surrounding environment where the farmer resides, the data shows that the income of the relevant social group of reference for each farmer, and the distance to the nearest capital of Department, increase both in value as we move from the *Altiplano*, through the *Valles*, into the *Llanos* region.

4.5 Estimation Results

Table 13 shows the results of the model for the sample selection or decision equations in columns A, and for the log-income equation estimates in columns B. The first set of results for both equations, under heading (1), include all the explanatory variables described in Section 4.4.2, over the 1,696 sample observations, with 956 of them uncensored. The decision model estimates show that two household characteristics are significantly related to the probability of remaining as a pure farmer: the age of the head (negative) and the gender of the household head (positive). There exists a higher probability of remaining as pure farmer if the head of the household is a male, which is consistent with the fact that usually men own and inherit production inputs such as land, more frequently than woman. The results show that the probability that Bolivian rural households remain as pure farmers decreases as the head becomes older.

Next, Column 1A also shows that one of the two human capital proxies included is statistically significant; the probability of specializing in farming decreases as the head receives more schooling. This result confirms previous evidence that households with higher education are induced to shift from farm to nonfarm activities.[157] The results from the first decision model further indicate that land assets per capita and agricultural equipment are, as expected, positively associated with the probability of being a pure farmer; the magnitude of the marginal effect of both variables is almost negligible, though. Access to infrastructure variables (i.e. electricity and pipe-borne water) have a negative impact on the probability of remaining as a pure farmer, which is coherent with the existing evidence that such factors facilitate income diversification (i.e. establishment of self-employed businesses).

In terms of the contextual variables, there are three peer's covariates included in the selection equation: average age of head, average number of households where the head is male and, average value of agricultural equipment. At the household level, these variables are statistically significant in the selection equation, but not in the log income one. Two of the peer's are statistically significant, and the result from a joint significant Wald test confirms the relevance of their inclusion in the decision equation with a ${\chi^2}_{[3]}$ test statistic

[157] The results of the Variance Inflation Factors (VIFs) statistic show values in all cases smaller than 5 for both explanatory variables. We used a linear specification to obtain the estimated coefficients.

equal to 25.37, and a p-value < 0.001. These variables play the role of exclusion restrictions that assist in the identification of the income equation.

Another contextual variable included in the probit equation, heading (1), is the income of the farmer's social group of reference. According to the results, as the wealth of the social group increases, the probability of having farm work as the solely source of income decreases. This is consistent with the notion that as the income of the social network increases, the farmer has a greater access to resources to start a business or engage in other entrepreneurial activities. Finally, the regional dummies for the *Valles* and *Llanos* regions are negatively associated with the probability of remaining as pure farmer, confirming the existence of intercept regional differences in term of the propensity to nonfarm employment diversification.

Under heading (1), column (B) shows the estimates for the outcome equation, i.e. the log income equation, for the 956 pure farmers in the sample. Four variables significantly explain the log income levels of farmers that do not diversify: Human capital proxied by the total years of education of the household and the education of the head, the value of the land assets owned by the farmers and, whether the household has access to electricity. As expected, all of them are associated with higher levels of income. Likewise, the regional dummy variables are significantly positive, confirming the advantage of wellbeing that households living in the *Valles*, *Llanos*, and where coca leafs are produced, enjoy compared to the base categories used in each case.

Table 13. Sample selectivity model estimates

Variable	(1) Decision Model (Pure farmers=1) (A) Coefficient		z	Pure farmers Log Income Estimates (Selection corrected) (B) Coefficient		z
Household Characteristics						
Number of children under age 5	-0.013		-0.33	0.021		0.45
Number of male over age 15	0.058		0.94	0.101		1.32
Age of the household head	-0.012	***	-4.31	0.004		1.17
Dummy for gender of the household head (1=male; 0=female)	0.471	***	4.02	0.156		0.92
Human Capital						
Number of years of schooling of the household head	-0.069	***	-5.68	0.034	*	1.81
Total number of years of schooling of the household	-1.86E-03		-0.44	0.015	***	2.77
Household assets						
Land assets				7.57E-06	**	2.40
Land assets per capita	5.99E-05	**	2.29			
Other productive assets	9.64E-05	*	1.68	5.00E-06		0.46
Access to infraestructure						
Dummy for access to electricity (1=yes; 0=no)	-0.398	***	-4.46	0.440	***	3.21
Dummy for access to pipe-borne water (1=yes; 0=no)	-0.407	***	-5.38	-0.020		-0.19
Contextual variables						
Income of the social group of reference	-2.06E-05	***	-3.06			
Distance to the nearest's capital of Department	-4.98E-04		-0.72	1.05E-03		1.21
Average age of head of the social group of reference	0.015	**	2.22			
Average number of households in the group of reference where the head is male	1.571	***	5.01			
Average value of productive assets of the social group of reference	4.90E-06		-0.25			
Regional dummy variables						
Valles	-0.412	***	-4.89	0.342	***	3.13
Llanos	-0.317	***	-2.90	1.188	***	9.29
Coca leaf production region	0.074		0.71	0.395	***	3.14
Constant	-0.85	*	-1.81	1.30	***	5.37
Log likelihood	-2507.04					
F (10,130)						
Wald chi2	322.14					
p-value	<0.001					
Residual standard error	1.218					
Lambda	0.487					
rho (ε1, ε2)	0.400	**	2.549			
LR (chi2) test of rho (ε1, ε2)=0	3.57					
p-value	0.059					
N	1,696					

Significance levels: *10%, **5%, ***1%

Table 13. Sample selectivity model estimates (Cont.)

	(2)				(3)			
	Decision Model (Pure farmers=1)		Pure farmers Log Income Estimates (Selection corrected)		Decision Model (Pure farmers=1)		Pure farmers Log Income Estimates (Selection corrected)	
	(A)		(B)		(A)		(B)	
Variable	Coefficient	z	Coefficient	z	Coefficient	t	Coefficient	t
Household Characteristics								
Number of children under age 5								
Number of male over age 15			0.111	1.62			0.054	0.82
Age of the household head	-0.011 ***	-4.48	0.004	1.15	-0.012 ***	-3.96	0.004	1.06
Dummy for gender of the household head (1=male; 0=female)	0.505 ***	4.94			0.471 ***	3.83		
Human Capital								
Number of years of schooling of the household head	-0.074 ***	-7.02	0.040 **	2.32	-0.082 ***	-5.99	0.027	1.52
Total number of years of schooling of the household			0.015 ***	2.85			0.018 ***	3.31
Household assets								
Land assets			8.13E-06 ***	3.74			9.84E-06 ***	3.58
Land assets per capita	6.76E-05 **	2.46			7.88E-05 *	1.94		
Other productive assets	1.01E-04 *	1.74			8.56E-05 *	1.89		
Access to infraestructure								
Dummy for access to electricity (1=yes; 0=no)	-0.364 ***	-4.24	0.472 ***	3.59	-0.363 ***	-2.65	0.395 *	1.91
Dummy for access to pipe-borne water (1=yes; 0=no)	-0.421 ***	-5.72			-0.456 ***	-4.93		
Contextual variables								
Income of the social group of reference	-2.17E-05 ***	-3.26			-2.13E-05 *	-1.90		
Distance to the nearest's capital of Department			1.20E-03	1.42			1.40E-04	0.10
Average age of head of the social group of reference	0.012 *	1.87			6.40E-03	0.70		
Average number of households in the group of reference where the head is male	1.555 ***	5.04			1.721 ***	4.32		
Average value of productive assets of the social group of reference	-4.54E-06	-0.23			-5.13E-06	-0.26		
Regional dummy variables								
Valles	-3.953 ***	-4.90	0.368 ***	3.50	-0.413	-3.37	0.481 **	2.53
Llanos	-0.356 ***	-3.53	1.204 ***	9.71	-0.421	-2.26	1.260 ***	6.53
Coca leaf production region			0.373 ***	3.06			0.547 ***	3.54
Constant	-0.76	-1.63	1.469 ***	7.71	-0.484	-0.76	1.534 ***	6.51
Log likelihood	-2514.37							
F (10,130)					17.970			
Wald chi2	317.06							
p-value	<0.001				<0.001			
Residual standard error	1.200				1.152			
Lambda	0.399				0.384			
rho (ε1, ε2)	0.333 **	2.356			0.334 *	1.833		
LR (chi2) test of rho (ε1, ε2)=0	4.200							
p-value	0.040							
N	1,696				1,696			

Significance levels: *10%, **5%, ***1%

In the first model there is evidence of sample selectivity even after the model controls for a number of observable characteristics of the households. The rho coefficient is significantly different from zero with a Likelihood Ratio test statistic of 3.57, and a p-value of 0.059. The positive sign of the coefficient explains that the sample is positively selected, or that the expected log income of pure farmers in the sample exceeds the value in the population. The use of the selection correction model is justified.

Table 13, i.e. columns 2A and 2B correspond to the sample selectivity model estimates that result from a restricted model, where explanatory variables from columns 1A and 1B, with p-values > 0.40, have been drooped. Wald tests of the joint significance of those variables were run, and the chi-squared distributed resulting statistic of 1.29 (p-value of 0.864) for the log income equation, and 2.14 (p-value of 0.829) for the decision equation, widely support the approach adopted. The resulting estimates from both equations, including the Likelihood Ratio test on rho, remain consistent and significant at conventional levels in the restricted model.

As explained in Section 4.5.1, the dataset comes from the 2000 MECOVI survey, in which multistage sampling procedure was implemented. As is customary in such cases, it is necessary to properly identify, and control for, the survey design and data collection procedure, in order to avoid point estimates that may be biased, and standard error estimates that may be misleading. The resulting estimates that incorporate the appropriate survey design are contained in columns 3A and 3B. In terms of the selection equation estimates, all but one of the peer's variables and the regional dummy variables, maintain their statistical significance when compared to the previous two sample selectivity models. In the log income equation, one of the two human capital variables included remain significant. Other than that, the model retains the significance of the covariates at conventional levels, and the selection correction is still justified.[158]

4.5.1 Income's prediction and nonfarm income impact on poverty and inequality

Having estimated the income of the rural farmers that do not diversify, it is straight forward to proceed with the prediction of the income of the farmers that actually diversify into nonfarm activities, as if they were pure farmers, following the method explained in Section 4.4.2. Based on those predictions, on the actual income of the diversifying farmers, and on the actual income of the pure farmers in the sample, comparisons can be made between the two following states of the rural farmers in terms of poverty and inequality: 1) the actual state in which there are farmers that remain as pure farmers and farmers that do diversify out from the farm, and 2) the predicted, or counterfactual, state in which the pure farmers maintain its non-diversifying condition, but the diversifying farmers are considered pure farmers. Table 14 shows the results. The table includes four poverty indicators to evaluate the impact of nonfarm income on poverty levels of the rural area. The poverty line used to determine which households' income falls below it, and hence is considered to be poor, was taken from the dataset, and put in terms of US dollars using the exchange rate prevailing at the time of the survey.

Comparing the two column (a) results, the Foster-Greer-Thorbecke coefficients suggest that there would be a higher percentage of rural residents being poor, or that the incidence of poverty would increase (85.97%% compared to 82.04%), that on average the poor's income shortfall would increase or that the depth of poverty would be higher (68.83% compared to 60.07%), and that the severity of poverty would be higher as well (58.84 compared to

---

[158] The peer's effects have also maintained their joint significance, with a $\chi^2_{[3]}$ test statistic equal to 6.77, and a p-value < 0.001.

49.12), if diversifying farmers would be modeled as non-diversifying. The median income value among the poor in the actual state compared to the predicted state, confirms these negative results. There is a 35% decrease in the value.

Table 14. Poverty and Inequality impact of nonfarm income

	State in which there are pure farmers and diversifiers (*Actual*)				State in which all rural farmers are hypothesized to be pure farmers (*Predicted*)			
	(a)	(b)	(c)		(a)	(b)	(c)	
Variable	Coefficient	Coefficients with weights [c]	95% confidence interval (Sample design adjusted)[d][e]		Coefficient	Coefficients with weights[f]	95% confidence interval (Sample design adjusted)[g][h]	
Poverty [a]								
Headcount ratio (%)	82.04	87.54	85.40	89.68	85.97	88.96	86.95	90.98
Poverty gap ratio (%)	60.07	65.59	62.44	68.74	68.83	73.26	71.00	75.53
Squared poverty gap	49.12	54.36	50.85	57.87	58.84	63.78	61.22	66.34
Median income of the poor (US$)	6.98	6.98			4.52	4.52		
Inequality								
Gini index	0.70	0.68	0.64	0.72	0.73	0.71	0.65	0.77
Theil index [b]	1.12	1.01	0.73	1.29	1.47	1.25	0.75	1.76
N	1,960	1960[i]			1,960	1960[j]		

[a] The poverty line was set at US$. 38.43 (Bs. 237.8817), based on the value reported in the data set. The same poverty line is used for both scenarios, the actual and the counterfactual.
[b] The entropy measure.
[c] [f] Coefficient estimates includes the required weights, i.e. household size.
[d][g] Following Howes and Lanjouw (1998), poverty measures correct standard errors for the stratification and multistage sample design of the MECOVI 2000. Concisely, it includes 9 strata and 148 Primary Sample Units.
[e][h] Following Jolliffe and Krushelnytskyy (1999), inequality measures include bootstrap estimates of the sampling variances for the stratified and multistaged sample design of the MECOVI 2000.
[i] Twelve observations are being dropped from the calculations since their value is zero.
[j] Final number of observations used in the inequality calculations was 1834, number for which there was complete information.

The values of the coefficients just discussed are in concordance with various poverty measures reported for the country.[159] Since the unit of analysis of the present work is, however, the household, and the income used to calculate poverty is measured in per capita terms, it becomes necessary to account for the size of the households in the calculations to obtain correct estimates. The resulting coefficients are reported in columns (b). Also, the use of survey data to measure poverty rates for the country as a whole gives rise to sampling errors. Measures of the standard deviations and confidence intervals are, as a consequence, required to be computed and reported for the poverty indices estimates. Finally, considering the multistage sampling procedure used in the survey data, it is often suggested to use sample weights to adjust the observations accordingly in order to get the

[159] See for example CEPAL (2006).

appropriate estimates.[160] The final estimated results are contained in columns (b) and (c). As the estimates show, all the FGT indices increase once again between the *actual* state and the *predicted* state, confirming the positive impact of nonfarm income on poverty (87.54 compared to 88.96, 65.59 compared to 73.26, and 54.36 compared to 63.78).[161] The bootstrapped standard errors for the poverty estimates are small and the 95 percent confidence intervals do not overlap for two of the indices; therefore, we can say with some confidence that poverty, as measured by the Poverty Gap and the Squared Poverty Gap, would have risen if nonfarm income diversification would have not existed. Nonfarm income helps decrease the severity of poverty, but it is not enough to significantly reduce the number of the poor.

Figure 5 shows the cumulative distribution of the actual and predicted incomes for the pure farmers and the diversifiers. Even though statistically not significant, it makes evident that a greater percentage of the farmers who diversify out from the farm work, but are predicted to be pure farmers, fall below the poverty line compared to the situation in which their actual income is used. The farthest northeast curve overlaps the cumulative distribution of the actual and predicted incomes for the pure farmers in the sample, since both are exactly the same.

The results in terms of the impact of income diversification on the poverty measures can also be observe through Figure 6. The discrepancy between the predicted incomes, in which all the farmers are assumed not to diversify into nonfarm activities, and are represented by the hollow circles, and the actual incomes that form the solid line, becomes wider as income

160 The sample has 146 Primary Sampling Units (PSU) and 9 strata (one for each department in the country).

161 When using data from a household survey, the resulting estimates are sample results that can be easily used to make inferences about the whole population, if each observation in the sample is appropriately weighted (World Bank, 2000). The weights are the total number of observation in the population that each observation in the sample represents. The sampling procedure in the MECOVI 2000 combined stratification by agglomerate population size and a two-stage sampling. The corresponding weights are part of the available data for each household.

increases. This gap highlights the positive and significant effect of nonfarm income in reducing the depth of poverty with reference to the adopted poverty line. To the right of the interception between this line and the one that plots the actual incomes, the scatter points below the poverty threshold represent the households that would have been considered poor under the predicted state, result that was not found to be statistically conclusive, though.

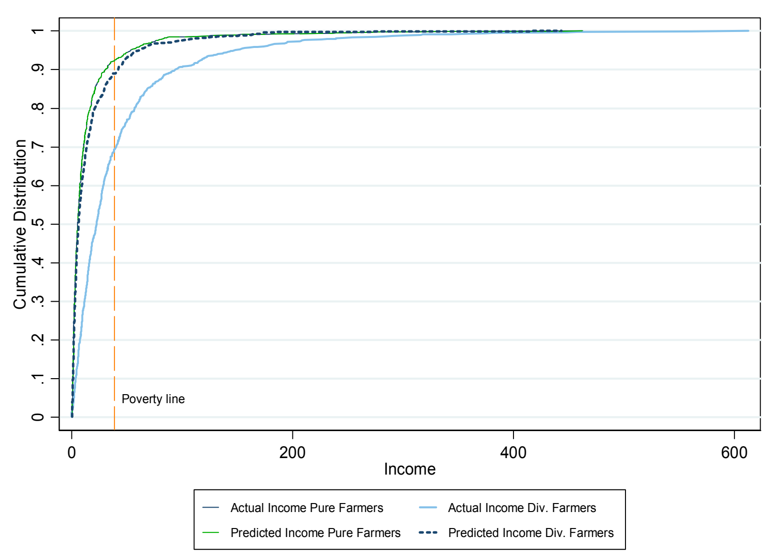

Figure 5 – Cumulative income distributions

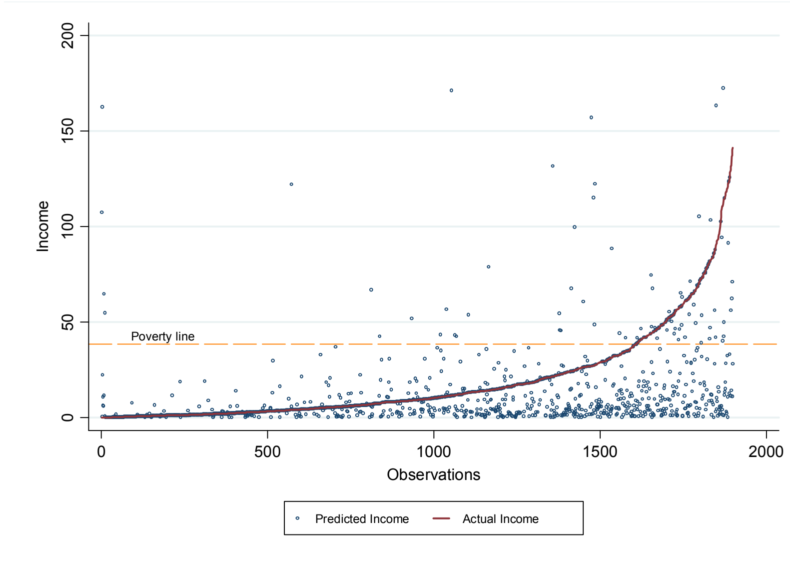

Figure 6 – Poverty effects of nonfarm income diversification

Table 14 also reports the results of the two inequality measures explained in Section 4.3.3.2, for the same two set of results: the actual and the predicted states. In terms of both estimates, the income distribution among rural farmers would be more unequal if farmers, who actually diversify income through nonfarm activities, would be modeled as if they would not. This is true for the coefficients in columns (a), as well as for the ones that include the already mentioned number of household member's weights (columns b). The bootstrapped standard errors estimates that observe the multistage sampling procedure used, confirm the statistical significance of the inequality indices, this time, however, the confidence intervals do overlap. Figure 7 shows the Lorenz curves for the actual and predicted incomes, and confirms the further position of the latter with respect to the perfect equality situation.

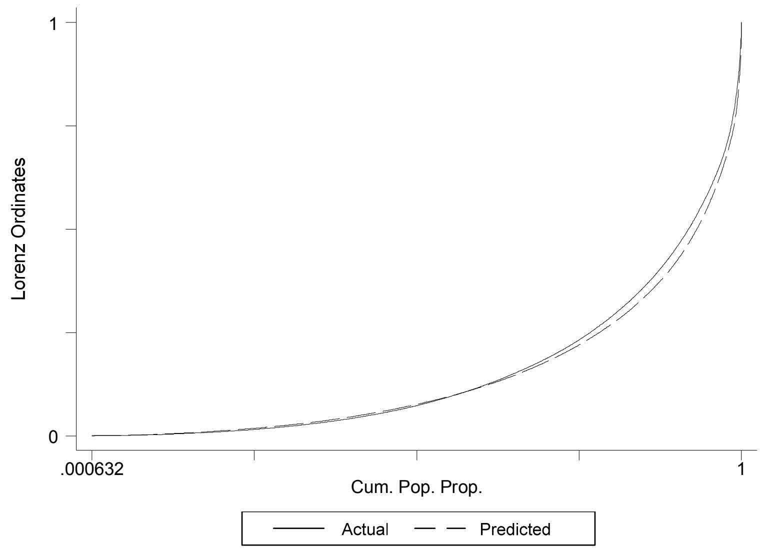

Figure 7 – Lorenz curves

4.6 Conclusions

It is well known that rural farmers in developing countries face extremely limiting condition, making agriculture a highly risky activity with low returns. Also, the existence of extremely high transaction costs prevents farmers, especially small farmers, have access to insurance. Access to credit markets is as well restrained for them, when the size of land ownership acts as collateral to access production financing. Additionally, it is also common that the distribution of endowments be highly unequal in poor countries, notably in terms of the distribution of primarily agricultural endowments such as land.

The most common strategy rural famers undertake to overcome market imperfections, and the low returns from the farm, is to develop sources of income that are not positively correlated with farm income.

Poverty and inequality, on the other hand, are among the most worrisome issues in the developing world. Since nonfarm income has been increasing its importance among its rural households, a large number of empirical studies have been conducted to determine the impact of income diversification activities on those two welfare indicators, with different regional patterns found.

Rural Bolivia is precisely one the worst performers in Latin America in terms of poverty and income distribution, with the rate of participation of nonfarm income in the total household income around the average of Latin America (40%) with an increasing trend. Farm work, hence, remains for now as the most important source of income across rural households, and a large number of them still rely solely in it.

This paper has shown that the depth and severity of poverty in Bolivia would have been higher if diversifying farmers would have not engaged into nonfarm activities. The existence of this additional income source has significantly increased the livelihood of the rural famers, shrinking the gap with respect to the poverty line. The results suggest, however, that the increase in income was not sufficient to abate the number of poor, which level would have statistically remained the same.

In terms of inequality there was no evidence of a significant effect. The indices of inequality computed for the observed incomes of pure farmers and the diversifying ones, and for the predicted incomes where the diversifying farmers are modeled as pure farmers, show an increase confirming the proposed hypothesis. Nevertheless, the overlapping of the estimated confidence intervals restrains us from drawing a statistically valid conclusion about the positive effect of nonfarm income on overall inequality, as measured by the Gini and Theil indices. In any case, it is clear that nonfarm income in Bolivia is innocuous as far as inequality is concerned, at least thus far. But if the nonfarm activities in which the poor are involved remain as the ones with low returns, as can be inferred from the results in poverty, the trend could be reversed.

In terms of policy, our results confirm the need to look beyond self-employed farm work when thinking in policies aimed to promote rural development in the developing world. According to our findings nonfarm activities represent an income source that increases overall income in the rural area, and decreases the gap between the income of the poor and the minimum level it is established as necessary to avoid poverty. Secondly, the likelihood that entry barriers may limit the access to income diversification activities of poor farmers, calls for direct public efforts to reduce those constraints by increasing the assets of the poor, e.g. human capital. This would not just facilitate diversification into the nonfarm sector, but would also allow the poor be able to access the high-returns nonfarm sectors. Income from agricultural activities will also benefit from this.

REFERENCES

Adams, Jr. R. (1991). The Effects of International Remittances on Poverty, Inequality and Development in Rural Egypt. Research Report 86. International Food Policy Research Institute, Washington, DC.

Adams, Jr. R. (2001). Nonfarm income, inequality and poverty in rural Egypt and Jordan. The World Bank Poverty Reduction and Economic Management Network: Poverty Division, Washington DC.

Adams, Jr. R. (2004). Remittances and poverty in Guatemala. World Bank Policy Research Working Paper 3418, Washington D.C.

Adams, Jr. R., A. Cuecuecha, and J. Page (2008). Impact of remittances on poverty and inequality in Ghana. World Bank Policy Research Working Paper 4732, Washington D.C.

Adams, R. (1995). Sources of Income Inequality and Poverty in Rural Pakistan. Research Report 102 Washington, DC: International Food Policy Research institute.

Adams, R. (2001). Nonfarm income, Inequality, and Poverty in Rural Egypt and Jordan. Policy Research

Adams, R. (2002) Nonfarm Income, Inequality and Land in Rural Egypt. Economic Development and Cultural Change 50(2): 339-363.

Aitchison, J. and C. G. G. Aitken (1976). Multivariate binary discrimination by the kernel method. Biometrika 63 (3): 413 - 420.

Andersen, L. (2002). Rural-urban migration in Bolivia: Advantages and disadvantages. Institute for Socio-Economic Research. La Paz Bolivia. (Manuscript).

Antezana, L.E. (1969). La Reforma Agraria campesina en Bolivia 1956-1960. Revista Mexicana de Sociologia, 31 (2), pp. 245-321.

Araujo, C., A. de Janvry and E. Sadoulet (2004). Peer Effects in Employment: Results from Mexico's Poor Rural Communities. Department of Agricultural and Resource Economics UCB, CUDARE Working Paper 991.

Babatunde, R. O. and Qaim, M. (2009). Patters of rural diversification in rural Nigeria: Determinants and impacts. Quarterly Journal of International Agriculture 48(4): 305 – 320.

Bardhan, P. K. (1989). The Economic Theory of Agrarian Institutions. Oxford: Oxford University Press.

Barraclough, S. and A. Domike (1966). Agrarian structure in seven Latin American countries. Land Economics, Vol. 17(4): 391-424.

Barret, C. B., T. Reardon, and P. Webb (2001). Nonfarm income diversification and household livelihood strategies in rural Africa: concepts, dynamics, and policy implications. Food Policy 26: 315-331.

Binger, B. and E. Hoffman (1998). Microeconomics with Calculus, Second Edition. New York: Addison-Wesley educational Publishers Inc.

Binswanger, H. P., K. Deininger and G. Feder (1995). Handbook of Development Economics, Volume III, Chapter 42, Edited by J. Behrman and T.N. Srinivasan, Elsevier Science B.V.

Binswanger, H.P. and M.R. Rosenzweig (1986). Behavioral and material determinants of production relations in agriculture. Journal of Development Studies, 22:503-539.

Brown, R. (1997). Estimating remittance functions for Pacific Island migrants. World Development, Vol. 25(4): 613-626.

Carter, M. R. (1984). Identification of the inverse relationship between farm size and productivity: An empirical analysis of peasant agricultural production. Oxford Economic Papers, 36, 131-145.

Carter, M.R. and J. Kalfayan (1989). A General Equilibrium Exploration of the Agrarian Question. U.W. Agricultural Economics Staff Paper 279, revised.

Carter, M.R. and K. D. Wiebe (1990). Access to capital and its impact on agrarian structure and productivity in Kenya. American Agricultural Economics Association 72(5): 1146-1150.

Carter, M.R. and May, J., (1999). Poverty, livelihood and class in rural South Africa. World Development 27(1):l-20.

Carter, M.R., D. Mesbah (1993). Can land market reform mitigate the exclusionary aspects of rapid agro-export growth? World Development 21 (7): 1085-1100.

CEPAL (Comisión Económica para América Latina y el Caribe)(2006). Anuario Estadístico.

Chinn, Dennis L. (1979), Rural poverty and the structure if farm household income in developing countries: Evidence from Taiwan. Economic Development and Cultural Change 27(2): 283-301.

Collier, P., Radwan, S. and Wangwe, S. (1986). Labour and Poverty in Rural Tanzania. Oxford: Clarendon Press, 1986)

Comisión Económica para América Latina y el Caribe (CEPAL) (2006). Anuario Estadistico. Web. May 2009.

Comision Economica para America Latina y el Caribe (CEPAL) 2000. De la urbanizacion acelerada a la consolidación de los asentamientos humanos en America Latina y el Caribe: El espacio regional. Santiago de Chile: Centro de las Naciones Unidas para los Asentamientos Humanos (Habitat).

Comisión Económica para América Latina y el Caribe (CEPAL) and Instituto Interamericano de Cooperación para la Agricultura (IICA)(2001). Panorama de la Agricultura de América Latina y el Caribe, 1990-2000.

Coudouel, A., J. Hentschel and Q. Wodon (2002). Poverty Measurement and Analysis, Chapter 1 in A Source Book for Poverty Reduction Strategies (PRS), Edited by Jeni klugman. Washington D.C.: The World Bank. Web. June 2009.

Crown, William H. (1998), Startistical models for the social and behavioral sciences: Multiple regression and limited-dependent variable models. Westport, Conn. : Praeger.

Davidson, R. and J. Mackinnon (2004). Econometric Theory and Methods. New York: Oxford University Press.

Davidson, R. and Mackinnon, J. (2004). Econometric Theory and Methods. New York: Oxford University Press.

Davis, B., Winters, P., Carletto, G., Covarrubias, K., Quinones, E., Zezza, A., Stamoulis, K., Bonomi, G. and DiGiuseppe, S. (2007). The Rural Income Generating Activities: A Cross Country Comparison. Background paper for the World Development Report 2008.

De Janvry, A. and Sadoulet, E. (1993). Rural Development in Latin America: Re-linking Poverty Reduction to Growth. In Including the Poor, ed. M. Lipton and J. van der Gaag. Washigton D.C.: World Bank.

De Janvry, A., F. Finan and E. Sadoulet (2005). Measuring the poverty reduction potential of land in rural Mexico. Journal of Development Economics 77: 27-51.

De Janvry, A., Sadoulet, E. and Zhu, N. (2005). The Role of Non-Farm Incomes in Reducing Rural Poverty and Inequality in China. Depactament of Agricultural & Resource Economics, UCB, CUDARE Working Papers, # 1001.

Dooley, D. and J. Prause (2004). The Social Cost of Underemployment. New York: Cambridge University Press.

Dorner, P. (1992). Latin American land reforms in theory and practice: a retrospective analysis. Madison: The University of Wisconsin Press.

Easterly, W. (2007). Inequality does cause underdevelopment: Insights from a new instrument, Journal of Development Studies 84: 755-776.

Echeverria, R., C. F. Jaramillo, J. Melmed-Sanjak, B. P. Reydaon, and F. Vogelgesang (1998). Perspectivas sobre mercado de tierras rurales en América Latina, Banco Interamericano de Desarrollo, Washington D.C.

Elbers, C. P. Lanjouw (2001). Intersectoral Transfer, Growth and Inequality in Rural Ecuador. World Development 29(3): 481-496.

Ermini, L. and Hendry, D. (2008). Log Income Vs. Linear income: An application of the Encompassing Principle. Oxford Bulletin of Economics and Statistics 70 (s1): 807-827.

Escobal, J. (2001). The Determinants of Nonfarm Income Diversification in Rural Peru. World Development 29(3): 497-508.

Evans, H.E., and Ngau, P. (1991). Rural-urban relations, household income diversification and agricultural productivity. Development and Change 22: 519-545.

Feder, Gershon (1985). The relation between farm size and farm productivity. Journal of Development Economics 18: 297-313.

Foster, J., Greer, J., and Thorbeckem E. (1984). A Class of Decomposable Poverty Measures. Econometrica 52(3): 761-766.

Gamba, P. and E. Mghenyi (2005). Rural Poverty Dynamics, Agricultural Productivity and Access to Resources. Food Security Collaborative Working Papers 55165, Michigan State University, Department of Agricultural, Food, and Resource Economics.

Gopal, R. and C. Tisdell (2002). Remitting Behavior of Nepalese Rural-to-Urban Migrants: Implications for Theory and Policy, Journal of Development Studies, Vol. 38 (3): 76-94.

Greene, L. and T. Seaks (1998). A hausman test for a dummy variable in probit. Applied Economics Letters, Vol. 5: 321-323.

Greene, W. (1998). Gender Economics Courses in liberal Arts Colleges: Further Results. Journal of Economic Education, Vol. 29 (4): 291-300.

Greene, W. (2008). Econometric Analysis, sixth ed. Upper Saddle River, N.J.: Pearson/Prentice Hall.

Gunning, J. W., J. Hoddinott, B. Kinsey, and T. Owens (2000). Revisiting forever gained: income dynamics in the resettlement areas of Zimbabwe, 1983–96. Journal of Development Studies 36 (6): 131– 154.

Hansen, M. H. and Hurwitz, W. N. (1943). On the theory of sampling from finite populations. The Annals of Mathematical Statistics, 14, 333-362.

Härdle, W., M. Müller, S. Sperlich, and A. Werwatz (2004). Nonparametric and Semiparametric Models. Berlin: Springer.

Hayfield, T. and J. S. Racine (2008). Nonparametric Econometrics: The np Package. Journal of Statistical Software 27(5).

Hoddinot, J. (1992). Modelling remittance flow in Kenya. Journal of African Studies, Vol. 1(2):206-232.

Hsiao, C., Q. Li, and J.S. Racine (2007). A consistent model specification test with mixed discrete and continuous data". Journal of Econometrics 140: 802-826.

Hurvich, C. M., J. S. Simonoff, and C. L. Tsai (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. Journal of the Royal Statistical Society Series B 60: 271-293.

Imai, K. and R. Gaiha (2002). Vulnerability, shocks and persistence of poverty: Estimates for semi-arid rural South India. Department of Economics Discussion Paper Series # 128, University of Oxford.

Instituto Nacional de Estadistica (INE) (2000). Documento Metodologico Encuesta Continua de Hogares ECH-MECOVI 2000. Web. July 2008.

Instituto Nacional de Estadistica (INE) (2008). Notas de Prensa No. 11, 34, 42, 60, 65, 71, 74 y 75. La Paz Bolivia. Web. July 2008.

Jiménez, W. and S. Lizárraga (2003). Ingresos y desigualdad en el área rural de Bolivia. Unidad de Análisis de Política Económica (UDAPE). La Paz: UDAPE.

Jiménez, W., S. Lizárraga (2003). Ingresos y Desigualdad en el Área Rural de Bolivia, Unidad de Análisis de Políticas Sociales y Económicas. Análisis Económico (UDAPE).

Keele, L. (2008). Semiparametric Regression for the Social Sciences. England: Wiley.

Kevane, M. (1996). Agrarian Structure and Agricultural Practice: Typology and Application to Western Sudan. American Journal of Agricultural Economics 78:236-245.

Kooijman-van, A. and Clancy, J. (2010). Impacts of Electricity Access to Rural Enterprises in Bolivia, Tanzania and Vietnam. Energy for Sustainable Development, doi:10.1016/j.esd.2009.12.004.

Lanjouw, P. (1999). Rural nonagricultural Employment and Poverty in Ecuador. Economic Development and Cultural Change, 48(1): 91.

Laserna, R. (2004). Coca y mercado en la economía campesina: Yungas y Chapare en Bolivia. Cochabamba Bolivia. Manuscrito.

Li, Q. and J.S. Racine (2004). Cross-validated local linear nonparametric regression. Statistica Sinica 14: 485-512.

Lokshin, M. (2006). Difference-based semiparametric estimation of partial liner regression models. The Stata Journal 6(3): 377-383.

López, R. and A. Valdés (2000). Fighting rural poverty in Latin America: new evidence of the effects of education, demographics, and access to land. Economic Development and Cultural Change 49(1): 197-211.

Lucas, R. (1985).Motivation to remit: Evidence from Botswana. Journal of Political Economy, Vol. 93 (5): 901-918.

Lucas, R. (1997). Internal Migration in Developing Countries, Handbook of Population and Family Economics, edited by M.R. Rosenzweig and O. Stark. London: Elsevier Science B.V.

Maddala, G. (1983). Limited-Dependent and Qualitative Variables in Econometrics, Cambridge: Cambridge University Press.

Malky, A. (2004). Sector Agropecuario en Bolivia: 1990-2004, Unidad de Analisis de Politicas Sociales y Economicas (UDAPE); downloaded from http://www.udape.gov.bo/diagnosticos/index/ by March 2008.

McCulloch, N. and B. Baulch (2000). Simulating the impact off policy upon chronic and transitory poverty in rural Pakistan. Journal of Development Studies 36(6): 100-130.

Minot, N., M. Epprecht, T.T. T. Anh, and L.Q. Trung (2006). Income diversification and Poverty in the Northern Uplands of Vietnam. Research report number 145. International Food Policy Research Institute, Washington, DC.

Monfardini, C. and R. Radice (2008). Testing Exogeneity in the Bivariate Probit Model: A Monte Carlo Study. Oxford Bulletin of Economics and Statistics, Vol. 70(2), pages 271-282, 04.

Morales, R., Choque, C. M., Galopo, E., Jemio, L. C., and N. Morales (2000). Bolivia: Geografía y Desarrollo Económico. Inter-American Development Bank: Washington, DC.

Muñoz, J. (1999). Los mercados de tierras rurales en Bolivia. Serie Desarrollo Productivo No. 61. Santiago: CEPAL.

Nadaraya, E. A. (1965). On nonparametric estimates of density functions and regression curves. Theory of Applied Probability 10:186-190.

Niimi, Y. and ç. Özden (2008). Migration and Remittances in Latin America: Patterns and determinants, in Remittances and Development: Lessons from Latin America, edited by P. Fajnzylber and J. H. Lopez. Washington DC: The World Bank.

Organización de Estados Amercianos (OEA). Secretaría General/INC. Proyecto Chapare: Propuesta de Estudio para el Desarrollo Integrado. La Paz, OEA, May 1975.

Racine, J. S. and L. Liu (2007). A partially linear kernel estimator for categorical data. Unpublished manuscript. McMaster University, Hamilton, Ontario.

Racine, J.S. (1997). Consistent significance testing for nonparametric regression. Journal of Business and Economics Statistics 15(3): 369-379.

Racine, J.S. (2008). Nonparametric Econometrics: A Primer. Quantile, Number 4, pp 7-56.

Ravallion, M. (2002). Poverty Comparisons: Harwood Fundamentals Of Applied Economics. Routledge.

Readon, T., Stamoulis, K., Balisacan, A., Cruz, M. E., Berdegue, J. and Banks, B. (1998a). Rural Non-farm Income in Developing Countries: Importance and Policy Implications, in The State of Food and Agriculture 1998. Rome: Food and Agriculture Organization of the United Nations.

Reardon, T. and J.E. Taylor (1996). Agroclimatic shock, income inequality and poverty: evidence from Burkina Faso. World Development, 24(4): 901-914.

Reardon, T. J. Taylor, K. Stamoulis, P. Lanjouw, and A. Balisacan (2000). Effects of non-farm employment on rural income inequality in developing countries: An Investment Perspective. Journal of Agricultural Economics, Vol. 51(2): 266-288.

Reardon, T., Cruz, M. E., and Berdegue, J. (1998b). Los pobres en el desarrollo del empleo rural no agrícola en América Latina: paradojas y desafíos. Invited Plenary Paper at the Tercer Simposio Latinoamericano de Investigación y Extensión en Sistemas Agropecuarios, Lima, 19-21 August.

República de Bolivia (2001). Bolivian Poverty Reduction Strategy (BPRE). La Paz 2001. Web. April 2009.

Rivero, F. and F. Mollinedo (2002). Diseno y construccion de los marcos de muestreo para las encuestas de hogares; documento presentado en el Taller #9 del Programa para el Mejoramiento de las Encuestas y la Medicion de las Condiciones de Vida en America Latina y el Caribe, auspiciado por la ONU. Lima Peru. Document can be downloaded from http://www.eclac.cl/deype/mecovi/taller9.htm

Robinson, P. M. (1988). Root-N consistent semiparametric regression. Econometrica, 56: 931-54.

Sanchez, V. (2006). The determinants of rural nonfarm employment and incomes in Bolivia. Michigan State University, Departament of Agricultural Economics.

Scott, C. D. (2000). Mixed fortunes: a study of poverty mobility among small farm households in Chile, 1968–86. Journal of Development Studies 36 (6): 155– 180.

Stark, O. and R. Lucas (1988). Migration, remittances and the family. Economic Development and Cultural Change, Vol. 36(3): 465-481.

Stiglitz, J.E. (1986). Economics of information and the Theory of Economic Development. NBER Working Paper No. 1566.

Tam, L. (1994). Rural-to-urban Migration in Bolivia and Peru: Association with child mortality, breastfeeding cessation, maternal care, and contraception. Demographic and Health surveys (DHS) Program, Working Paper # 8, Macro International Inc.: Calverton, MD.

Tannuri-Pianto, M., D. Pianto, and O. Arias (2004). Rural-Urban Migration in Bolivia: An Escape Boat? Brazilian Association of Graduate Programs in Economics in its series Proceedings of the 32th Brazilian Economics Meeting with number 120.

Taylor, E. (1999).The New Economics of Labour Migration and the Role of Remittances in the Migration Process, International Migration Vol. 37(1):63-88.

Taylor, E. and T. J. Wyatt (1996).The shadow value of migrant remittances, income and inequality in a household-farm economy. The Journal of Development Studies, Vol. 32(6): 899-912.

Taylor, E., R. Adams, J. Mora, and A. Lopez-Feldman (2005).Remittances, Inequality and Poverty: Evidence from rural Mexico. ARE Working Papers, University of California.

Taylor, J. E. and Yunez-Naude, A. (2001). Selectivity and the Returns to Scholling in a Diversified Rural Economy. Mimeo. University of California at Davis.

Thorp, R. (1998). Progreso, pobreza y exclusión: Una historia económica de América Latina en el siglo XX. Banco Interamericano de Desarrollo/Unión Europea, Washington DC.

Todaro, Michael (1995). Internal Migration in Developing Countries: A survey. In Reflections on Economic Development, Edited by Michael Todaro. England: Cornwall.

Unidad de Análisis de Política Económica (UDAPE) (2004). Objetivos de Desarrollo del Milenio en Bolivia: Quinto informe de progreso. Web. May 2009.

Unidad de Análisis de Política Económica (UDAPE) (2004). Sector Agropecuario Bolivia. La Paz: UDAPE. Web. May 2009.

United Nations (UN) 2004. World Urbanization prospects: The 2003 Revision. New York: Department of Economic and Social Affairs' Population Division.

United Nations (UN) 2008. World Urbanization prospects: The 2007 Revision. New York: Department of Economic and Social Affairs' Population Division.

Urioste, M. (2003). The abandoned Agrarian Reform: Valleys and High Plains. La Paz: Fundación Tierra. Web. April 2009.

Valdivia, J. (2004). Las micro-finanzas: un mundo de pequeños que se agrandan. Fundación Milenio-GTZ. La Paz- Bolivia.

Wang, M. C. and J. van Ryzin (1981). A class of smooth estimators for discrete distributions. Biometrika 68: 301-309.

Warriner, D. (1969). Land reform in principle and practice. Oxford: Clarendon Press.

Watson, G. S. (1964). Smooth regression analysis. Sankhya 26(15):359-372.

White, H. 1982. Maximum likelihood estimation of misspecified models. Econometrica 50: 1–25.

Wilde, J. (2000). Identification of multiple equation probit models with endogenous dummy regressors. Economics Letters, 69: 309-312.

Winkelmann, R. and Boes, S. (2006), Analysis of Microdata. Heidelberg: Springer-Verlag Press.

Winkelmann, S. and S. Boes (2006). Analysis of Microdata. Berlin: Springer-Verlag.

Wooldridge, J. (2002), Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press.

Working Paper No. 2572, The World Bank, March.

World Bank (1995). Bolivia: National land Administration Project. Washington, DC, United States: World Bank.

World Bank (1996). Bolivia poverty, Equity and Income: Selected policies for expanding earning opportunities for the poor. Report No. 15272-BO. Washington, DC, United States: World Bank.

World Bank (2005). Introduction to Poverty Analysis. Washington DC.

World Bank (2006). Global Economic Prospects: Economic implication of remittances and migration. Washington: The World Bank.

Wourtese, F. S. (2008). Migration, Poverty and Inequality: Evidence from Burkina Faso. Discussion Paper 00786 International Food Policy Research Institute (IFPRI), Washington D.C.

Yatchew, A. (1997). An elementary estimator of the partial linear model. Economic Letters 57:135-143.

Yatchew, A. (2003). Semiparametric Regression for the Applied Econometrician. Cambridge: Cambridge University Press.

Yörük, B. (2009). How responsive are charitable donors to requests to give? Journal of Public Economics, 93: 1111-1117.

APPPENDICES

Appendix A

Table A.1 Bivariate probit model: SUEST specification tests

		Coefficient	p-value
Testing jointly significance of regional dummy variables (Ho: regional dummies are all zero)			
Likelihood-ratio test	$\chi^2_{[6]}$	27.84	0.00
Testing equality of coefficients across regions			
Ho: coefficients in Valleys and Lowplains are equal			
Diversification equation	$\chi^2_{[10]}$	22.24	0.01
Remittances equation	$\chi^2_{[10]}$	8.38	0.59
Ho: coefficients in Highland and Lowplains are equal			
Diversification equation	$\chi^2_{[9]}$	20.31	0.02
Remittances equation	$\chi^2_{[9]}$	10.68	0.30

Appendix B

B.1) Household income maximization: The small farmers

Following Carter and Kalfayan (1989), the optimal choice problem faced by small farmers that are capital constrained and need to provide working capital by supplying part of their labor-time out from the farm, can be written as:

$$\max Y = pQ(L_f, I, T; z) - p_i I + w\,\Omega(L_s) \tag{b.1}$$

$s.t.$

$$w\,\Omega(L_s) - p_i I - S_0 \geq 0 \tag{b.2}$$

$$L_s + L_f = \bar{L} \tag{b.3}$$

$$L_s \geq 0; L_f \geq 0 \tag{b.4}$$

where we have a constrained maximization problem with an inequality capital constraint, and the nonnegativity constraints on the household labor allocation, that need to be solved using the Kuhn-Tucker conditions. Following Binger and Hoffman (1998) approach to solve this set-up, we introduce a set of new variables, also called *slack variables*, which ensure nonnegativity (g^2), and reformulate the inequality constraint to hold with equality (t^2). This allows us to write the Lagrangian function with two constraints and their corresponding multipliers (μ, τ) as:[162]

$$\mathcal{L} = Q(L_f, I, T; z) - p_i I + w\,\Omega(\bar{L} - L_f) + \mu[t^2 - (w\,\Omega(\bar{L} - L_f) - p_i I - S_0)] + \tau(L_f - g^2) \tag{b.5}$$

And the first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial L_f} = q_l - w\Omega' + \mu w\Omega' + \tau = 0 \Rightarrow q_l - w\Omega' + \mu w\Omega' = 0 \tag{b.6}$$

$$\frac{\partial \mathcal{L}}{\partial I} = q_i - p_i + p_i\mu = 0 \tag{b.7}$$

$$\frac{\partial \mathcal{L}}{\partial \mu} = t^2 - w\,\Omega(\bar{L} - L_f) + p_i I + S_0 = 0 \Rightarrow w\,\Omega(\bar{L} - L_f) + p_i I + S_0 = 0 \tag{b.8}$$

$$\frac{\partial \mathcal{L}}{\partial t} = 2\mu t = 0 \Rightarrow \mu = 0, t = 0, \text{or both} \tag{b.9}$$

[162] The price of the good produced by the farmers has been set equal to 1.

$$\frac{\partial \mathcal{L}}{\partial g} = -2\tau g = 0 \Rightarrow \tau = 0, g = 0, \text{or both} \qquad \text{(b.10)}$$

where the final expression of the first first-order condition (b.6) comes from the fact that even though τ can be either equal to zero or non equal to zero (equation b.10), if we assume an interior solution for family labor allocation $g \neq 0$, implying $\tau = 0$. Also, by the first-order conditions given above, we know that $\mu = 0, t = 0$, or they are both 0; however, given that we assume small farmers are capital constrained[163], the corresponding constraint is binding implying $\mu \neq 0$, or more specifically $\mu > 0$.[164]

B.2) Household income maximization: The large farmers

Once more the characterization of the maximization problem faced by the larger capital-constrained farmers follows Carter and Kalfayan (1989), where these agents are theorized to use their land endowment as collateral to obtain credit of size $\beta(T)$, for which they pay a financial cost r, to devote all their time to farm work, and to hire labor force which needs to be supervised. This optimization can be formally represented by:

$$\max Y = pQ(L, I, T; z) - p_i I - w\, L_d - r\beta(T) \qquad \text{(b.11)}$$

$s.t.$

$$L = L_f - \delta(L_d) + \gamma_0 L_d \qquad \text{(b.12)}$$

$$\beta(T) - p_i I - wL_d \geq 0 \qquad \text{(b.13)}$$

$$L_f = \bar{L}, L_d \geq 0 \qquad \text{(b.14)}$$

[163] Assumption also set for such farmers by Carter and Kalfayan 1989.

[164] It can be showed that based on the second-order conditions for a constrained maximization problem, $\mu > 0$ if the constraint is binding, Binger and Hoffman (1998), pp. 81.

Applying again the *slack variables* for the nonnegativity constraint (u^2), and in order to reformulate the inequality constraint to hold with equality (t^2), the Lagrangian function with two constraints and their corresponding multipliers (μ, τ) becomes:[165]

$$\mathcal{L} = Q(\bar{L} - \delta(L_d) + \gamma_0 L_d, I, T; z) - p_i I - \ wL_d - r(p_i I + wL_d) + \mu[t^2 - (\beta(T) - p_i I - wL_d) + \ \tau\ (L_d - g^2) \tag{b.15}$$

$$\frac{\partial \mathcal{L}}{\partial L_d} = q_l(\gamma_0 - \delta') - w - rw + w\mu + \tau = 0 \Rightarrow \ q_l(\gamma_0 - \delta') - w - rw + w\mu = 0 \tag{b.16}$$

$$\frac{\partial \mathcal{L}}{\partial I} = q_i - p_i - rp_i + p_i\mu = 0 \tag{b.17}$$

$$\frac{\partial \mathcal{L}}{\partial \mu} = t^2 - (\beta(T) - p_i I - wL_d) = 0 \Rightarrow \beta(T) - p_i I - wL_d = 0 \tag{b.18}$$

$$\frac{\partial \mathcal{L}}{\partial t} = 2\mu t = 0 \Rightarrow \ \mu = 0, t = 0, \text{or both} \tag{b.19}$$

$$\frac{\partial \mathcal{L}}{\partial g} = -2\tau g = 0 \Rightarrow \ \tau = 0, g = 0, \text{or both} \tag{b.20}$$

where the final expression of the first-order condition comes from the fact that even though τ can be either zero or non-zero (Equation b.20), if we assume an interior solution for the choice variables g$\neq$ 0, which implies that $\tau = 0$. Also, given the large farmers' first-order conditions, we know that $\mu = 0, t = 0$, or they are both 0; however, since we assume these agents are also capital rationed, the corresponding constraint is again binding, implying $\mu \neq 0$, or $\mu > 0$. This last result determines the final expression for the third first-order condition.

[165] Again here the price of the good produced by the farmers has been set equal to 1.

Appendix C

In order to be able to evaluate how the optimal values of the households' income equation are affected by changes in the parameters of the model, we just need to differentiate Y^* with respect to the required parameter. Since we are interested in the income effect of an increase in household's landholdings, we need to differentiate Y^* with respect to T in Equation (3.4) of section 0, and Equation (3.11) in section 3.3.2, for the small and large farmer's cases, respectively.

C.1) Small farmers

In order to accomplish the task, we make use of the fact that the optimal value of the Lagrangean function is simply the optimal value of the income equation (objective function) that is:[166]

$$\mathcal{L}({L_f}^*, I^*, \mu^*) = Y({L_f}^*, I^*) \tag{c.1}$$

So, Equation (3.4) can be rewrite in the following manner:

$$Y^* = Y\big({L_f}^*(w, p_i, \bar{L}, S_0, T; z), I^*(w, p_i, \bar{L}, S_0, T; z), T; z\big) \tag{c.2}$$

Which in terms of Equation (3.1) in Chapter 3, we get:

$$Y^* = Q^*({L_f}^*, I^*, T; z) - p_i I^* + w\,\Omega\big(\bar{L} - {L_f}^*\big) \tag{c.3}$$

Next, we differentiate Y^* with respect to T using the chain rule:

$$\frac{\partial}{\partial \mathrm{T}} \mathrm{Y}\big({L_f}^*, I^*, T; z\big) = q_{l_f} \frac{\partial L_f^*}{\partial T} + q_i \frac{\partial I^*}{\partial T} + q_T - p_i \frac{\partial I^*}{\partial T} - w\,\Omega' \frac{\partial L_f^*}{\partial T} \tag{c.4}$$

[166] See Binger and Hoffman (1998).

Which, rearranging terms, is equal to:

$$\frac{\partial}{\partial \mathrm{T}}\mathrm{Y}\left(L_f{}^*, I^*, T; z\right) = \; q_T + \frac{\partial L_f^*}{\partial T}\left(q_{l_f} - w\;\Omega'\right) + \frac{\partial I^*}{\partial T}(q_i - p_i) \tag{c.5}$$

which is precisely Equation (3.5) in section 0.

C.2) Large farmers

In this case we just need to follow the same methodology, and so the Lagrangean equation at the optimum choices of labor demand and non-labor inputs, is equal to:

$$\mathcal{L}(L_d{}^*, I^*, \mu^*) = Y(L_d{}^*, I^*) \tag{c.6}$$

Therefore, we can rewrite the income equation (Equation (3.11) in Chapter 3) as being ultimately a function of the model's parameters, or:

$$Y^* = Y(L_d{}^*(w, p_i, \overline{L}, r, T; z), I^*(w, p_i, \overline{L}, r, T; z), T; z) \tag{c.7}$$

Next, if we express Equation (3.7) in terms of its optimal value, we get:

$$Y^* = Q^*(L_d{}^*, I^*, T; z) - p_i I^* + wL_d{}^* - r(p_i I^* + wL_d{}^*) \tag{c.8}$$

We can now differentiate Y^* with respect to T, and using the chain rule finally obtain:

$$\begin{aligned}\frac{\partial}{\partial \mathrm{T}}\mathrm{Y}(L_d{}^*, I^*, T; z) = -q_l\delta'\frac{\partial L_d^*}{\partial T} + q_l\gamma_0\frac{\partial L_d^*}{\partial T} + q_i\frac{\partial I^*}{\partial T} + q_T - p_i\frac{\partial I^*}{\partial T} - \\ w\frac{\partial L_d^*}{\partial T} - rP_i\frac{\partial I^*}{\partial T} - rw\frac{\partial L_d^*}{\partial T}\end{aligned} \tag{c.9}$$

Which, rearranging terms, is equal to:

$$\frac{\partial}{\partial \mathrm{T}}\mathrm{Y}(L_d{}^*, I^*, T; z) = \; q_T + \left(q_l(\gamma_0 - \delta') - w(1+r)\right)\frac{\partial L_d^*}{\partial T} + (q_i - p_i(1+r))\frac{\partial I^*}{\partial T} \tag{c.10}$$

which is precisely Equation (3.12) in section 3.3.2.

Appendix D

D.1) Differencing estimator method

Due to its simplicity and versatility difference-based estimation has been used commonly to estimate a variety of settings of the semiparametric PLR model. The method, proposed by Yatchew (1997), consists of two steps. First, we primarily need to estimate the parametric variables by removing the nonparametric regression effect through a *differencing* procedure. Then, conventional smoothing methods can be applied to finally obtain the estimates of f.

Concisely, for the estimation of the function $f(t)$ in Equation (3.14), we first take first-order differences over the rearranged data, i.e. $t \leq \cdots \leq t_n$, which yields:

$$y_i - y_{i-1} = (x_{1i} - x_{1i-1})\beta_1 + \cdots + (x_{ki} - x_{ki-1})\beta_k + \big(f(t_i) - f(t_{i-1})\big) + (\varepsilon_i - \varepsilon_{i-1}) \qquad i = 2, \ldots, n \tag{d.1}$$

This procedure removes the nonparametric effect of t as the values of the x's become close, and as a result a consistent Ordinary Least Square (OLS) estimator of the vector β can readily be obtained.[167] It has been shown that the linear estimators $\hat{\beta}_1, \ldots, \hat{\beta}_k$ have a relative efficiency of 1/1.5 compared to the most efficient estimator, such as Robinson's (1988) method. The performance can be improved, though, if higher-order differences are applied

[167] The approximate sampling distribution of the parametric estimates, by convention denoted with $\hat{\beta}_{\text{diff}}$, is $\rightarrow N(\beta, \frac{1}{n}\frac{1.5\sigma_\varepsilon^2}{\sigma_u^2})$, where σ_u^2 is the conditional variance of x given t. Yatchew (1997).

(Yatchew, 2003)[168]. Note that for the error terms in Equation (d.1) to have the same variance as those in Equation (3.14), all the terms in the former need to be weighted by a factor corresponding to the order of the differencing applied.[169]

Having estimated the vector β, we proceed to subtract the parametric estimated effects from both sides of Equation (3.14), and obtain:

$$y_i - x_{1i}\hat{\beta}_1 - \cdots - x_{ki}\hat{\beta}_k = f(t_i) + (x_{1i}\beta_1 - x_{1i}\hat{\beta}_1) + \cdots + (x_{ki}\beta_k - x_{ki}\hat{\beta}_k) + \varepsilon_i \qquad i = 1, \ldots, n \tag{d.2}$$

expression that allows us finally to estimate f using a smoothing method, for which a variety of options exist.[170]

D.2) The differencing-based estimate results

We implement differencing estimation method which applies, by default, the local regression smoother *lowess* (a variation of *loess*) to estimate the nonlinear function f.[171] Table D.1 shows the resulting estimates for the parametric component. These differencing-based semiparametric estimates and the semiparametric Robinson's method estimates of Table 7, (located in the main body of this work) are overall qualitatively similar. The latter, however, provides a better fit of the data and returns standard error estimates that are consistently smaller. Table D.1 also reports the statistic for a significance test of the variable

---

168 It is possible to obtain an asymptotic efficiency estimator if higher-order differences are applied, for which samples that contains at least 30,000 observations are needed. For smaller sample sizes, higher-order differences might result in biased estimates, though. Lokshin (2006).

169 Yatchew (2003), provides optimal differencing weights up to order 100.

170 This is possible since $\hat{\beta}$ converges sufficiently quick to β. For details see Yatchew (2003).

171 The command by default applies the tricube weighting function, and a bandwidth of 0.8 (Lokshin, 2006)

land that enters the consumption specification nonlinearly. With a p-value less than 0.001, we have enough evidence to reject the restricted model in which $f(t)$ is a constant.[172]

Table D.1 Semiparametric PLR model estimates: Differencing estimates [a]

Variable	Coefficient		SE
Head's years of schooling	4.96	***	1.48
Household total years of education	2.68	***	0.35
Number of adults	2.80		3.05
Number of children	2.14		2.68
Livestock assest	-0.71	*	0.04
Distance to the nearest's capital of Department	-0.10	**	0.05
Valley (1=yes; 0=no)	33.68	***	5.73
Lowplains (1=yes; 0=no)	70.25	***	7.22
N	1374		
R^2	0.23		
Significance test on land: f is a linear function	9.01		
p-value	<0.001		

Significance levels: *10%, **5%, ***1%

[a] Order of differencing m=1

D.3) Regional differencing-based estimate results

In terms of the three agroclimatic regions, Table D.2 reports the differenced-based parametric results for each sub-population. The estimates have an overall similarity to the more efficient ones presented in Table 9. The smaller standard error estimates obtained with the Robinson's method, however, allow increases in the level of significance of some of the variables in the *Altiplano* and *Llanos* regions. When we compare the fit of both sets of regional estimates, the explanatory power of the ones resulting from Robinson's method are considerably higher for each region.

172 According to Yatchew (2003), the specification test statistic is equal to $V = (m * n)^{1/2} (s^2_{res} - s^2_{diff})/s^2_{diff}$, with $V \xrightarrow{D} N(0,1)$, where s^2_{res} is the variance of the residual under the null, and s^2_{diff} the variance of the residual under the alternative.

Table D.2 Regional Semiparametric PLR model estimates: Differencing estimates [a]

Variable	Highland		Valleys		Lowplains	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Head's years of schooling	-0.51	1.10	0.03	2.33	9.08 ***	2.62
Household total years of education	2.03 ***	0.31	4.95 ***	0.67	1.10 *	0.62
Number of adults	-0.43	2.90	-5.99	5.50	11.04 *	5.97
Number of children	2.94	2.28	-1.43	5.05	9.87 *	5.46
Livestock assest	0.08 *	0.05	0.21 ***	0.07	0.32 ***	0.10
Distance to the nearest's capital of Department	-0.16 ***	0.05	0.01	0.09	-0.26 ***	0.09
N	592		484		298	
R^2	0.19		0.23		0.26	
Significance test on land: is f a linear function	1.3		3.5		27.6	
p-value	0.091		<0.001		<0.001	

Significance levels: *10%, **5%, ***1%

[a] Order of differencing m=1

VITA

NANEIDA REGINA LAZARTE ALCALA

Candidate for the Degree of

Doctor of Philosophy

Thesis: ECONOMIC ANALYSIS OF SUBSISTENCE FARMERS IN BOLIVIA: THREE ESSAYS

Major Field: Economics

Biographical:

Personal Data: Born in Potosí, Bolivia, the daughter of Elia Alcalá Cruz and Jorge Lazarte Rojas. Married to Mauricio Subieta Benito and mother to Mateo Sebastian Subieta Lazarte.

Education: Completed the requirements for the Doctor of Philosophy in Economics at Oklahoma State University, Stillwater, Oklahoma in December, 2010; received the Master of Arts degree in International Cooperation and Management of Development Projects from the Complutense University of Madrid, Madrid, Span in 2002; received the Bachelor of Science degree in Economics from the Bolivian Catholic University, La Paz, Bolivia in 1997.

Experience: Instructor of Principles of Macroeconomics at the Economics Department, Oklahoma State University, Fall semester of 2008 and from August 2009 to August 2010; Teaching Assistant for Microeconomics at Oklahoma State University for the Spring 2008 and Spring 2009 semesters; Instructor of Introduction to Economics at Oklahoma State University for the Spring and Fall semesters of 2007; Financial Project Manager for the Bolivian National Transportation Agency from 2002 to 2003; Fellow Researcher for the United Nations Development Program from 2000 to 2001; Operations Manager for Prepaid Services for Millicom Cellular Internation in Bolivia from 1997 to 1999; Research Assistant for the Institute of Surveys and Studies of the Bolivian Catholic University during 1996.

Professional Memberships: American Economic Association

Name: Naneida Regina Lazarte Alcalá Date of Degree: December, 2010

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: ECONOMIC ANALYSIS OF SUBSISTENCE FARMERS IN BOLIVIA: THREE ESSAYS

Pages in Study: 139 Candidate for the Degree of Doctor of Philosophy

Major Field: Economics

Findings and Conclusions:

The first essay examines the role that remittances plays in income diversification strategies in the developing world. Using a large and nationally representative survey for Bolivia, we find that remittances alleviate production constraints and market failures that are commonly faced by rural farmers in agrarian economies. They represent an additional income source that relax credit constraints, and hence facilitate further diversification of rural households into other nonfarm activities. The results are based on an endogenous bivariate probit model where the probability of diversification is in part determined by the decision to remit.

The second essay demonstrates that the access to larger land endowments may not translate into income gains under conditions in which farmers may not have access to relevant rural markets, and hence are hypothesized to be confined to low (and even decreasing) marginal income values of land. Large farmers, on the contrary, freely access the market using their large land endowments and enjoy increasing marginal returns of the asset. A nationally representative survey data from Bolivia is primarily used to study the issue, and a semiparametric partially linear (PLR) regression model is applied to estimate the model's parameters. The results obtained confirm the derived marginal income effects of land for small farmers and large farmers, i.e. the relationship between land and wealth is nonlinear. The regional analysis carried out in this regard, ratifies the findings. Two estimation techniques were applied to verify the robustness of the results.

The third essay analyzes the impact of nonfarm income activities on the country's rural poverty and income inequality, using the MLE with a careful incorporation of exclusion restrictions that ensures the identification of the model. In order to differentiate the estimation among the three contrasting agroclimatic regions existing in the country, and the ones in which there are also coca leaves' cultivations, regional dummy variables are conveniently incorporated. The resulting consistent and efficient estimates from the sample selectivity model were used to predict a counterfactual income for rural households that diversify out from farm activities, as if they would be pure farmers. The paper concludes that nonfarm income in rural Bolivia has helped reduce the depth and severity of poverty, but the number of poor would have statistically remained the same. In terms of inequality there was no evidence of any significant effect.

ADVISER'S APPROVAL: Dr. Lee Adkins