UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Income recovery in urban and rural areas in Colombia

Abstract

This thesis investigates whether the household income recovery after a negative income shock, caused by a health or death shock, differs between urban and rural areas in Colombia. The hypothesis is that urban areas are financially worse off compared to rural areas after a shock due to differences in availability of formal and informal insurances. Uninsured households in urban areas are not covered by either a formal or informal insurance which thus makes them worse off compared to their rural counterpart. By using linear OLS regressions and the difference-in-difference-in-differences identification strategy this thesis was able to confirm the research question that there is a statistically significant difference in income recovery between urban and rural households, between those affected and not affected by the shock, before and after the shock occurred. The results also seem to confirm the hypothesis due to the insufficient performance of formal insurances in urban areas. This because the findings suggest that formal insurances are unable to compensate the loss of the informal insurance system as they cannot counteract the negative impact of the income shock. The results also show that regardless of area, poor households are more likely to experience a shock compared to wealthy households.

Authors:
LINDA HAMP
CLARA SEGERSTEDT

Supervisor:
SIMON SCHÜRZ

Bachelor Thesis in Economics (15hp)
Department of Economics
June, 2017

Contents

Chapter 1

Introduction

A working insurance system provides several benefits to a society as, for instance, more people can access better medical care when they have a health insurance to cover most of the costs. Insurance also brings security, as for instance in a development country, a farmer could receive some monetary compensation if his/her land is struck by disasters, such as drought or heavy rainfalls. These insurance systems can be provided in two separate ways, namely by formal contracts between an official institution and an individual, or by informal arrangements between individuals.

Formal insurance systems are defined as explicit attempts to create an insurance market where individuals trade in risk (Besley, 1995). Gains from such trade are made when individuals with different risk preferences trade with each other. Formal insurances are therefore provided by the government or private insurance firms, and insurance access is granted through official contracts. However, in almost all developing countries these services are provided by the government, either by directly funded national health services or by enforcing employers to finance the insurance (Pauly et al., 2006). These contracts are set up between the insurer and the insured and promises that a pre-specified monetary amount will be payed if an uncertain event is realized. The prospect of such contracts is thus limited by the ability to specify and enforce them.

Informal insurance arrangements exist due to an imperfect formal insurance market. They denote trade in risk between individuals in a community without the involvement of official agencies. The existence of these systems can be argued from two conflicting perspectives rooted in different beliefs of human behavior and interaction (Besley, 1995). First, that the trade is facilitated by altruistic feelings between members of a neighborhood or social class. Second, that the trade is motivated by self-interested individuals who expect reciprocal behavior sometime in the future. These two perspectives are linked with the possibility to socially enforce an informal insurance arrangement as an individual's risk behavior is thus modified by either norms of honesty or norms of reciprocity.

When comparing the level of formal and informal insurance between the urban and rural areas, it is likely that informal insurance arrangements do not exist in the same extent in urban areas as in rural areas. This would be because in urban areas there are less close-knit communities and higher anonymity than in rural areas which do not promote the creation of informal structures to the same extent as in rural areas. In turn the access to formal insurance in rural areas is likely to be much lower than in urban areas as contracts are less enforceable as distance increases.

The degree to which formal and informal insurances can be measured differs significantly. Informal insurances are difficult to capture due to the structure of social networks, while formal insurances are more traceable and reliable due to the realization of contracts. Since formal insurances can be measured in actual numbers, descriptive information about them can be collected through surveys and used for analysis. Based on the presumed distribution of formal and informal insurances in urban and rural areas, information regarding formal insurances will in this thesis be used to discuss the possible effects of informal insurances. Therefore, this thesis will study the effects on income caused by an exogenous shock, and discuss how formal and informal insurances contribute to income recovery in urban and rural areas.

There are various types of formal insurances an individual can buy, for example crop insurance, health insurance or life insurance, which all provide the insured with some pre-specified monetary amount if a specific event comes to pass. The formal insurances that are considered in this thesis are health insurance and life insurance. It can be argued that a well-functioning health insurance improves access to high quality health care and accordingly lowers health risks (Besley, 1995). Hence, health insurance improves an individual's quality of life as it protects a households' income and consumption possibilities. Life insurance on the other hand, can improve the quality of life for the other members of a household as it provides the household with a considerable monetary amount in the case of a household member's death.

Since health and life insurances will be of central interest in this thesis, the corresponding two types of shocks, namely health shocks and death shocks, will be considered. A health shock is when a member of a household suddenly and unexpectedly becomes physically ill, while a death shock is when a member of a household suddenly and unexpectedly dies. Health shocks are therefore disruptive to a household's income as labor supply decreases and health expenditures increase, which can force consumption to fall (Mohanan, 2013) and accordingly cause a negative income shock. Intuitively, death shocks have a more permanent damage to household income compared to health shocks, as the death of a household member results in a permanent loss of income. Baeza and Packard (2006) list health shocks as an essential field of study since they are persistent factors to why low-income households become poor. Lost income is therefore an important aspect of this thesis, since it will focus on investigating the difference in income levels after a health or death shock, and more importantly the income recovery effect for households with a health or life insurance.

1.1 Purpose and research question

The research question of this thesis is: *does the income of urban and rural households recover differently from an income shock caused by an exogenous health or death shock?*. The purpose of this thesis is hence to explore if there is a statistically significant difference in income recovery between urban and rural areas after they are exposed to an exogenous health or death shock. Our theoretical prediction is that this relationship varies between the areas as a result of different access to and level of formal health and life insurance.

The hypothesis is thus that the income recovery gap after a health or death shock, between those formally insured and uninsured in urban areas is larger than the corresponding gap in rural areas. This would mean that the formally uninsured in urban areas have a more negative income effect than those formally uninsured in rural areas. The idea is that those in rural areas that do not have a formal insurance are instead covered by an informal insurance arrangement. However, those without a formal insurance in urban areas would not have this advantage and would therefore be completely without any form of health or life insurance.

1.2 Scope

The main focus of this thesis is the difference in income recovery caused by a health or death shock in urban and rural areas. Additionally, the thesis briefly discusses the difference in the level of access to health and life insurances between the urban and rural areas. In order to make an accurate investigation and reliable comparison, this thesis does not contain any cross-national research. In fact, simply one country is chosen for the study, namely Colombia in South America. Colombia is classified as an upper middle-income country (The World Bank, 2015), which is suitable for this thesis since health insurance is widely adopted both in urban and rural areas. Life insurance is not as widely adopted as health insurance, but it is on the rise as premium sales are driven by newly formed life insurance arrangements (Beresford and Rubio, 2015). This is also suitable for this thesis since it means that our thesis can contribute to the collected knowledge on how such insurance affects household income. Since the investigation is not a cross-country study, differences such as political policies, culture or ethnicity, which could affect household income after a health or death shock, are not taken into consideration.

Our thesis is limited by two inherent problems with insurance systems. First, the existence of incomplete contracts with a persistent disability to enforce them, and second, the prevalence of asymmetric information where the risk seller has an information advantage compared to the risk buyer (Besley, 1995). In turn, asymmetric information takes two forms. There might be some pre-existing characteristics of the individual which makes him/her more prominent to risk (adverse selection). Once the insurance is acquired, the individual might change his/her behavior which makes him/her more prominent to risk (moral hazard). We therefore assume that the risk-taking behavior caused by the possession of a health or life insurance is equal in the target areas as well, and consequently this factor is not accounted for in our investigation.

In this thesis we make three assumptions regarding health and death shocks. First, that a health or death shock is sudden and unexpected. Second, that the fraction of cases where the health or death shock is in fact expected, is equal in both urban and rural areas. Third, that the shock is evenly distributed across age as the severity of an income loss might differ due to the age of the individual exposed to the health or death shock.

Chapter 2

Theory and background

Some pre-existing knowledge regarding the actual situation in Colombia in terms of insurance coverage and health care system is needed for essential understanding in this thesis. In addition, the effect of formal insurances on economic growth is key to fully grasp the importance of evolving insurance systems in developing countries. This is covered in this chapter, which also describes previous work that is informative within this field of study.

2.1 The health care system in Colombia

Colombia is a country with a radical development when it comes to its health care system. In 1990, only approximately 24% of the population had health insurance, and it was notably beneficial for the richer part of the population (OECD, 2015). In fact, 47% of the wealthiest population quintile had health insurance, but only 4.3% of the poorest quintile had any form of financial protection against physical illness. According to Alvarez et al. (2011) other problems Colombia faced in the early 1990's were inadequate distribution of the health care system in troublesome urban and rural areas, and a population growth larger than what the health system could cover. These problems resulted in a reformation of the health law, with the main goal of achieving universal health care (Chapman, 2016). The structural reform was called *Law 100*, and one of the significant changes that it entailed was that hospitals and health centers both in urban and rural areas were privatized, meaning that the public health systems were replaced by individual private health insurance systems. In addition, a primary plan for medical aid was conducted to ensure that insurance companies provided insurance holders with certain health care services (Alvarez et al., 2011).

Over the last 20 years, the reforms have resulted in a positive change for the Colombian health system. Some of the improvements that Colombia has experienced are for instance shorter waiting times for an appointment and increased health care service standards (OECD, 2015). An increase in access to health care in the poorest areas in Colombia has also had major positive effects on, inter alia, the health conditions of the inhabitants. Today, citizens pay less for health care, and free services have become more accessible. Furthermore, the effects of the reforms seem to favor those in rural and poor urban areas the most (Giedion and Uribe, 2009; Trujillo et al., 2005). OECD (2015) considers the health care system in Colombia to be well designed and states that the country is making progress towards universal health care. Despite the fact that 76.3% of the population lived in cities in 2014, the access to health care in Colombia is somewhat equalized.

Despite this positive development of the health system, there are several factors that still need improvement, such as the supply of medical aid and the overall quality of health care services (Giedion and Uribe, 2009). In addition, there are other problems to be resolved in Colombia to further optimize the health system. Colombia has over the last 40 years had an armed ongoing conflict that has resulted in numerous dreadful actions, such as kidnapping, homicide, sexual violence and multiple injuries, affecting the access to health care. According to OECD (2015), this conflict has affected rural areas more than urban areas.

Critics in regard to the health care system, such as Webster (2012), claim that the Colombian health system might not be as good in practice as it is in theory, and that in reality, inequalities and long queues still remain in Colombia today. Furthermore, he states that the health system is beginning to break down, and that the Colombian government has been accused of taking bribes. Others doubt Law 100, and consider the effects of the health system unclear, and that the system is still unbalanced and discriminating against the poor (Giedion and Uribe, 2009).

2.2 Formal insurance in Colombia

The Colombian insurance market is dominated by compulsory insurances, workers' compensation and life insurance (Beresford and Rubio, 2015). The compulsory insurances are over 50 in total, and they are frequently expanded by the Colombian legislators. These compulsory insurances affect different sectors such as motor liability, employers' liability and environmental liability, which cover any liability which may arise from negligence, accidents or environmental damage. Despite this mandatory legislation, evasion rates are high and for the employers' liability approximately 62% of the employers are not covered by an insurance if an employee gets injured at the workplace.

Foreign insurance companies have a strong foothold in the Colombian insurance market, and in 2014, approximately 58% of the life and non-life insurance market was dominated by foreign insurers (Beresford and Rubio, 2015). However, by legislation, life insurance can only be provided from regulated companies within the country. In order for foreign insurance companies to provide this they must therefore establish branch offices within the country and be approved and regulated by the Financial Superintendency.

In 2015, 97% of the Colombian population were covered by some form of formal health insurance arrangement, compared to the significantly smaller number of approximately 24% in 1990 (OECD, 2015). In 2014 the growth of premium sales by insurance companies was 8.9% and led by life insurance arrangements, meaning that more and more Colombians increase their insurance coverage (Beresford and Rubio, 2015).

Beresford and Rubio (2015) state that a major issue for the Colombian insurance market is to raise their reputation as many Colombians view insurance as a luxury only available to high income households. They further predict that the market will continue to benefit from an increased competition due to new market entries which will stimulate economic growth. This would mean that the insurance market has not yet reached market equilibrium where there is a balance between insurance supply and demand.

New company entries can thus still improve the current market as market competition tend to lower prices, something that seems crucial for changing the current beliefs of insurances being a luxury.

2.3 Effect of formal insurances on economic growth

A key economic theory is that an improvement in health increases an individual's productivity and thus also increases his/her income (Bardhan and Udry, 1999). In turn, higher productivity leads to higher economic growth. Hence, a well-functioning formal health insurance system is important for developing countries wishing to increase their economic growth.

Furthermore, a well-functioning insurance system can be used to reduce income risk, meaning that a household's income would be negatively affected by some exogenous shock. Insurance thus smooths the household's consumption over time, meaning that if faced with some hardship a household can still to some degree consume the same amount as before (Dercon, 2000). This is especially important in the case of life insurances as a household's permanent income loss can be somewhat recovered through a formal insurance plan. Furthermore it is often the poorest households who are left without any form of insurance, meaning that a negative income shock would to a very large extent be passed on to the household's current consumption (Morduch, 1999). Targeting these households with an improved access to health and life insurance would therefore also increase their spending possibilities as insurance protects current consumption.

Imperfect insurance markets are also an inefficient use of resources which might yield a misallocation of resources within a country. This in turn lowers returns and may lower the overall investment rate as risk aversion tend to lead to inefficient investments (Banerjee and Duflo, 2005). Improving such insurance markets is thus important for economic growth as investments might increase.

2.4 Previous work

A former study performed by Wagstaff (2007) examines the economic effects of health shocks in Vietnam. The conclusion of this work shows that with respect to income, urban households are more vulnerable when exposed to a health shock than rural households. He also finds that health shocks affect income with a larger negative effect in urban households in comparison to rural households. The reason for this differentiation between urban and rural households, Wagstaff (2007) argues is due to the fact that rural areas are better at adjusting labor supply if one family member dies or cannot contribute to the workforce anymore. He also looks at how health shocks affect medical spending, which he claims to be largely dependent on whether the household has a formal insurance. He finds that the medical spending if a household member is hospitalized is larger for those uninsured than insured. One thing to take into account however, is that formal insurances can largely influence how household income is affected by health shocks. Wagstaff (2007) describes the health insurance coverage as limited in the study conducted in Vietnam 1993-1998. In addition, the health insurance scheme used in the investigation was introduced 1993, and therefore relatively new when the data was collected. In contrast, and as mentioned in section 2.2, 97% of the Colombian population had health insurance in 2015 (OECD, 2015). This thesis will therefore contribute to Wagstaff's work by investigating the income recovery from health and death shocks in urban and rural areas in Colombia, and then based on these results and insurance statistics, analyze the effect of health and life insurances.

There is an endogeneity problem when studying the causal effect of health or death shocks on economic outcomes as health and wealth are often correlated. One way to avoid this is done by Mohanan (2013) who studies the exogenous health shock of injuries from bus accidents on a specific route in India. This is then compared to the control group who travel the same bus route but is unexposed to the shock. He finds that shock-affected households were able to smooth their consumption in terms of food and housing. However, households were only able to pay their health-care related bills by borrowing, meaning that the main effect of the health shock was increased debts. However, one year after the shock Mohanan (2013) was unable to find any difference between the treatment and control groups in terms of labor supply, meaning that both groups had the same on average monthly labor income. Although not measured by Mohanan (2013), it is reasonable to assume that a death shock might lead to similar household behavior, meaning that if the bus accident led to someone's death the household would have to increase its debt in order to smooth its consumption. Our thesis will contribute to this study by acknowledging the endogeneity problem that comes with studying health, and by extending the use of a health insurance. However, as previously mentioned, we assume that these problems are normalized as we compare the urban and rural areas. The fact that borrowing is important for households if exposed to a health shock is also acknowledged, but our thesis will only focus on household income and instead add the importance of health and life insurances. Further studies could perhaps include a factor of household debts.

Townsend (1995) does not test variations between urban and rural areas, however, he shows that there is a significant variation within and between rural villages in Thailand in the terms of their informal structures. The village closest to Bangkok in the sample deviates from the others as this village has integrated with the cash economy of the urban areas. However, despite being in a rural area the village seem to lack any internal informal insurance arrangements. The non-existence of an informal system can therefore be explained by the proximity to the urban areas. This proximity thus makes the village a less close-knit community which in turn makes the enforcement of informal insurance arrangements more difficult (Banerjee and Duflo, 2005). The other villages vary in terms of the extent of the informal arrangements, but they still display some tendencies of common risk-sharing arrangements (Townsend, 1995). As stated before, informal insurance arrangements are not measured in this thesis, however, they are still highly relevant when analyzing the results. Further studies might also want to replicate the study made by Townsend (1995) and measure variation withing rural areas, especially concerning their distance to urban areas.

Chapter 3

Methodology

This chapter covers the practical issues of the empirical study conducted in order to answer our research question, *does the income of urban and rural households recover differently from an income shock caused by an exogenous health or death shock?*. All decisions and strategic choices throughout the process are accounted for here. The data is described, such as its origin and the specific variables analyzed, followed by a brief discussion regarding the limitations the data has. Finally, some summary statistics are presented, describing the distribution of the observations.

3.1 Approach

In order to test our research question we exclusively used linear OLS regressions, however, these regressions were divided into two groups with different approaches and interpretations. First, we have what we from here on will refer to as a naive OLS regression (see equation 3.1), to test the individual effect of a health or death shock on households in urban and rural areas. Then, we used the difference-in-difference-in-differences (DDD) identification strategy (see equation 3.2) to answer the research question, namely to investigate the difference in income recovery between urban and rural areas caused by a health shock or death shock. The dependent variable Y in these equations represents monthly household income measured in Colombian pesos (COP) throughout this thesis. Shock and time are dummy variables, the first equal to one if a household has experienced a shock, and the latter equal to one for the later time period (meaning the time period after the shock). Note that time is hence a dummy and not a continuous variable. The naive OLS also includes a vector for a set of controls as illustrated by **X'**.

With the first naive OLS regression we examined the causal effects of a shock on monthly household income. We performed this regression with two different shock variables: health shock and death shock. Furthermore, this regression was run on data collected from two different periods of time in order to compare to what degree households were affected by the shock. In addition, we investigated differences between urban and rural areas, in both time periods. By running these regressions, we were able to see how the shock affected the households' monthly income in the two areas and in two different time periods.

The reason we refer to the first linear OLS regression(s) as naive is because of the assumptions that have to be made in order to apply the OLS regression technique. First, the errors have to be normally distributed, which we found out, that in this case they were not. This matter was resolved by taking the logarithm of monthly household income, as can be seen in all regressions throughout this thesis. Second, as mentioned in section 1.2, we have assumed that the shock is exogenous and that there is zero correlation between the regressors and the error term. This means that poor and rich people with the same probability experience a health or death shock and there is therefore no correlation between unobserved variables and monthly household income.

$$log(Y) = \alpha_0 + \alpha_1 Shock + \mathbf{X'}\beta + U \tag{3.1}$$

Although the naive OLS regression is useful in order to investigate the effect on monthly household income caused by a shock, it does not illustrate a significant *difference* in the effect after a shock between urban and rural areas. Furthermore, there is no guarantee that the shock in fact is exogenous, which can cause an endogeneity problem. In order to avoid the potential endogeneity problem and to provide an answer to our research question, the DDD identification strategy composes a good method of investigation. It provides the opportunity to examine how the different areas are affected by a shock in relation to each other. Thus, the DDD allowed us to test two regions (urban/rural), in two time periods (before/after the shock), between those that did and did not experience a shock. The DDD estimator is displayed in equation 3.2 and the regression in 3.3. In equation 3.2, the delta represents the treatment effect (the effect of the shock) and is the main coefficient of interest. t_1 and t_0 are the two time periods, t_0 being a time period before the shock occurred, and t_1 being a time period after the shock has occurred. As with the naive OLS, the DDD includes a vector of controls illustrated by $\mathbf{X'}$, illustrated in equation 3.3.

$$\begin{aligned}\hat{\delta} = {} & [(\bar{Y}_{t_1} - \bar{Y}_{t_0})^{Urban,Shock} - (\bar{Y}_{t_1} - \bar{Y}_{t_0})^{Urban,No\ shock}] - \\ & -[(\bar{Y}_{t_1} - \bar{Y}_{t_0})^{Rural,Shock} - (\bar{Y}_{t_1} - \bar{Y}_{t_0})^{Rural,No\ shock}]\end{aligned} \tag{3.2}$$

$$\begin{aligned}log(Y) = \alpha_0 + \alpha_1 Time + \alpha_2 Area + \alpha_3 Shock + \alpha_4 (Time * Area) + \\ + \alpha_5 (Time * Shock) + \alpha_6 (Area * Shock) + \delta(Time * Area * Shock) + \mathbf{X'}\beta + U\end{aligned} \tag{3.3}$$

In order to clarify the effects of including area in the regression, we additionally did a regression using a difference-in-differences (DD) identification strategy which consists of the same variables and relationships as the DDD, but excludes area (see equations B.1 and B.2 in Appendix B). We thus compared the difference in income recovery between households that have experienced and not experienced a shock. Correspondingly, a significant interaction term between time and shock tells us that the effect of a shock on monthly household income differs depending on time. The results from the DD are not included in chapter 4 since they do not answer our research question. However, they will be mentioned briefly in the evaluation of the results in chapter 5, and are therefore included in Appendix B.

An alternative approach to using the DDD would be to use two DD identification strategies, one for each area. However, in contrast to the DDD, these would not control for area fixed effects. All single variables in the DDD contain observable and unobservable characteristics correlated with household monthly income, controlled for as fixed effects. This means that we control for variables that are not included in our study but might still correlate with income. Accordingly, the DDD was considered the best choice.

When using the DD and the DDD identification strategies one needs first to control for the OLS assumptions and then the parallel time trend assumption. The parallel time trend assumption states that absent the treatment (shock), both groups (areas) would exhibit parallel time trends in the dependent variable (monthly household income). This would give weight to the fact that any conclusions drawn from the study were due to the shock and not some other time trend which is unaccounted for.

3.2 Data

The data used for our analyses was retrieved from Encuesta Longitudinal Colombiana (ELCA) conducted by the Universidad de los Andes in Bogotá, Colombia. ELCA is a survey that was constructed with the aim of establishing a panel database of Colombian households during twelve years, with new data gathering every three years (Universidad de los Andes, n.d.). ELCA has previously been used for instance by Fernández et al. (2014) to test how the labor market in rural Colombia cope with violent shocks. Iregui-Bohórquez et al. (2016) have also used the ELCA survey to test the relationship between health status and labor participation.

In this thesis we used the two rounds that have been published this far (June 5, 2017), namely 2010 and 2013. Since the data is collected at household level between two different areas we have throughout all regressions used clustered standard errors at household level. This is necessary as some events or shocks might affect groups of households within one area in the same way. For example, a health shock might affect multiple households in a rural village if the cause of the shock is an epidemic related to the cattle which several households tend. Accordingly, we assume independence across the regions (sick households in rural areas do not affect households in urban areas), but we allow for some correlation within a region (sick households within rural areas might affect other households within the rural areas).

3.2.1 Organization of data and variable description

When we received the data it was originally divided into several different datasets, split into the categories *Households*, *Shocks* and *People* for each area and year. The original Spanish names can be found in table A.1 in Appendix A. We combined all datasets into one, using household id as key. The *Households* datasets contained data regarding for instance the number of people in a household, whether a household had a health or life insurance, etc. The *Shocks* datasets contained information about what kind of shocks a household had experienced, affected household members, etc. Finally, the *People* datasets contained information about the household head, such as age and sex, which were used as control variables. Since all data was originally in Spanish, the first step was to translate and identify key variables that would be used in the analyses in this thesis. The key variables can be found in table A.2 in Appendix A, together with the original Spanish names of the variables. In addition, the shock variables in the *Shocks* datasets needed adjustment, and their change of structure can also be seen in table A.2.

Our dependent variable, monthly household income measured in Colombian Pesos (COP), was a bit troublesome to organize with the given data. As it turned out, the variable that represented monthly household income from labor (which was of most interest for us), was missing in the dataset called *Rural Households 2013*. Instead, the rural questionnaire for 2013 included two new income variables: income from agricultural labor and income from non-agricultural labor. Our first assumption was that the variable for income from labor from 2010 had been divided into two separate income variables in 2013. Therefore, we added the income from agricultural labor and income from non-agricultural labor together in order to get a representative income variable to compare to 2010. However, we were not sure that this would be a good representation of the income variable, since the different structure of the income variables in rural areas 2013 could imply that the distribution of income had been measured differently than in the other datasets. Apart from income from labor, other income variables were income from pensions, income from leases, income from interests or dividends and other income. Therefore, we created a new income variable in which we added all income variables together in 2010 and 2013 respectively. This allowed us to also use this representation of the income variable to try and receive comparable results. In the end, we decided to analyze our results based on the income variable that included all variations of income as we believe it to be a more credible measure. The reason we believe so is because if the income variables have been defined differently in 2013 compared to 2010, combining all income variables reduces the chance of mismatching labor income. However, another way to reconstruct the household monthly income would be to look at household expenses, either in comparison to the present income variables or as a standalone dependent variable. Due to time constraint this was not possible for our study, but could for future research be an interesting approach.

Health shock is in the datasets defined as "accident or illness of a household member that prevented him/her from performing his/her daily activities", and death shock as the "death of the head of household or the head's spouse". There was an additional death shock variable in the ELCA datasets that could have potentially been used (death of other household member(s)), but we decided that the death of the household head or the head's spouse would be the most destabilizing shock for the household and thus give a greater effect on the loss of income.

However, the formulation of the shock question in the survey differs between the 2010 and 2013 round. In 2010 the ELCA survey asked whether the household had experienced any of these shocks within the last twelve months. In 2013 households were asked whether they had experienced any of the shocks between 2010 and 2013. The fact that the time frame differs between the rounds must therefore be taken into account when evaluating the effect of the shocks.

Furthermore, the ELCA survey was designed to exclude single-households and to not register a household member over 65 years old as the household head. The reason for not registering older household heads was due to the fact that they are nearing the end of their work cycle and might not yield interesting income results in twelve years time (Centro de Estudios Sobre Desarrollo Economico - CEDE, 2016). This means that all households that have experienced a death shock where a household head has died, has lost a household head younger than 65 years old.

Regarding the insurance variables, the health insurance variable was missing in the datasets from 2010. Accordingly, we had to assume that if a household had a health insurance in 2013, they also had one in 2010. For our analyses we handled the life insurance variable in the same way as the health insurance variable.

The urban areas were defined as the following regions: Atlántica, Oriental, Central, Pacífica and Bogotá, and the rural areas were defined as: Atlántica Media, Cundi-Boyacense, Eje Cafetero, and Centro-Oriente. In both areas these regions were in turn divided into subareas to ensure a proper distribution of observations (Centro de Estudios Sobre Desarrollo Economico - CEDE, 2010).

Additionally, our regressions included several control variables, namely sex of household head, age of household head, number of people in the household, and whether the household had a health or death shock in the twelve months before 2010. Whether the head is male or female is of interest because targeting female headed households in poverty reliving efforts might yield better results than targeting male headed households (Morrison et al., 2007). For us this would mean that perhaps a health or death shock would affect a household's income differently depending on the sex of the household head. The age of the head might affect household income because an individual is only expected to provide income to the household during certain years of his/her life. Once an individual enters the labor market, income tends to increase with working experience, which an older person would have more of than a younger one. The number of people in the household might affect household income as more people can contribute to (if they are old enough), but are also dependent on, the collective income. The final control variable regards whether the household had experienced a health or death shock in the last twelve months before 2010. The idea is hence that if a household has already experienced a shock, it can already be disadvantaged compared to non-shock households. All regressions in this thesis were run first without control variables, and then with these four controls. However, when running the naive OLS for 2010 it does not make sense to add a control variable for those who had a shock in 2010, since that variable is already included as an independent variable. Thus, only three control variables (sex, age and the number of people in the household) where used for this specific regression.

In general, some overall adjustments had to be made to the observed households in the datasets. First, there were duplicates of household identification numbers in the datasets containing observations from 2013. This was due to divisions of households where one (or more) members of the household moved and consequently formed their own household. This statistical issue was dealt with by simply removing all split households from the datasets. The number of split households were relatively few (118 households) compared to the fraction of households that had remained the same. Therefore, we concluded that removing split households from the datasets would not affect the outcome of our results.

Another factor to account for was that some households moved between the target areas between 2010 and 2013. To be able to conduct a fair analysis where the two areas are compared, this thesis only considers households that remain in their urban or rural areas in both years. Households that have moved between areas were therefore removed from the datasets.

As with all panel data surveys there is a possibility of households not wishing to complete all rounds, and a possibility of adding new households in later rounds that were not the original subjects of the survey. In this study these households were consequently also removed.

3.2.2 Limitations with the data

In our monthly income variable, we chose not to include income from money aid. The main reason for this is that income from money aid is a bit unstructured in the ELCA datasets, and we were not able to track why households received money aid, or by whom the aid was provided. Furthermore, this variable did not exist in the *Households* dataset for rural areas in 2010, which left us without the possibility to compare the income from money aid between 2010 and 2013. We were thus unable to test whether households that had experienced a shock received extra money aid between 2010 and 2013 or not.

Another variable which we might have wanted to include in our data is the collection from insurance policies over the last twelve months. The main reason why we did not include collections from insurance policies was that the number of observations were extremely few (27 in total for both years and areas combined).

As previously mentioned, only formal insurance data has been collected in the ELCA survey. The extent to which the households might have informal insurance arrangements will in this thesis therefore be evaluated based on the regression results and the statistical information regarding the level of formal insurance arrangements. Further studies might try to collect information about the informal insurances in order to fully explain the relationship between formal and informal insurance in certain areas. However, as mentioned in chapter 1, informal insurance is hard to capture.

In section 3.1 we mention the parallel time trend assumption as an important assumption for the DD and DDD identification strategies. One way to argue that the parallel assumption is fulfilled is to compare the two groups (urban/rural areas) in earlier time periods and observe no significant differences in their time trends. Any differences in the next period would thus be due to the treatment. For this thesis, however, no such data was available as we have used the first and the second round of the ELCA survey. This might therefore affect our conclusions as there is a possibility that other time-trending factors have been obtained and measured in the model.

3.2.3 Summary statistics

The sample size was, after the removal of missing cases and inconsistencies, 16,532 observations for both years combined. Since households that had moved between areas, or that did not participate in the survey both years, were removed, we had 8,266 households that completed the survey questions of interest in both 2010 and 2013. The overall statistics can be seen in table 3.1. There might be some power issues as this can be considered quite a small sample to be representative of the whole Colombian population. However, we are confident that the sample to some extent can reflect the population and that our analyses therefore are valid. Of course, generalizations must be made with care in regards to this sample size.

When looking at each area separately in table 3.1 we have the same number of households in for example the urban areas in 2013 as in 2010 due to the elimination of household inconsistencies. The sample size of the urban areas was thus for each year 4,373 observations (52.90%), and for the rural areas 3,893 observations (47.10%). That is, our sample was almost evenly divided between the urban and rural areas. In 2013, 3,967 households in total replied to the question regarding health or life insurance. Only 353 (8.90%) of the households replied that they had a health insurance, and 849 (21.40%) households claimed they had a life insurance.

As illustrated in table 3.1, 3,157 (38.19%) out of 8,266 households said they had experienced a health shock between 2010 and 2013. In the 2010 round, 1,307 (15.81%) out of 8,266 households replied that they had experienced a health shock within the twelve months of the survey in 2010. For the death shock between 2010 and 2013, 182 (2.20%) households had experienced a death shock where their household head or the head's spouse had died. In the twelve months before 2010, 36 (0.44%) out of 8,266 households had the same death shock. The remaining control variables contain the 2010 values, 6,002 (72.61%) out of 8,266 households had a male household head, the head's average age was approximately 44.6 years old, and the households consisted on average of 4.4 people.

Table 3.1: Summary statistics

Variable	**Mean**	**Std. Dev.**	**N**
Urban	0.5290	0.499	8266
Health insurance	0.0890	0.285	3967
Life insurance	0.2140	0.41	3967
Health shock 2013	0.3819	0.486	8266
Health shock 2010	0.1581	0.365	8266
Death shock 2013	0.0220	0.147	8266
Death shock 2010	0.0044	0.066	8266
Male	0.7261	0.446	8266
Age	44.567	12.122	8266
Household size	4.3687	2.003	8266

When sorting this by area, the summary statistics can be found in table 3.2 for the urban areas. 171 (6.87%) out of the 2,489 urban households that answered the insurance question had health insurance, and 678 (27.24%) households had life insurance. Furthermore, 1,648 (37.69%) out of 4,373 households in total had experienced a health shock between 2010 and 2013. On the contrary, 670 (15.32%) of the households had experienced a health shock in the twelve months before the survey in 2010. Between 2010 and 2013, 88 (2.01%) households had experienced a death shock, and in the twelve months before 2010 this number was 17 (0.39%) of the households. The remaining control variables in 2010 had the following values, 2,814 (64.35%) of the households had a male household head, the average age of the household head was 43.5 years old, and a household consisted on average of 4.2 people.

Table 3.2: Summary statistics in urban areas

Variable	**Mean**	**Std. Dev.**	**N**
Health insurance	0.0687	0.253	2489
Life insurance	0.2724	0.445	2489
Health shock 2013	0.3769	0.485	4373
Health shock 2010	0.1532	0.36	4373
Death shock 2013	0.0201	0.14	4373
Death shock 2010	0.0039	0.062	4373
Male	0.6435	0.479	4373
Age	43.517	11.968	4373
Household size	4.2045	1.995	4373

For the rural areas the summary statistics are illustrated in table 3.3, and as can be seen, 182 (12.31%) out of 1,478 rural households had health insurance, and 171 (11.57%) households had life insurance. Between 2010 and 2013, 1,509 (38.76%) out of 3,893 households had experienced a health shock, and in the twelve months before 2010, this number was 637 (16.36%) of the households. For the death shock between 2010 and 2013, 94 (2.41%) had experienced the death of their household head or the head's spouse. In the twelve months before 2010, 19 (0.49%) of the households had experienced the death shock. The remaining control variables for 2010 have the following values, 3,188 (81.89%) households had a male household head, the average age of the household head was 45.7 years old, and the household consisted on average of 4.5 people.

Table 3.3: Summary statistics in rural areas

Variable	Mean	Std. Dev.	N
Health insurance	0.1231	0.329	1478
Life insurance	0.1157	0.32	1478
Health shock 2013	0.3876	0.487	3893
Health shock 2010	0.1636	0.37	3893
Death shock 2013	0.0241	0.154	3893
Death shock 2010	0.0049	0.07	3893
Male	0.8189	0.385	3893
Age	45.747	12.187	3893
Household size	4.5430	2.002	3893

As seen by these summary statistics, in the rural areas the ratio of households that had health insurance is approximately the same as the ratio that had life insurance. However, for the urban areas almost four times as many households had life insurance compared to health insurance. We can also see that approximately the same percentage of households in either area had experienced a health or death shock in both the twelve months before 2010 and in between 2010 and 2013.

Chapter 4

Results

This chapter is structured so that we first present the results from our naive OLS regressions divided by year, shock and area. Then, the chapter ends with the results and interpretations of the regressions using the DDD identification strategy.

4.1 Naive OLS regressions

Table 4.1 illustrates the results from the naive OLS regressions for 2013, and table 4.2 illustrates the same regressions but run on data from 2010. As displayed, twelve regressions were run in total for each year, and will hereby be referred to as models 1 to 12.

The regressions contain the effect of a shock on monthly household income and can be divided into three different groups: general, urban and rural. General includes all households in both areas, while urban contains specifically the urban households and rural specifically the rural households. For each group, the regression were run two times: one without control variables and one with. In the first six models the shock is categorized as a health shock, and in the final six as a death shock.

A result described as significant in this section is significant at the 10% level, the 5% level or the 1% level. For further information of what level of significance that occur in each case, see tables 4.1 and 4.2.

4.1.1 Results from 2013

As seen in table 4.1, when investigating the effects of a health shock on monthly household income in the different areas we get the following results. Models 1 to 4 are all significant. For the first model this means that if a household has had a health shock between 2010 and 2013, the income level is negatively affected by 11.0% compared to those that did not experience a health shock within these years. This general case thus applies to households in both urban and rural areas, and the significance holds when adding the controls in model 2, increasing the negative effect to 12.3%. The interpretation of model 3 is that if a household in the urban areas has had a health shock between 2010 and 2013, the income level is negatively affected by 17.4% compared to those in the urban areas that did not experience a health shock within these years. This significance and interpretation holds when adding controls in model 4, and changes the negative effect to 24.2%. As models 5 and 6 are not significant for the health shock variable we cannot reject the null hypothesis that in the rural areas there is no difference in income between those who did have a health shock and those who did not.

For the control variables in models 2, 4 and 6 we can see that being male has a positive income effect and is consistently significant in all three models. Age is significant both the general case (model 2) and the rural case (model 6), and has in both cases a negative effect on income with approximately 1%. Household size is consistently significant with a positive coefficient, whereas having a shock in the twelve months before 2010 is not significant in either one of the models.

For the death shock, on the other hand, all models except model 12 are significant. The interpretation of model 7 is that if a household (regardless of area) has had a death shock between 2010 and 2013, the household's monthly income is negatively affected by 91.7% compared to those that did not experience a death shock between these years. Adding control variables does not affect the significance level but slightly reduces the negative effect on monthly income to 83.5%. The interpretation of model 9 is that if a household in the urban areas has had a death shock between 2010 and 2013, the income level is negatively affected by 112.4% compared to those that did not experience a death shock in the urban areas between these years. Adding the control variables to this in model 10, the significance of death shock decreases and the negative effect on household income changes to 100.1%. For model 11, the results indicate that if a household in the rural areas has had a death shock between 2010 and 2013, the income level is negatively affected by 61.4% compared to those that did not experience a death shock within these years. This significance for the death shock is however lost when adding control variables in model 12.

For the control variables in models 8, 10 and 12, again being male is consistently significant with a positive effect on a household's monthly income. Furthermore, age is again significant with a negative coefficient in both the general case (model 8) and in the rural case (model 12) but not in the urban one (model 10). Household size has also consistent significance in all cases with a positive effect on the household's income. Lastly, whether or not the household had a death shock in the twelve months before 2010 is only significant in the rural model with controls (model 12) and has a positive income effect. The interpretation is thus that households who have had a previous death shock in 2010 in the rural areas have had a positive income effect with 115.7% compared to those who did not experience the death shock.

4.1.2 Results from 2010

When it comes to 2010, we have the following results in table 4.2. Models 1, 2, 5 and 6 are significant, while models 3 and 4 are not. Regarding the control variables, the interpretation is the same as in 2013, but now age is significant in all models. In model 1, regardless of area, if a household has experienced a health shock in the twelve months before 2010, the income level is negatively affected by 32.9% compared to those that did not experience a health shock in the same time period. When adding control variables to this in model 2 all controls are significant and the coefficient for the health shock has changed to 31.5%. Furthermore, models 5 and 6 imply that if a household in the rural areas has had a health shock before 2010, the income level is negatively affected by 49.1% and 46.9% respectively, compared to those that did not experience a health shock in rural areas in the same time period. However, we cannot reject the null hypothesis that in the urban areas there is no difference in monthly household income before and after a health shock if it occurred in the twelve months before 2010. The control variables in model 4 are all significant and have positive coefficients.

The effect of a death shock on monthly household income in 2010 is remarkably different from the rest of the results. All models except model 9 and 10 are significant for the death shock variable, however with positive coefficients. The interpretation for model 7 is that if a household (regardless of area) has had a death shock before 2010, the income level is positively affected by 85.1% compared to those that did not experience a death shock before 2010. All controls added to this in model 8 are significant and the effect of the death shock on the monthly income has increased to 94.2%. For rural areas in model 11 the results imply that if a household has had a death shock in the twelve months before 2010, the income level is positively affected by 166.4% compared to those in the rural areas that did not experience a death shock. Again, when adding the controls in model 12 all control variables are significant and have increased the coefficient for the death shock now to a positive income effect of 220.4%. However, for urban areas we cannot reject the null hypothesis that there is no difference in monthly household income before and after a death shock has occurred, if the shock occurred in the twelve months before 2010.

4.1.3 Summary of results from the naive OLS regressions

First of all we can see some similarities between the two regressions. The effect of a health shock for the general case is significant both with and without controls in both 2010 and 2013. In both cases the coefficient is negative and a health shock thus seems to overall have a negative effect on monthly household income. This negative impact seems to be greater in the 2010 round as this coefficient is larger and more significant. However, we also see some variations for the health shock between the regressions. For 2013 the urban models are significant, but in 2010 we have significance for the rural models. For 2013 the effect of a health shock is thus negative in urban areas, but nothing can be said about the effect in rural areas. In 2010, on the contrary, the effect of a health shock is negative in rural areas, but nothing can be said about the effect in urban areas.

For the death shock in 2013 we see a clear negative effect on monthly income, regardless of area. However, for 2010 the effect of a death shock is very much positive in both the general case and in the rural case, however nothing can be said about the effect in urban areas. The positive effect on household income after a death shock is not very plausible and seems to contradict the interpretation of the other results.

Overall it seems that these naive OLS models, regardless of area and shock in either of the years, have very little explanatory power of household monthly income, as seen by the low R^2. However, as previously mentioned in section 3.1, the results from these naive OLS regressions could be biased as there is a possibility that the assumption of an exogenous shock is not fulfilled.

Table 4.1: Naive OLS regressions for 2013 per area

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	General	General	Urban	Urban	Rural	Rural	General	General	Urban	Urban	Rural	Rural
	Income	Income	Income	Income	Income	Income	Income	Income	Income	Income	Income	Income
Health shock	-0.110**	-0.123**	-0.174***	-0.242***	-0.0109	0.0227						
	(-2.16)	(-1.98)	(-2.82)	(-3.11)	(-0.14)	(0.24)						
Male		0.362***		0.501***		0.924***		0.357***		0.492***		0.926***
		(5.76)		(7.72)		(7.26)		(5.71)		(7.65)		(7.27)
Age		-0.00926***		0.00106		-0.00997***		-0.00844***		0.00191		-0.00911***
		(-4.26)		(0.41)		(-3.01)		(-3.86)		(0.74)		(-2.73)
Household size		0.111***		0.112***		0.151***		0.109***		0.109***		0.150***
		(9.19)		(7.69)		(7.80)		(9.08)		(7.65)		(7.75)
Health shock 2010		0.0495		0.145		-0.0177						
		(0.61)		(1.42)		(-0.15)						
Death shock							-0.917***	-0.835***	-1.124***	-1.001**	-0.614*	-0.609
							(-3.50)	(-2.81)	(-2.78)	(-2.38)	(-1.84)	(-1.50)
Death shock 2010								0.325		-0.273		1.157**
								(0.54)		(-0.24)		(2.55)
Constant	13.10***	12.76***	13.69***	12.85***	12.42***	11.43***	13.07***	12.72***	13.65***	12.79***	12.43***	11.41***
	(436.32)	(105.62)	(418.92)	(97.83)	(254.41)	(51.78)	(542.06)	(105.29)	(494.04)	(97.25)	(325.53)	(51.79)
N	8266	8266	4373	4373	3893	3893	8266	8266	4373	4373	3893	3893
R^2	0.0006	0.0197	0.0020	0.0346	0.0000	0.0434	0.0037	0.0218	0.0071	0.0380	0.0016	0.0448

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 4.2: Naive OLS regressions for 2010 per area

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	General	General	Urban	Urban	Rural	Rural	General	General	Urban	Urban	Rural	Rural
	Income	Income	Income	Income	Income	Income	Income	Income	Income	Income	Income	Income
Health shock	-0.329***	-0.315***	-0.0411	-0.0463	-0.491**	-0.469**						
	(-2.98)	(-2.86)	(-1.03)	(-1.18)	(-2.52)	(-2.49)						
Male		0.169*		0.293***		1.480***		0.179**		0.296***		1.508***
		(1.94)		(10.09)		(7.31)		(2.05)		(10.19)		(7.41)
Age		-0.0203***		0.00624***		-0.0253***		-0.0206***		0.00619***		-0.0256***
		(-5.95)		(4.88)		(-4.31)		(-6.03)		(4.85)		(-4.35)
Household size		0.0717***		0.0274***		0.214***		0.0709***		0.0274***		0.213***
		(3.76)		(3.70)		(6.32)		(3.72)		(3.69)		(6.31)
Death shock							0.851***	0.942***	0.218	0.358	1.664***	2.204***
							(4.34)	(4.38)	(0.78)	(1.39)	(8.06)	(7.70)
Constant	12.37***	12.84***	13.62***	13.04***	11.07***	10.04***	12.32***	12.80***	13.61***	13.03***	10.99***	9.949***
	(312.21)	(74.49)	(904.41)	(206.42)	(152.01)	(28.66)	(330.17)	(74.02)	(974.98)	(205.81)	(161.37)	(28.28)
N	7904	7904	4011	4011	3893	3893	7904	7904	4011	4011	3893	3893
R^2	0.0013	0.0089	0.0003	0.0362	0.0018	0.0357	0.0003	0.0081	0.0002	0.0364	0.0008	0.0353

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

4.2 DDD regressions

Table 4.3 shows the results from the DDD. As illustrated, we only had to run four models. Model 1 uses health shock on monthly household income, as does model 2 but adds control variables. Model 3 is thus death shock on monthly household income, with model 4 adding the controls.

As we can see in table 4.3, the treatment effect (the triple interaction coefficient) for the health shock is significant on the 10% level and yields a negative effect on the households monthly income in both model 1 and 2. Without control variables, the negative treatment effect is 33.4%, and with controls it is 32.5%. Having a health shock in the twelve months before 2010 gives a 12.2% negative income effect compared to those that did not have a shock. The interpretation of the above is that urban areas have a larger difference in monthly household income between those who did and did not experience a health shock, before and after the shock occurred, compared to the rural areas. That is, urban areas recover less than rural areas from a health shock occurring between 2010 and 2013, meaning that households with a health shock in the urban areas are worse off financially than those in the rural areas.

As mentioned in section 3.1 the single variables contain observable and unobservable characteristics, controlled for as fixed effects. The results in table 4.3 therefore tell us that people in Colombia earn more money in 2013 compared to 2010, as can be seen by the positive and significant coefficients for the time variable. Furthermore, urban households earn consistently about 257-272% more than rural households, which is also significant on the 1% level. The single variable for a health shock between 2010 and 2013 is negative and significant at the 10% level without controls, which is interpreted as poor households being more likely to have a health shock than rich households.

In table 4.3, we also observe that the treatment effect of a death shock is significant at the 10% level. In fact, the p-value for the treatment effect is 0.05, which means that it is very close to being significant at the 5% level. The negative effect is 163.3% without control variables and 160.6% with controls. All variables in models 3 and 4 are significant on at least the 10% level, except for the interaction of time and death shock, and the control variable for death shock in 2010. The interpretation of the treatment effect of a death shock is the same as the one for a health shock. We see that urban areas do not recover as well as rural areas from a death shock, and that those who had a death shock in the urban areas in between 2010 and 2013, are worse off financially than those who had the shock in the rural areas.

As with the health shock, the single variables for the death shock control for fixed effects. We can therefore see that Colombian households have a higher income in 2013 compared to 2010. We also see that urban households earn 259-274% more than rural households. Furthermore, the single variable containing death shock in between 2010 and 2013 is significant at the 1% level and negative, which means that poor households are more likely to have a death shock than rich households.

In table 4.3 we see a quite high degree of explanatory power. Adding the controls improves these models as the R^2 increases to 16.7% and 16.9% for each shock respectively.

Table 4.3: DDD table

	(1)	(2)	(3)	(4)
	Income	Income	Income	Income
Year 2013	1.347***	1.358***	1.405***	1.410***
	(15.84)	(15.98)	(19.08)	(19.18)
Urban	2.569***	2.722***	2.591***	2.737***
	(32.21)	(33.63)	(37.36)	(38.63)
Health shock 2013	-0.278*	-0.237		
	(-1.78)	(-1.54)		
Year 2013*Urban	-1.316***	-1.319***	-1.382***	-1.382***
	(-14.51)	(-14.58)	(-17.60)	(-17.66)
Year 2013*Health shock 2013	0.262	0.252		
	(1.56)	(1.50)		
Urban*Health shock 2013	0.180	0.135		
	(1.13)	(0.86)		
Year 2013*Urban*Health shock 2013	-0.334*	-0.325*		
	(-1.85)	(-1.81)		
Health shock 2010		-0.122*		
		(-1.93)		
Household size		0.129***		0.128***
		(11.63)		(11.50)
Age		-0.00694***		-0.00616***
		(-3.57)		(-3.17)
Male		0.696***		0.700***
		(12.90)		(12.96)
Death shock 2013			-1.712***	-1.616***
			(-2.77)	(-2.66)
Year 2013*Death shock 2013			0.884	0.953
			(1.25)	(1.35)
Urban*Death shock 2013			1.433**	1.365**
			(2.23)	(2.16)
Year 2013*Urban*Death shock 2013			-1.633*	-1.606*
			(-1.96)	(-1.92)
Death shock 2010				0.487
				(1.60)
Constant	11.07***	10.24***	11.03***	10.15***
	(141.88)	(76.36)	(162.30)	(78.91)
N	16170	16170	16170	16170
R^2	0.1460	0.1674	0.1480	0.1688

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Chapter 5

Discussion

This chapter analyzes the results in chapter 4. We return to the research question and hypothesis, evaluate our work and assumptions and discuss future work.

5.1 Evaluation of results

Our research question in this thesis was *does the income of urban and rural households recover differently from an income shock caused by an exogenous health or death shock?*. Our results show a statistically significant difference for the treatment effect between time, residential area, and those affected and unaffected by a shock. This is consistent for both the health and death shock. The interpretation of our results is that households in urban areas that have experienced a shock are worse off financially than those experiencing the shock in rural areas. Thus, our research question is affirmatively justified, and there seems to be a statistically significant difference in income recovery between the two areas in Colombia after a health or death shock.

Another way in which our results give evidence to suggest that area is important for the relationship for income recovery between shock and time is by comparing the results from the DD regression in Appendix B with the DDD regression in section 4.2. The DD tests only shock and time with non-significant results, whereas the DDD adds area to this relationship and yields statistically significant results. It thus seems that whether the household is urban or rural is very important when testing the effects of health or death shock on monthly household income.

Our results from the DDD regression in section 4.2 further show that poor households are more likely to have a larger negative income effect from a health or death shock. Combining this result with the criticism of Giedion and Uribe (2009) regarding the Colombian health care system, mentioned in section 2.1, we see that poor households are more affected by the shock, but are also discriminated by the health system. This combination thus shows the disadvantage of being poor in Colombia, that you are more affected by shocks but are less treated by the system.

Furthermore, our naive OLS regressions overall showed that the effects from either a health or death shock had a negative impact on monthly household income. This is hence in line with the results from the DDD. However, for the death shock in the 2010 regressions we got positive and statistically significant coefficients. In other words, the results indicate that experiencing a shock where the household head or the head's spouse died within twelve months of the survey in 2010, would have a positive effect on monthly household income. These results seem unreasonable, not only compared to the results from 2013, but also from an empirical perspective. Again, this is likely due to the naive nature of the OLS which is why we progressed to the DDD. It is likely that the estimators might be biased as the OLS assumption of an exogenous shock might not be fulfilled. This is confirmed by previous studies such as Mohanan (2013) which have showed that there is a consistent endogeneity problem between health and wealth, which is mentioned in section 2.4. The fact that our results show that poor households are more affected by the shock indicates that this endogeneity problem is also apparent in this study. However, we tried to investigate if households that had experienced a death shock had received a lump sum from some insurance company, which could explain a temporary positive effect on monthly income, but found no information supporting this belief in the ELCA datasets. We thus conclude that the results from the naive OLS might not be reliable for further analysis.

An interesting feature of the ELCA data is that relatively few households in both urban and rural areas say that they have a health or life insurance. Our results further show that in the urban areas more households had life insurance than health insurance. This is very much in contrast to what OECD (2015) states, mentioned in section 2.1, namely that 97% of the Colombian population had a health insurance in 2015. Even though our insurance data is from 2013 it is unlikely that such a big increase in insurance coverage can happen up to 2015. This difference can thus be due to a faulty variable, meaning that households did not know that they were covered by a health insurance. It might also be that the households thought the survey question meant some private, additional health insurance and chose not to include this official health insurance in their reply. A third option is that the official numbers are not representative of the actual situation in Colombia.

5.2 Evaluation of hypothesis

Our hypothesis as to why there might be a difference in income recovery after a shock between the urban and rural areas was in short that there would be a difference in the level of and access to insurances between the areas. We considered it likely that formal insurances were stronger in urban areas and informal insurances were stronger in rural areas. Our idea was therefore that those affected by a shock and formally uninsured in urban areas would be financially worse off than those affected by a shock and formally uninsured in the rural areas. This difference in gap would thus be due to a greater informal insurance sector in the rural areas than the urban areas.

Our study was unable to test the prevalence of informal insurance in the two areas due to the nature of our data. We also had to first establish that there in fact was a difference in income recovery between the areas before we could test insurance as a mechanism. This is therefore not explicitly considered in this study, however, our results show that the shock-affected households in urban areas are financially worse off than those in rural areas. When comparing this to the summary statistics of those who were and were not formally insured between the areas we can see that urban areas tend to have a higher ratio of households with life insurance compared to the rural areas. However, although this ratio is quite high, the life insurance in the urban areas does not seem to compensate the loss of the informal insurance system as those affected by a shock are more negatively affected in the urban areas than the rural areas.

There might be other factors that caused urban households to have a larger gap in income recovery between those formally insured and uninsured compared to rural households. As mentioned in section 2.4, Wagstaff (2007) states that rural areas are better at adjusting the labor supply after a death shock. This is in line with our results, which thus show that urban households are worse at adjusting their labor after a shock compared to rural households. It is likely that urban areas have a higher degree of wage jobs with higher specialization skills compared to rural areas, where one farm worker is more easily substituted for another. As our results are in agreement with previous studies we deem them to be reliable. Our contribution to this is therefore that we have confirmed earlier theories and added the importance of insurance.

The overall contribution of our thesis is thus to confirm that there is a difference in income recovery between urban and rural areas before and after a health or death shock occurred. We deduct that this difference is likely to be due to differences in informal insurances. With regards to our sample size it is possible that this difference can be seen in all urban and rural areas in Colombia. However, one must note that it is likely that the variation in distance between urban and rural areas might affect the structure of the formal and informal insurance systems, see the work of Townsend (1995) in section 2.4.

5.3 Future work

As discussed in subsections 3.2.1 and 3.2.2 our income variables were a bit problematic. Future work might thus want to recreate our study but with household expenditures as the dependent variable as this might be more reliable and consistently measured between the rounds of the survey.

The ELCA datasets had plenty of interesting variables that we due to the time constraint could not investigate further. However, future studies could look at different outcome variables such as health status to see for instance whether a health shock affects health status differently between the areas. Another interesting variable to include would be households debts, as previous studies showed that after a health shock consumption can be smoothed using increased debts, see Mohanan (2013) in section 2.4. Whether or not the level of indebtedness due to a shock differs between the areas could therefore also yield interesting insights.

Finally, our study is heavily based on the assumptions made in section 1.2, namely that the health and death shocks are exogenous. Furthermore, in cases where this first assumption was not fulfilled, the ratio of "expected shocks" would be equal in both areas. If these assumptions do not hold, our results from the OLS regressions may not be valid. However, by using a regression with a DDD identification strategy we have captured some of the effects that may be caused by a not fully exogenous shock. If it is deemed that the assumptions indeed do not hold, future work can progress to an IV regression which uses another variable as an instrument for the shock.

Chapter 6

Conclusions

The main conclusion we draw in this thesis is that the income of urban and rural households recover differently from an income shock caused by a health or death shock.

We further deduct that this might be due to differences in formal and informal insurances, however further studies on the subject are required. Our results show that even though life insurance in urban areas are quite common, they seem to work poorly. This as the households affected by a shock in the urban areas are worse off financially compared to those affected by the shock in the rural areas.

Our results also confirm previous problems between health and wealth, that poorer households in fact are more negatively affected by the shock than wealthier households. Previous studies also indicate that these poorer households are disadvantaged in terms of the health care system in Colombia. This would mean that poor households in Colombia are more likely to be affected by a shock but also less likely to receive any medical health care if needed.

Bibliography

Alvarez, L. S., Salmon, J. W. and Swartzman, D. (2011), 'The colombian health insurance system and its effect on access to health care', *International Journal of Health Services* **41**(2), 355–370.

Baeza, C. and Packard, T. (2006), *Beyond survival: protecting households from health shocks in Latin America*, Stanford University Press.

Banerjee, A. V. and Duflo, E. (2005), 'Growth theory through the lens of development economics', *Handbook of economic growth* **1**, 473–552.

Bardhan, P. and Udry, C. (1999), *Development microeconomics*, OUP Oxford.

Beresford, N. and Rubio, R. (2015), *The Insurance and Reinsurance Law Review - Colombia*, Law Business Research.

Besley, T. (1995), 'Savings, credit and insurance', *Handbook of development economics* **3**, 2123–2207.

Centro de Estudios Sobre Desarrollo Economico - CEDE (2010), *Encuesta Longitudinal Sobre Diamica de los Hogares Colombianos, PH-D1-R1: Diseño de la Muestra*, [pdf] Bogotá: Universidad de los Andes Facultad de Economia. Available at: *https://encuestalongitudinal.uniandes.edu.co/es/component/phocadownload/category/12-otros-documentos?download=77:diseno-muestral-pdf* [Accessed 15 April 2017].

Centro de Estudios Sobre Desarrollo Economico - CEDE (2016), *Documento de Cambios Bases de 2010*, [pdf] Bogotá: Universidad de los Andes Facultad de Economia. Available at: *https://encuestalongitudinal.uniandes.edu.co/es/component/phocadownload/category/12-otros-documentos?download=74:cambios-bases-2010-pdf* [Accessed 15 April 2017].

Chapman, A. R. (2016), 'The contribution of human rights to universal health care', *Health & Human Rights Journal* **18**(2).

Dercon, S. (2000), 'Income risk, coping strategies and safety nets'. Background paper for the World Development Report 2000/01, WPS/2000.26. Oxford University.

Fernández, M., Ibáñez, A. M. and Peña, X. (2014), 'Adjusting the labour supply to mitigate violent shocks: Evidence from rural colombia', *The Journal of Development Studies* **50**(8), 1135–1155.

Giedion, U. and Uribe, M. V. (2009), 'Colombia's universal health insurance system', *Health Affairs* **28**(3), 853–863.

Iregui-Bohórquez, A. M., Melo-Becerra, L. A. and Ramírez-Giraldo, M. T. (2016), 'Health status and labor force participation: evidence for urban low and middle income individuals in Colombia', *Portuguese Economic Journal* **15**(1), 33–55.

Mohanan, M. (2013), 'Causal effects of health shocks on consumption and debt: Quasi-experimental evidence from bus accident injuries', *Review of Economics and Statistics* **95**(2), 673–681.

Morduch, J. (1999), 'Between the state and the market: Can informal insurance patch the safety net?', *The World Bank Research Observer* **14**(2), 187–207.

Morrison, A. R., Raju, D. and Sinha, N. (2007), *Gender equality, poverty and economic growth*, Vol. 4349, World Bank Publications.

OECD (2015), *OECD Reviews of Health Systems: Colombia 2016*, OECD Publishing, Paris.

Pauly, M. V., Zweifel, P., Scheffler, R. M., Preker, A. S. and Bassett, M. (2006), 'Private health insurance in developing countries', *Health Affairs* **25**(2), 369–379.

The World Bank (2015), 'Colombia'. [online] Available at: *http://data.worldbank.org/country/Colombia* [Accessed 6 April 2017].

Townsend, R. M. (1995), 'Financial systems in northern thai villages', *The Quarterly Journal of Economics* **110**(4), 1011–1046.

Trujillo, A. J., Portillo, J. E. and Vernon, J. A. (2005), 'The impact of subsidized health insurance for the poor: evaluating the colombian experience using propensity score matching', *International journal of health care finance and economics* **5**(3), 211–239.

Universidad de los Andes (n.d.), 'ELCA project'. [online] Available at: *https://encuestalongitudinal.uniandes.edu.co/en/elca/elca-project* [Accessed 19 April 2017]".

Wagstaff, A. (2007), 'The economic consequences of health shocks: evidence from vietnam', *Journal of health economics* **26**(1), 82–100.

Webster, P. C. (2012), 'Health in colombia: a system in crisis', *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* **184**(6), E289–90.

Appendices

Appendix A

Variables

Since the datasets are originally in Spanish, this appendix contains both the original Spanish names of the variables and datasets, together with the English names they are given in this thesis and a short explanation of what they contain and represent.

Table A.1: Datasets used in this thesis

Original Spanish name	**Given English name**	**Content**
Rhogar_2010	Rural Households 2010	All household variables except shock variables and household head control variables.
Rchoques_hogar_2010	Rural Shocks 2010	Shock variables for households (health shock and death shock).
Rpersonas_2010	Rural People 2010	Control variables for household head (sex and age).
Uhogar_2010	Urban Households 2010	All household variables except shock variables and household head control variables.
Uchoques_hogar_2010	Urban Shocks 2010	Shock variables for households (health shock and death shock).
Upersonas_2010	Urban People 2010	Control variables for household head (sex and age).
Rhogar_2013	Rural Households 2013	All household variables except shock variables and household head control variables.
Rchoques_2013	Rural Shocks 2013	Shock variables for households (health shock and death shock).
Uhogar_2013	Urban Households 2013	All household variables except shock variables and household head control variables.
Uchoques_2013	Urban Shocks 2013	Shock variables for households (health shock and death shock).

Table A.2: Key variables used in this thesis

Original Spanish name	**Given English name**	**Description of variable**
consecutivo	Household id	Identification number of household.
llave	Key	Another identification number used mainly for identification of split households in datasets from 2013.
zona	Urban	Dummy variable equal to 1 if the household lives in the urban areas and 0 in the rural areas.
t_personas	Household size	Number of people in the household.
sexo	Male	Dummy variable equal to 1 if the household head in 2010 was male.
edad	Age	Age of household head in 2010.
act_segsal	Health insurance	If a household had health insurance or not in 2013.
act_segvida	Life insurance	If a household had life insurance or not in 2013.
ing_trabajo		Average monthly income from labor per household. Non-existent in the *Rural Households 2013* dataset.
ing_trabnoagr		Income from non-agricultural labor. Only existent in the *Rural Households 2013* dataset.
ing_trabagr		Income from agricultural labor. Only existent in the *Rural Households 2013* dataset.
ing_pensiones		Income from pensions.
ing_arriendos		Income from leases.
ing_intereses_div		Income from interests or dividends.
ing_otros_nrem		Other income.
	Monthly household income	New income variable, non-existent in original Spanish datasets. Combines ing_trabajo, ing_trabnoagr, ing_trabagr, ing_pensiones, ing_arriendos, ing_intereses_div and ing_otros_nrem. Main variable of interest, used as Y in all regressions.
	Year 2013	New dummy variable equal to 1 if the data was collected in 2013, and 0 if the data was collected in 2010.

tuvo_choque	Combined with choque to create health shock 2013, health shock 2010, death shock 2013 and death shock 2010	Whether the household had experienced the shock or not.
choque	Combined with tuvo_choque to create health shock 2013, health shock 2010, death shock 2013 and death shock 2010	The different types of shocks the household might have had.
	Health shock 2013	New variable, non-existent in original Spanish datasets. Indicates if a household had experienced a health shock or not.
	Death shock 2013	New variable, non-existent in original Spanish datasets. Indicates if a household head or the head's spouse died during the time period.
	Health shock 2010	Indicates if a household has had a health shock before 2010 or not.
	Death shock 2010	Indicates if a household has had a death shock before 2010 or not.

Appendix B

Results from DD regressions

The regressions using a difference-in-differences (DD) identification strategy were made to illustrate the difference in income between those who were affected and unaffected by the shock, before and after the shock occurred. This is thus expressed in equations B.1 and B.2 which compared to the other regressions in our study do not take area into account. Again **X'** represents a vector of controls.

$$\hat{\delta} = (\bar{Y}_{t_1} - \bar{Y}_{t_0})^{Shock} - (\bar{Y}_{t_1} - \bar{Y}_{t_0})^{No\ shock} \tag{B.1}$$

$$log(Y) = \alpha_0 + \alpha_1 Time + \alpha_2 Shock + \delta(Time * Shock) + \mathbf{X'}\beta + U \tag{B.2}$$

The regressions were run in four models, the first two for a health shock with and without controls, and the last two for a death shock with and without controls. In table B.1 we see that the treatment effect between year and health shock is not statistically significant, regardless of the inclusion of control variables. This means that the DD does not tell us much as we cannot reject the null hypothesis that there is no difference between affected and unaffected households, before and after the shock occurred. This conclusion can also be drawn for the treatment effect with death shock as this is not significant at any of the three significance levels either with or without controls. Since these results therefore do not contribute to the overall results in this thesis, they are placed here in the appendix. However, they are briefly mentioned in 5.1.

Table B.1: DD table

	(1) Income	(2) Income	(3) Income	(4) Income
Year 2013	0.706***	0.712***	0.728***	0.733***
	(15.50)	(15.69)	(18.48)	(18.64)
Health shock 2013	-0.184**	-0.158*		
	(-2.16)	(-1.86)		
Year 2013*Health shock 2013	0.0907	0.0871		
	(1.00)	(0.96)		
Health shock 2010		-0.163**		
		(-2.28)		
Household size		0.0923***		0.0911***
		(7.99)		(7.88)
Age		-0.0146***		-0.0139***
		(-6.68)		(-6.33)
Male		0.268***		0.270***
		(4.54)		(4.57)
Death shock 2013			-1.145***	-1.014***
			(-3.03)	(-2.71)
Year 2013*Death shock 2013			0.167	0.228
			(0.38)	(0.51)
Death shock 2010				0.190
				(0.67)
Constant	12.37***	12.45***	12.34***	12.36***
	(288.62)	(105.89)	(332.01)	(106.55)
N	16170	16170	16170	16170
R^2	0.0173	0.0278	0.0193	0.0288

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01