ESSAYS IN DECISION MAKING UNDER UNCERTAINTY AND THE POLITICAL ECONOMY OF INSTITUTIONS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Esteban José Méndez Chacón

August 2019

ESSAYS IN DECISION MAKING UNDER UNCERTAINTY AND THE POLITICAL ECONOMY OF INSTITUTIONS

Esteban José Méndez Chacón, Ph.D.

Cornell University 2019

This dissertation consists of two lines of research. The first line focuses on decision making under uncertainty. Specifically, I analyze the nature of risk preferences in life-insurance choices, and I examine adverse selection and moral hazard in private-healthcare markets in a context where the government provides universal healthcare. The second line focuses on the political economy of institutions, and I study the impact of a multinational company on the economic development of its host country during and after its tenure.

In Chapter 1, I estimate a structural model of risk preferences from the choice of insured amount in life-insurance contracts. The literature investigating life-insurance purchases has mostly focused on survey data to estimate how demand for insurance depends on demographic and socioeconomic factors, leaving aside the task of estimating an underlying model of decision making that generates this demand. Using proprietary data on life-insurance choices, I estimate a model of risky choice to explain households' decisions. My results indicate that, in addition to standard risk aversion (decreasing marginal utility for wealth), including decision weights that might differ from the actual probabilities improves the fit of the model. These weights can be interpreted as a combination of state-dependent utility and probability distortions. Moreover, I find support for the existence of heterogeneity in preferences. Women are more risk averse than men, and the decision weights vary depending on gender and age.

In Chapter 2, we investigate adverse selection and moral hazard in private healthcare in markets where the government provides universal healthcare. We use proprietary data on health-insurance choices and medical expenditures to examine asymmetric information in this institutional setting. We disentangle adverse selection and moral hazard by leveraging variation in copays and deductibles implemented by the insurance company, and the fact that, for a sample of customers, the insured amount was exogenously assigned, shutting down the adverse selection channel. We find evidence for the presence of asymmetric information, both in the form of adverse selection and moral hazard. Moreover, we develop a model that incorporates sources of heterogeneity that could potentially explain selection in this institutional setting, namely risk aversion, risk type, concern for health, and taste for convenience services offered in the private healthcare sector.

In Chapter 3, we analyze the impact of large-scale FDI on economic development by studying the case of the United Fruit Company (UFCo) in Costa Rica from 1889 to 1984. We implement a geographic regression discontinuity design that exploits a quasi-random assignment of land, and census data geo-referenced at the census-block level for 1973, 1984, 2000 and 2011. These allow us to identify the company's effect during its tenure, and after it stopped production. We find that former "UFCo lands" have had higher living standards and better economic outcomes than counterfactual areas without the company's presence, and that convergence is slow. These findings are validated using nighttime lights data. Detailed historical data suggest that the mechanisms behind our results are investments in physical and human capital carried out by the UFCo.

BIOGRAPHICAL SKETCH

Esteban José Méndez Chacón was born in San José, Costa Rica in July 1989. He earned his B.Sc. in Economics from Universidad de Costa Rica in 2012. Following his undergraduate studies, he was awarded a doctoral scholarship from Banco Central de Costa Rica. In 2014, he started his graduate work at Cornell University, that he completed under the supervision of Professor Ted O'Donoghue. Esteban is an applied microeconomist whose research explores decision making under uncertainty and the political economy of institutions. He earned his M.A in Economics in 2017, and after completing his Ph.D., he will work as a Research Economist at the Banco Central de Costa Rica.

To everybody who has supported me in any part of the way.

ACKNOWLEDGEMENTS

I could not have completed my Ph.D. without many people that along the way have selflessly provided their advice, encouragement, help, and trust. Having all these people has amplified my efforts and has increased my chances to succeed. To all of them, thanks. At the risk of forgetting someone, and I apologize if I do, I proceed to enumerate some of them.

I am very grateful for my family and their constant love and prayers. Despite the distance, I never felt alone. I thank my parents, Deyanira and Elías, who have been supportive in every step of my education and my life. It is their sacrifices and their lessons on hard work and perseverance that got me until here. I also thank my brother, Josué, for his support and for being with my parents when I could not be around. I thank Krissya, for her companion and inspiration, and for representing all the amazing things that will come in the next chapter of my life.

I am grateful for a cast of educators who have prepared me and taught me the abilities needed to accomplish this goal. In particular, I thank Carlos Alpízar, Ronulfo Jiménez, Danilo Montero, Erlend Muñoz, Juan Robalino, and Adolfo Rodríguez; they all fostered my interest in economics during my undergraduate, and they encouraged me to pursue a Ph.D. Moreover, I keep many great friends from my undergraduate studies that always have supported me: Diego Astorga, Gabriela Espinoza, Tabaré Capitán, Jonathan Garita, Juliana Mesén, Angelo Orane, Allechar Serrano, Melissa Vega, Alberto Vindas, Rodrigo Villalta, and Alvaro Zuñiga.

This dissertation was written under the supervision of Ted O'Donoghue, Levon Barseghyan, and Marco Battaglini, and I thank them for their guidance and support. I am especially grateful to Ted for his encouragement through graduate school, for reading countless permutations of the manuscript, for being always

available for meetings, and for providing his feedback on any research idea; all his dedication has made me a better researcher.

I thank Abhishek Ananth, Jee Hun Choi, Tilahun Emiru, Yu She, Armand Sim, Luoyi Su, Lucy Xiaolu Wang, Qi Wu, Nahim Bin Zahur, and Xinrong Zhu. More than colleagues and friends, they all became my family in Ithaca. Thank you for all Friday dinners, all FIFA matches, and your continuous support which got me through Cornell.

I am grateful to the insurance companies who generously provided the data for the first two chapters. I thank Sarah Schlagter, Stephanie Jacobs, and Elena Goloborodko for assistance in obtaining and working with the restricted data. For helpful comments which greatly improved the first two chapters, I thank Francesca Molinari, Nathan Lindberg, and conference participants at Banco Central de Costa Rica (BCCR), Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Cornell University, the Behavioral Economics Reading Group (BERG), and the Works in Progress Seminar (TWIPS). For the second chapter of my dissertation, I am greatly indebted with my coauthors Jason Somerville and Matthew Thirkettle. They have been excellent coauthors, and I have learned a lot from our interactions.

The third chapter of my dissertation is a collaboration with Diana Van Patten, a Ph.D. Candidate at UCLA. She has been a wonderful friend since I met her in undergraduate, and she has been pivotal in many of my accomplishments. I also thank Keith Jenkins, GIS Librarian at Mann Library, for helping me with all the questions I had while using QGIS. I thank Jorge León and Adolfo Quesada-Román for valuable discussions on historical background and assistance in obtaining maps. For the third chapter I am grateful for feedback from Douglas L. Miller and seminar and conference participants at Cornell University, 2018 North East

Universities Development Consortium (NEUDC), 2018 Empirics and Methods in Economics Conference (EMCON), EEA-ESEM 2018, the Behavioral Economics Reading Group (BERG), and the Works in Progress Seminar (TWIPS).

The support provided by the Banco Central de Costa Rica (BCCR) was fundamental for completing my graduate studies and this dissertation. I thank Janilú Azofeifa, Felipe Castañeda, Róger Madrigal, William Mora, Evelyn Muñoz, and Ricardo Rodríguez. Besides financial support, they were very helpful in providing access to data. I hope one day I can give back to my country everything it has invested in my education. The findings and conclusions expressed are solely those of the author(s) and do not represent the views or policy of the Banco Central de Costa Rica (BCCR) or any other institution.

"Quo magis res singulares intelligimus, eo magis Deum intelligimus."

Benedictus de Spinoza. Eth. V. Prop. XXIV.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

RISK PREFERENCES IN LIFE-INSURANCE DECISIONS

1.1 Introduction

Most choices in life are made without fully knowing their outcomes, and, consequently, the analysis of risk preferences has played an important role in numerous studies in economics. Some of these studies have restricted attention to life-insurance purchases. In this context, the usual approach is to evaluate how demand for life insurance depends on demographic and socioeconomic variables. An area that remains relatively unexplored is to estimate an underlying preference model that generates this demand.

In this study, I estimate a structural model of risk preferences from the choice of insured amount in life-insurance contracts. In addition to standard risk aversion (diminishing marginal utility for wealth), the model allows for weights on all states that might differ from the actual probabilities. These weights can be interpreted as a combination of probability weighting and state-dependent utility. For the estimation, I use data from an insurance company that operates in Costa Rica. The core sample includes 3,164 households that bought insurance against death, incapacity, and serious disease (i.e., cancer, stroke, renal insufficiency or myocardial infarction) between 2010 and 2015. Households chose an insured amount, where they receive that amount in case of death or incapacity and half of that amount in case of a serious disease.

My results indicate that, in addition to standard risk aversion (as in the expected-utility model), including decision weights that might differ from the ac-

tual probabilities improves the fit of the model. The decision weights suggest that the adverse event changes the marginal utility of consumption, or that people distort the actual probability of the adverse event. In particular, the marginal utility of consumption increases as a consequence of death or incapacity, and decrease following a serious disease. Alternatively, the results can be interpreted as individuals overweighting the probability of death and incapacity relative to the probability of serious disease. Moreover, I find support for the existence of heterogeneity in preferences. Women are more risk averse than men, and the decision weights vary across gender and age categories. The main message is robust to sensitivity checks, such as alternative assumptions on the utility function or the relevant wealth.

These results contribute to a strand of literature addressing life-insurance choices, and more broadly to research on the estimation of risk preferences using field data. To the best of my knowledge, this study is the first that estimates the nature of risk preferences using data on life-insurance choices. In the life-insurance literature, most empirical studies rely on household income or consumer-expenditure surveys. In my case, I have actual individual-level market data that reflects agents' life-insurance decisions given the available choice set. Moreover, most studies in the life-insurance context are concerned with how macroeconomic, demographic, and socioeconomic variables matter for the purchasing behavior. For example, with few exceptions, the studies have found that variables such as income, employment rates, education level, family size, and net worth are positively related to life-insurance demand. Overall, the literature, as well as the number of variables considered, are extensive. Zietz (2003), Liebenberg et al. (2012), Outreville (2013), and Outreville (2014) review these papers and their main results. Different from this line of research, I explicitly estimate the structure of preferences that explain the households' choices. In this sense, my study is close to Halek and Eisenhauer

(2001). They estimate the relative-risk-aversion parameter for 2,376 households using survey data on life-insurance purchases and analyze how this measurement changes across demographic groups. However, unlike their analysis, I allow for non-expected-utility models and the possibility of state-dependent preferences.

The estimation of the structure of risk preferences using field data has been a topic receiving increasing attention in economics. Barseghyan et al. (2018) review and assess the relevant literature, which includes field contexts such as insurance (e.g., Cicchetti and Dubin 1994; Cohen and Einav 2007a; Barseghyan et al. 2013), labor markets (e.g., Chetty 2006), and horse race betting (e.g., Jullien and Salanie 2000; Snowberg and Wolfers 2010). Nevertheless, an important reason to extend the analysis to the life-insurance context is that, in life insurance, high-stakes lotteries are involved, unlike other types of insurance already examined in previous studies, such as inside-wire (Cicchetti and Dubin, 1994) or property insurance (Cohen and Einav 2007a; Barseghyan et al. 2013).

While the literature that estimates risk preferences using field data has considered both expected-utility theory and non-expected-utility models, most assume that the shape of the utility function does not change across states. Although this state-independence assumption can be justified for some environments, it might be non-credible in cases where the enjoyment of pecuniary payments depend on the final state once the uncertainty was resolved. This feature arises in life, health, or incapacity insurance, and more generally, in lotteries involving the loss of non-replaceable objects, such as species extinction (Kelsey, 1992).

Despite the variety of possible applications, there has been little empirical work on state-dependent preferences (Finkelstein et al., 2009). The existing work mainly focuses on the case of health status and is based on survey responses (e.g., Viscusi

and Evans 1990; Evans and Viscusi 1991; Sloan et al. 1998) and panel data (e.g., Lillard and Weiss 1997; Edwards 2008; Finkelstein et al. 2013). Moreover, except for Sloan et al. (1998), which allows for overweighting of small probabilities, state-dependent preferences are modeled within an expected-utility framework omitting non-expected-utility alternatives. The evidence of state-dependent preferences is inconclusive; studies have found negative state dependence (e.g., Viscusi and Evans 1990; Sloan et al. 1998; Finkelstein et al. 2013), positive state dependence (e.g., Lillard and Weiss 1997; Edwards 2008), and no evidence of state dependence (e.g., Evans and Viscusi 1991). By using insurance choices and considering the potential role of non-expected-utility alternatives, my analysis provides a different exploration to the study of state-dependent preferences. Moreover, by finding evidence of changes in the marginal utility of consumption due to the adverse event, my results contribute to a matter that continues unsettled in the literature.

To extend the analysis of risk preferences to the life-insurance context entails some limitations. First, claim rates show low variation across households. Without variability in the probability of an event, it is not possible to distinguish state-dependent utility from probability distortions. Hence, the estimated weight associated with each state has an open interpretation. In the analysis, I discuss both interpretations. A second limitation is that the data do not contain an explicit loss variable, for example, the monetary loss due to a permanent incapacity. One way to solve this issue that I implement in the analysis is to estimate the loss variable directly from the data.

The rest of the document is divided as follows. Section 1.2 describes the insurance product and the data. Section 1.3 develops an expected-utility model for life insurance and introduces further modifications to incorporate state-dependent

preferences and probability distortions. Moreover, it describes the model's identification and estimation. Section 1.4 presents and discusses the results, first under the assumption of homogeneous preferences, and later extends the model to allow for observable heterogeneity. Section 1.5 summarizes different robustness checks that confirm the results. Finally, Section 1.6 concludes the chapter.

1.2 Data

The data come from an insurance company that operates in Costa Rica. The company sells products that provide coverage in case of death, total and permanent disability, some types of diseases, medical expenses due to accidents, and funeral expenses. In this chapter, I restrict attention to a product that offers payment in three possible events: death, incapacity, and serious disease. The insured amount is chosen by the customer and can be any quantity between 1 million and 60 million Costa Rican Colones (CRC).[1] In the case of death or incapacity, the beneficiary would receive the insured amount. Incapacity is the loss or declining of physical or intellectual abilities as a result of an accident or illness that interferes with the person's earning capacity. The definition includes cases like total permanent paralysis, the anatomical or functional loss of both hands or both feet, total loss of vision and incurable insanity. To receive the compensation for this coverage the loss of earning capacity should be higher than 66.67%. Finally, a serious disease refers to cancer, stroke, renal insufficiency, or myocardial infarction. In this case, the payment is half of the insured amount.

[1]For comparison, this means an amount between US$1,871 to US$112,233. Costa Rican GDP per capita based on purchasing power parity (PPP) was on average 0.27 of that in the United States during the considered period.

The dataset is a cross-section of all customers who had a policy active in December 2015. The full dataset includes a total of 5,236 households, who between 2010 and 2015 bought 5,419 insurance policies. For each policy, the data record the issue date, the insured amount, and the premium paid at the time when it was first purchased. Moreover, the data contain for each policyholder their date of birth, gender, and place of residence. The data also include the claims associated with each policy between 2010 and 2015 and specify the event that happened to the policyholder.

The coverage is annual, and it renews automatically every year. I use choices at the time when the insurance was first issued, thus restricting attention to active choices to reduce the component of inertia in the decision.[2] The protection in case of death and serious disease can be extended to the immediate family of the policyholder (partner and children until 21 years old). I focus on households that bought insurance for just one person. If households that provide insurance for more than one member are considered, the number of possible states would increase depending on each insured member's possible states of the world, and possibly their choices do not reflect own risk preferences.

With these restrictions, the total number of households in the sample is 3,164, who first purchased the insurance between August 2010 to December 2015. Table 1.1 provides descriptive statistics for the sample as at year of first purchase.

As explained above, the insurance allows one to choose any insured amount between CRC 1 million and CRC 60 million; however, people tend to pick certain focal values. This fact will be useful to discretize the choice set. I include a value

[2]In other contexts, inertia has proven to have an important impact on people's choices. Examples are saving behavior (Madrian and Shea, 2001) or (more in line with this chapter) health insurance (Handel, 2013a).

Table 1.1: Descriptive Statistics

Variable	Mean	SD	Min	Max
Age (years)	38.38	9.73	19	76
Female (=1)	0.77			
Costa Rican (=1)	0.997			
Year of first purchase				
2010 (=1)	0.01			
2011 (=1)	0.14			
2012 (=1)	0.29			
2013 (=1)	0.17			
2014 (=1)	0.17			
2015 (=1)	0.22			

Note: Core sample of 3,164 households.

in the choice set if it was chosen by at least 1% of the sample.[3] The resulting choice set includes 12 options: 2 million, 3 million, 4 million, 5 million, 6 million, 7 million, 7.5 million, 8 million, 10 million, 12 million, 15 million, and 20 million. This set covers 93.14% of the observations. For the other 6.86%, the following criterion is used to accommodate the observations in the new set: for values greater than 20 million, the assigned choice is 20 million (0.3% of the observations). For values lower than 2 million, the assigned choice is 2 million (0.82% of the observations). For other values, the nearest option is assigned (for example 3.2 million becomes 3 million and 3.7 million becomes 4 million), and for values in the middle, the nearest lower option is assigned (for example 3.5 million becomes 3 million). Table 1.2 compares the percentage of households choosing the selected insured amounts before and after the reassignment.[4]

Table 1.3 summarizes the annual premium for the insured amounts chosen by at least 1% of the sample. For the analysis, I need for each household the pre-

[3]In total, 52 values are chosen, but 19 values contain one person each, and 31 values concentrate less than ten people each.

[4]As a robustness check, Appendix A.2 presents the analysis only including the 93.14% of households who choose one of the focal options. The main message is unchanged.

Table 1.2: Summary of Insured Amounts: Actual Choices vs. Discretized Choices

Insured Amount (in CRC)	Percentage of Households	
	Actual Choices	Discretized Choices
2 million	2.34	3.60
3 million	4.80	5.25
4 million	5.91	6.32
5 million	20.76	21.05
6 million	7.65	8.00
7 million	4.05	4.08
7.5 million	1.58	1.61
8 million	7.93	9.13
10 million	27.56	28.00
12 million	4.11	4.96
15 million	4.93	5.97
20 million	1.52	2.02

Note: Core sample of 3,164 households.

mium associated with each one of the 12 options in the choice set. Although the insurance company did not provide the exact price rule they follow, they indicated that premiums depend only on age, occupation, gender, and the respective insured amount. Except for occupation, the dataset contains all the price-relevant variables, and thus I use a regression approach to estimate the premiums that each household faced on its unchosen options. An important consideration is that in July 2014 the premium decreased by about 2% for each million. The core sample contains a total of 2,150 households (67.95% of total sample) before the price change and 1,014 afterward.

In addition to prices paid by the households in the sample, the company provided 112 examples of how different combinations of demographics translate to premiums. I include these examples in the regression data to get better estimates. Table 1.4 presents the results. The R^2 is close to one in both estimations, a good indication of the predictive power of the variables, even with one omitted variable (due to the absence of the occupation variable).

Table 1.3: Summary of Annual Premiums

Insured Amount (in CRC)	Mean	SD	Min	Max
2 million	60,835.79	59,236.40	18,828	520,440
3 million	77,392.19	27,096.93	33,912	164,460
4 million	97,459.51	31,809.35	42,120	199,332
5 million	105,041.00	54,559.10	43,212	1,145,196
6 million	120,501.20	41,396.77	56,580	353,748
7 million	132,009.60	36,020.62	66,012	318,564
7.5 million	122,009.30	28,983.58	70,728	244,644
8 million	142,173.40	51,699.71	75,432	376,764
10 million	142,399.20	57,141.03	55,344	1,061,652
12 million	156,872.20	46,044.19	115,860	455,592
15 million	191,122.60	72,665.04	69,144	928,728
20 million	268,605	154,958.10	71,976	910,188
Full Core sample	132,177.61	101,896.98	11,064	2,502,144

Notes: Core sample of 3,164 households. The values correspond to the actual premiums paid when the insurance was first purchased.

Table 1.4: Annual Premiums on Individual Characteristics

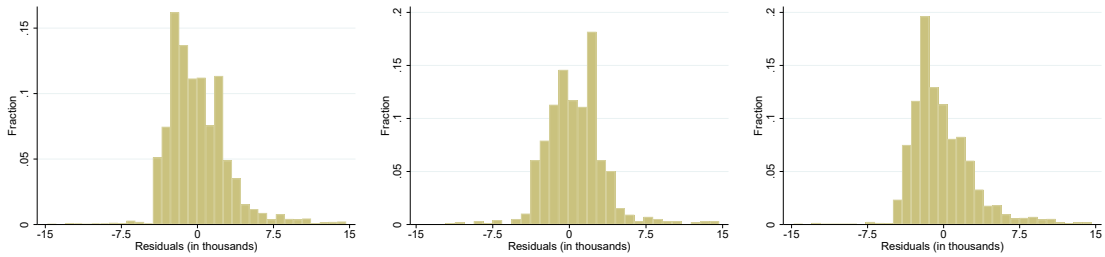
	(1) Log premium, pre-July 2014	(2) Log premium, post-July 2014
Log of insured amount	0.971***	0.989***
	(0.016)	(0.007)
Age	0.006**	0.007***
	(0.003)	(0.002)
Age^2/100	0.040***	0.038***
	(0.004)	(0.003)
Female	-0.206***	-0.242***
	(0.009)	(0.006)
Intercept	-4.310***	-4.635***
	(0.278)	(0.124)
Observations	2,183	1,093
Adjusted R^2	0.92	0.98

Note: Robust Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To test the predictions derived from the regression for the core sample of 3,164 households, Figure 1.1 presents a histogram of the residuals (defined as the difference between actual and fitted premiums at each household's chosen insured amount). For the pre-July 2014 group, the mean absolute error equals 6,156.99, that represents 4.70% of the average premium. For the post-July 2014 group, the mean absolute error equals 4,297.66, that represents 3.19% of the average premium. Overall, the mean absolute error equals 5,561.11 that represents 4.21% of the average premium. Furthermore, there is no systematic relationship between the size of the residuals and the insured amounts. The correlation coefficient is -0.007 for the whole sample, and 0.005 when considering only insured amounts with at least 1% of observations. To estimate the model, I construct the household-specific menu using the prices obtained from the regressions in Table 1.4, i.e., I use fitted premium values even for the actual choices.

Figure 1.1: Histograms of residuals.



All sample | Pre-July 2014 sample | Post-July 2014 sample

Notes: Core sample of 3,164 households. Residuals correspond to the difference between actual and fitted premiums at each household's chosen insured amount.

The data record the claim history for each policy and describes the reason. For the estimation of the model, I obtain empirical frequencies from the claim data, and I assign the same value to each household.[5] One limitation to estimate the claim probabilities is the fact that the data do not contain information about

[5] As a robustness check, Appendix A.3 presents the analysis using empirical frequencies estimated from national level data. The main message is unchanged.

people who were customers of the company and left. One way to think about this limitation is that the whole group is composed of three subgroups: (a) the customers who are still active in December 2015 and did not suffer any mishap, (b) the people who suffer any occurrence while they were customers, and (c) the people who were customers and left before December 2015. I cannot see what happened to the people in the last subgroup, and it might bias the claim rates upwards.

One additional limitation in the estimation of individual-specific claim rates based on observables is the small number of occurrences of the adverse events. The 5,236 households in the full dataset comprise 9,152 individuals and 29,880 individual-year observations. Among all the individual-year observations, the data report 15 deaths, one incapacitation, and 16 serious diseases. In such cases, a maximum likelihood estimation of a logistic model can suffer from downward bias (King and Zeng, 2001). As a solution to this problem, a penalized maximum likelihood, as proposed by Firth (1993), can be estimated. However, for 99.67% of observations, at least one predicted probability is lower than 0.01. This variation in probabilities seems to be so low for people to take into consideration the differences. Therefore, to estimate the model, I use the empirical frequency of the adverse event. Moreover, because the insurance pays the same amount in case of death or incapacity, I combine both states, and I consider two adverse events in my estimation: death or incapacity and serious disease. The empirical frequency in both cases correspond to $\frac{16}{29880} \approx 0.0005$.

1.3 Model and Estimation Strategy

1.3.1 Theoretical Model

In this section, I present a theoretical model for how a household chooses from the menu of premium-insured-amount options.

Although the insurance offers coverage in three possible states, in the cases of death and of incapacity the payoff is the same, and thus it is not possible to separately identify the corresponding parameters. Therefore, in the model, both states are combined. The model's possible states are denoted as life (L), death or incapacity ($D\&I$), and serious disease (S). Strictly speaking, by life I mean a state where none of the other events happened. Let μ_L, $\mu_{D\&I}$, and μ_S be the probability of life, death or incapacity, and serious disease, respectively. I assume that the events are mutually exclusive and $\mu_L + \mu_{D\&I} + \mu_S = 1$.

For the model, I need a monetized loss associated with each state. This loss plays an analogous role to property damage in a model of home or auto insurance. In the case of life insurance, the loss includes the present value of the future labor income that is foregone due to the adverse event. For simplicity, I assume the loss is the same across the two adverse states, i.e., $l_{D\&I} = l_S \equiv l$.

The insurance company offers to the household a set of pairs $\{(p(m), m) : m \in \mathcal{M}\}$ where $p(m)$ is the premium associated with insured amount m and $\mathcal{M}$ is the set of insured-amount options. In the case of death or incapacity, the household receives the amount m, and in the case of a serious disease, the household receives the amount $0.5m$. The lottery implied by choice m takes the form:

$$L(m) \equiv (-p(m), \mu_L; -l - p(m) + m, \mu_{D\&I}; -l - p(m) + 0.5m, \mu_S).$$

Given an initial wealth w and a utility function $u(w)$, the standard expected utility of the lottery $L(m)$ is:

$$\begin{aligned} U(L(m)) = \mu_L u(w - p(m)) + \mu_{D\&I} u(w - l - p(m) + m) \\ + \mu_S u(w - l - p(m) + 0.5m). \end{aligned} \tag{1.1}$$

The assumption that the utility function is state-independent is difficult to justify in the life-insurance context. For instance, the same amount of money will have different levels of enjoyment depending if the individual is healthy rather than in a vegetative state. Therefore, the states might have the effect of causing the whole utility function to change. To introduce state-dependent preferences, I follow Viscusi and Evans (1990) and include a multiplicative parameter that changes the marginal utility of consumption across states. Let $\alpha_{D\&I}$ and α_S be the respective state-dependent terms relative to the life state. The expected utility of the lottery $L(m)$ becomes:

$$\begin{aligned} U(L(m)) = \mu_L u(w - p(m)) + \mu_{D\&I} \alpha_{D\&I} u(w - l - p(m) + m) \\ + \mu_S \alpha_S u(w - l - p(m) + 0.5m). \end{aligned} \tag{1.2}$$

If $\alpha_i > 1$ then the marginal utility of consumption is greater in the adverse-event state i (i=$D\&I$, S) relative to the life state. On the other hand, if $\alpha_i < 1$ then the marginal utility of consumption deteriorates in the adverse-event state i relative to the life state. Finally, if $\alpha_i = 1$ the marginal utility of consumption is

the same across states. A priori, the magnitude of the state-dependent parameter is ambiguous. As pointed out by Finkelstein et al. (2009), it is conceivable that the adverse event lowers the marginal utility of consumption of goods that are complements to good health, such as travel, but increases the marginal utility of consumption of goods that are substitutes to good health, such as nursing care. In fact, the evidence provided by empirical studies is also inconclusive in the direction of the effect.

In the expected-utility model, the concavity of the utility function generates risk aversion. However, an alternative source of risk aversion to consider in the domain of life insurance is the probability-weighting feature of prospect theory (Kahneman and Tversky, 1979). The probability-weighting feature suggests that people transform probabilities into decision weights that might differ from the objective probabilities. Therefore, each adverse event has a transformed probability, and the remaining weight summing up to one belongs to the life state, i.e., $\Omega_L(\mu_L)+\Omega_{D\&I}(\mu_{D\&I})+\Omega_S(\mu_S) = 1$, where $\Omega_i(\mu_i)$ represents the weight associated with each state and reflects how people perceive the corresponding probabilities in the decision. When considering probability weighting and continuing to assume state-dependent utility, the expected utility of the lottery implied by the lottery $L(m)$ is given by:

$$\begin{aligned} U(L(m)) = {} & (1 - \Omega_{D\&I}(\mu_{D\&I}) - \Omega_S(\mu_S))u(w - p(m)) \\ & + \Omega_{D\&I}(\mu_{D\&I})\alpha_{D\&I}u(w - l - p(m) + m) \\ & + \Omega_S(\mu_S)\alpha_S u(w - l - p(m) + 0.5m) \end{aligned} \tag{1.3}$$

In principle, because α_i is independent of μ_i, while $\Omega_i(\mu_i)$ depends on μ_i, the state-dependent parameters and the probability distortions might be separately

identified. However, because my data effectively contains no variation in the adverse-event probabilities, in practice $\Omega_i(\mu_i)$ is not distinguishable from α_i.

Hence, I rewrite equation (1.3) as:

$$\begin{aligned} U(L(m)) = u(w - p(m)) + \Gamma_{D\&I} u(w - l - p(m) + m) \\ + \Gamma_S u(w - l - p(m) + 0.5m) \end{aligned} \tag{1.4}$$

where $\Gamma_{D\&I} = \dfrac{\Omega_{D\&I}(\mu_{D\&I})\alpha_{D\&I}}{1 - \Omega_{D\&I}(\mu_{D\&I}) - \Omega_S(\mu_S)}$ and $\Gamma_S = \dfrac{\Omega_S(\mu_S)\alpha_S}{1 - \Omega_{D\&I}(\mu_{D\&I}) - \Omega_S(\mu_S)}$.

The analysis can recover the parameters $\Gamma_{D\&I}$ and Γ_S. Equation (1.4) encompasses both the state-dependent-utility and probability-weighting models. If the conditions $\Omega_S(\mu_S) = \mu_S$ and $\Omega_{D\&I}(\mu_{D\&I}) = \mu_{D\&I}$ are assumed, then $\alpha_{D\&I} = \Gamma_{D\&I}\dfrac{\mu_L}{\mu_{D\&I}}$ represents the change in the marginal utility due to death or incapacity (similarly for the serious-disease state). On the other hand, under the condition $\alpha_{D\&I} = \alpha_S = 1$ then $\dfrac{\Gamma_{D\&I}}{\Gamma_S}$ represents the relative probability weights between the adverse events. Due to this issue, in the analysis I present the estimates for $\Gamma_{D\&I}$ and Γ_S and discuss both interpretations.

In the estimation, for the utility function $u(\cdot)$ I assume the constant absolute risk aversion (CARA) utility function $u(x) = -e^{-rx}$ for $r \neq 0$ and $u(x) = x$ for $r = 0$, where r is the coefficient of absolute risk aversion. A disadvantage of this utility function is that the family of constant relative risk aversion (CRRA) often gives a better fit (Wakker, 2008). In my case, the CRRA approach requires additional assumptions because the data do not contain any wealth measure.[6]

[6] As a robustness check, Appendix A.4 presents the analysis assuming CRRA utility. The main message is unchanged.

Using the CARA form equation (1.1) can be written as:

$$U(L(m)) = \mu_L(-e^{-r(w-p(m))}) + \mu_{D\&I}(-e^{-r(w-l+m-p(m)}))$$
$$+ \mu_S(-e^{-r(w-l+0.5m-p(m))}). \tag{1.5}$$

After a monotonic transformation equation (1.5) becomes:

$$U(L(m)) = \mu_L(-e^{-r(-p(m))}) + \mu_{D\&I}(-e^{-r(m-p(m)-l)})$$
$$+ \mu_S(-e^{-r(0.5m-p(m)-l)}). \tag{1.6}$$

Similarly, with CARA utility function and after a monotonic transformation, equation (1.4) becomes:

$$U(L(m)) = (-e^{-r(-p(m))}) + \Gamma_{D\&I}(-e^{-r(m-p(m)-l)})$$
$$+ \Gamma_S(-e^{-r(0.5m-p(m)-l)}). \tag{1.7}$$

1.3.2 Identification

The identification of the model consists of two pieces. The first piece is whether it is possible to identify the parameters in the model whenever the maximum willingness to pay (WTP) for a particular insured amount is known. The answer is affirmative to both the standard expected-utility model and the model that incorporates state-dependent utility and probability distortions as long as the choice set includes more than three insured amount options.

First, consider the case of the standard expected-utility model, where the parameters to be estimated are the coefficient of absolute risk aversion r and the size of the loss l. Proposition 1.1 shows that the parameter vector $\theta \equiv (r, l)$ is identified (the proof is in Appendix A.1). Consider a household with a coefficient of absolute risk aversion $r \neq 0$, a monetary loss l, and a vector of probabilities $(\mu_L, \mu_{D\&I}, \mu_S)$:

Proposition 1.1. *Take any three insured amount options:* $0 \leq m_a < m_b < m_c$, *with premium* p_a *for insured amount* m_a. *Denote by* $\tilde{p}_j$ $(j = b, c)$ *the premium that makes the household indifferent between insured amounts* m_a *and* m_j, *and assume that* $0.5(m_j - m_a) > \tilde{p}_j - p_a$. *Under standard expected utility, any* $(p_a, \tilde{p}_b, \tilde{p}_c)$ *is consistent with a unique* (r, l).

Table 1.5 illustrates the intuition. Multiple combinations of r and l are consistent with a particular WTP to move from insurance amount m_a to m_b. However, each combination delivers a different implication for the WTP to move from insurance amount m_a to m_c, where a lower WTP corresponds to higher levels of r and lower size of l.

Table 1.5: Intuition for Standard Expected-Utility Model Identification

r	l	WTP	
		m_a to m_b	m_a to m_c
0.09	27.3799	2.7528	4.5891
0.10	25.0000	2.7528	4.5071
0.11	23.0513	2.7528	4.4288

Note: $\mu_{D\&I}$=0.05, μ_S=0.03, m_a=0, m_b=8, and m_c=16.

The assumption $0.5(m_j - m_a) > \tilde{p}_j - p_a$ in Proposition 1.1 is a regularity condition that reasonable premiums will satisfy. As Table 1.3 shows, the condition always holds in the data.

Now consider the model that includes state-dependent utility and probability

distortions, where the parameters to be estimated are r, l, $\Gamma_{D\&I}$, and Γ_S. It is possible to show that for any r, the parameters l, $\Gamma_{D\&I}$, and Γ_S are not separately identified. Consequently, in the estimation, I choose a value for l based on the results from the standard expected-utility model. Once l is fixed, the vector $\theta_{|l} \equiv (r, \Gamma_S, \Gamma_{D\&I} \mid l)$ is identified. While the complexity of the problem prevents me from giving a formal proof of the identification argument, I provide the intuition in Table 1.6. Different from the standard expected-utility model, in this case, knowing the WTP for three insured amounts is not sufficient to identify the parameters. Table 1.6 shows that multiple combinations of the parameters can explain simultaneously the WTP to move from insurance amount m_a to m_b and from m_a to m_c. However, as the number of insured amount options available increases, the different parameters will begin to generate different WTP.

Table 1.6: Intuition for State-Dependent-Utility and Probability-Weighting Model Identification

r	$\Gamma_{D\&I}$	Γ_S	WTP			
			m_a to m_b	m_a to m_c	m_a to m_d	m_a to m_e
0.09	0.0674	0.0192	0.7638	1.4532	2.6247	4.2720
0.10	0.0500	0.0300	0.7638	1.4532	2.6252	4.2790
0.11	0.0371	0.0346	0.7638	1.4532	2.6255	4.2829

Note: l=25, m_a=0, m_b=2, m_c=4, m_d=8, and m_e=16.

The logic behind this identification strategy comes from Barseghyan et al. (2013). In my case, given the higher number of parameters, more insured amount options are needed. As explained in Section 1.2, the insurance allows any insured amount between CRC 1 million and CRC 60 million, so in principle I have a continuum of options. In practice, people's choices are concentrated on 12 options. The existence of more than three insured amounts options allows me to separately identify the parameters in my model.

The second piece of identification is whether from the data it is possible to

identify these WTPs. To identify WTPs requires variation in premiums that is exogenous to the households' risk preferences. In my setting, one relevant source of exogenous variation in premiums is due to the price change in July 2014. For the homogeneous model, not only the price change contributes to generating variation in premiums, but also the changes in demographics (age and gender). On the other hand, for the heterogeneous model, I assume that the utility parameters depend on demographics, and therefore the variation in premiums is due to the price change, that in practice might be insufficient to identify the WTP. Hence, while the model converges, the results from the heterogeneous model should be interpreted with caution.

1.3.3 Estimation

To estimate the models described above, I use a McFadden (1974) random-utility specification. Specifically, given an underlying utility function $U(L(m), \theta)$ (e.g., from equation (1.6) or equation (1.7)) the utility associated with insured amount m in the choice set $\mathcal{M}$ is given by:

$$\mathcal{U}(m, \theta) \equiv U(L(m), \theta) + \varepsilon(m)$$

where $\varepsilon(m)$ is a noise term, assumed to be i.i.d. with a type-1 extreme value distribution with scale parameter σ.

To construct the likelihood function for the estimation, note that a household chooses an insured amount m if $\mathcal{U}(m, \theta) > \mathcal{U}(m', \theta)$ for all $m' \neq m$, and therefore the probability of choosing the insured amount m is given by:

$$
\begin{aligned}
Pr(m \mid \theta, \sigma) &\equiv Pr(\varepsilon(m') - \varepsilon(m) < U(L(m), \theta) - U(L(m'), \theta) \quad \forall m' \neq m) \\
&= \frac{exp(U(L(m), \theta)/\sigma)}{\sum_{m' \in \mathcal{M}} exp(U(L(m'), \theta)/\sigma)}.
\end{aligned}
$$

The estimates in the standard expected-utility model are the coefficient of absolute risk aversion r, the size of the loss l, and the scale of choice noise σ. For the state-dependent-utility and probability-weighting model, I assume a value for the loss l based on the results from the standard expected-utility model. In this case, the estimates are the coefficient of absolute risk aversion r, the parameters $\Gamma_{D\&I}$ and Γ_S, and the scale of choice noise σ.

1.4 Results

1.4.1 Estimation under Homogeneous Preferences

In this section, I estimate the model assuming homogeneous preferences. I begin my analysis using the standard expected-utility model to estimate the parameter vector $\theta \equiv (r, l)$ and the scale of choice noise σ. As stated in Section 1.2, all households are assumed to face the same probabilities, equal to the empirical frequencies. The corresponding probability is 0.0005 for both events. Table 1.7 presents the results.

The estimate for $\ln(l)$ implies that the monetary loss l is about CRC 25,146,944. According to the Costa Rican National Household Survey (*Encuesta Nacional de Hogares [ENAHO]*), between 2010 and 2015 the annual per capita household income was CRC 3,933,668. Thus, this estimate would imply that an adverse event

Table 1.7: Estimation of the Model Assuming Standard Expected Utility ($U(L(m))$ Specified by Equation (1.6))

	Estimate	95 percent bootstrap confidence interval	
$\ln(r)$	-15.6044	-15.6689	-15.5349
$\ln(l)$	17.0402	16.9875	17.0901
σ	0.0046	0.0040	0.0055
LL	-7081.52		

Note: Core sample of 3,164 households.

represents a loss equivalent to more than six years of income. This fact highlights the high-stakes nature of the lottery considered in this chapter.

The estimate for $\ln(r)$ implies that r is 0.000000167. This coefficient of risk aversion might seem close to zero and therefore of little relevance. However, this perspective omits the fact that the insured amounts are in millions. To evaluate the importance of the risk-aversion coefficient, I calculate the maximum willingness to pay (WTP) for a household to increase its insured amount from CRC 2 million to CRC 4 million when the premium for a CRC 2 million coverage is CRC 37,473.[7] Table 1.8 displays how the WTP varies for different values of the risk-aversion coefficient. The results show that differences in r generate meaningful differences in the WTP.

Regarding the fit of the model, without choice noise, the model predicts 23% of observations correctly. When I consider choices within one rank (i.e., one step up or down on the menu of insured amount options) the number of correct predictions increases to 57%. Given that the choice set consists of 12 options, these results indicate that the model provides a good fit to life-insurance choices.

Finally, consider the impact of the estimated scale of choice noise of 0.0046.

[7]This amount corresponds to the average premium for a CRC 2 million coverage for the core sample.

Table 1.8: Economic Significance Standard Expected-Utility Model

r	$\ln(r)$	WTP
0	$-\infty$	1,498.50
0.000000157	-15.6689	49,668.01
0.000000167†	-15.6044	62,307.93
0.000000179	-15.5349	80,701.69

Note: WTP denotes (for a household with the utility function in equation (1.6), and the utility parameters from Table 1.7) the household's maximum willingness to pay to increase its insured amount from CRC 2 million to CRC 4 million when the premium for the CRC 2 million coverage is CRC 37,473.
† Estimate from Table 1.7

Over 10,000 simulations, the households' "noiseless" choices coincide with the "noisy" choices 21% of the time. This number increases to 56% if I consider the choices within one rank. Because with extreme noise, the probability that a household's "noisy" choice would equal any insurance amount would be around 8%, the noise seems to be not dominant in the estimations.

I next estimate the model with state-dependent utility and probability weighting to obtain the parameter vector $\theta_{|l} \equiv (r, \Gamma_S, \Gamma_{D\&I} \mid l)$ and the scale of choice noise σ. Given the results from Table 1.7, I assume that the loss is equal to CRC 25 million. Table 1.9 reports the results.

Table 1.9: Estimation of the Model Assuming State-Dependent Utility and Probability Weighting ($U(L(m))$ Specified by Equation (1.7))

	Estimate	95 percent bootstrap confidence interval	
$\ln(r)$	-15.8962	-15.9543	-15.8025
$\Gamma_{D\&I}$	0.0021	0.0015	0.0024
Γ_S	0.0000	0.0000	0.0003
σ	0.0034	0.0030	0.0041
LL	-7079.75		

Notes: Core sample of 3,164 households. The loss is assumed to be equal to CRC 25 million.

The estimate for $\ln(r)$ implies that r is 0.000000125, lower than the one corresponding to the expected-utility case. The parameter $\Gamma_{D\&I}$ is equal to 0.0021 and Γ_S is close to zero. Recall from Section 1.3 that the parameters $\Gamma_{D\&I}$ and Γ_S can represent state-dependent utility or probability weighting. According to the state-dependent-utility interpretation, the marginal utility of consumption changes across states, and people do not engage in probability distortions. Then $\alpha_{D\&I} = \Gamma_{D\&I}\dfrac{\mu_L}{\mu_{D\&I}} \approx 4.1610$ represents the change in the marginal utility due to death or incapacity; and $\alpha_S = \Gamma_S\dfrac{\mu_L}{\mu_S} \approx 0.00001$ represents the change in the marginal utility due to serious disease. The results suggest that the death or incapacity state increases the marginal utility of consumption, while the serious-disease state reduces it.

According to the probability-weighting interpretation, people transform probabilities into decision weights that do not necessarily coincide with the probabilities. Then $\dfrac{\Gamma_{D\&I}}{\Gamma_S} \approx 538{,}195.47$ is the relative weight people assign to the death or incapacity state relative to the serious-disease state. Because both events have the same objective probabilities, the result suggests a relative overweighting of the death or incapacity state.

Without choice noise, the state-dependent-utility and probability-weighting model predicts 22% of observations correctly. This number increases to 56% if I consider the choices within one rank. Regarding the role of the noise in the simulations, over 10,000 simulations, the households' "noiseless" choices coincide with the "noisy" choices 22% of the times. This number increases to 57% if I consider the choices within one rank. Similar than before, the noise seems to be not dominant.

1.4.2 Estimation under Heterogeneous Preferences

Previous papers that estimate risk preferences using individual-level insurance data find substantial preference heterogeneity (e.g., Cohen and Einav 2007a; Barseghyan et al. 2013). In this section, I extend the model to allow for observed heterogeneity in r, $\Gamma_{D\&I}$ and Γ_S. In particular, I assume for individual j:

$$\ln(r_j) = \boldsymbol{\beta}_r \mathbf{X}_j$$
$$\Gamma_{D\&I,j} = \boldsymbol{\beta}_{\Gamma_{D\&I}} \mathbf{X}_j$$
$$\Gamma_{S,j} = \boldsymbol{\beta}_{\Gamma_S} \mathbf{X}_j$$

where the elements in the vector $\mathbf{X}_j$ are the following: a constant, a dummy variable that is equal to one if the policyholder is a female, and three dummy variables that indicate an age interval. For the age intervals, I divide the sample into four categories: less than 29 years old (22% of households), from 30 to 39 years old (34%), from 40 to 49 years old (28%), and older than 50 years old (16%).

The model is estimated via maximum likelihood and the parameter vectors to be estimated are: $\theta_{HET} \equiv (\boldsymbol{\beta}_r, \sigma)$ for the standard expected-utility case and $\theta_{HET} \equiv (\boldsymbol{\beta}_r, \boldsymbol{\beta}_{\Gamma_{D\&I}}, \boldsymbol{\beta}_{\Gamma_S}, \sigma)$ for the model assuming state-dependent utility and probability weighting. In the estimations, I continue assuming that the loss is equal to CRC 25 million. It is worth remembering that, in the heterogeneous model, identification is more suspect because the variations in premiums necessary to identify the maximum willingness to pay are due exclusively to the price change in July 2014 (as discussed in Section 1.3.2).

Table 1.10 presents the results for the standard expected-utility model under

observed heterogeneity. For convenience, the first column restates the relevant estimates from Table 1.7. The mean fitted value of r is 0.000000146, and along with the scale of choice noise, the results are similar to the estimates in Table 1.7.

Table 1.10: Estimation of the Model Assuming Standard Expected Utility. Observed Heterogeneity

		Homogeneous preferences	Observed heterogeneity		
			Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_r$	Constant	-15.6044	-15.5585	-15.5757	-15.5385
	Female		-0.0184	-0.0337	-0.0039
	$30 \leq$ Age ≤ 39		-0.0285	-0.0435	-0.0138
	$40 \leq$ Age ≤ 49		-0.0555	-2.0241	-0.0404
	$50 \leq$ Age		-1.8480	-2.0241	-0.0016
Mean fitted value	$\ln(r)$	-15.6044	-15.7398	-16.0530	-15.5902
σ		0.0046	0.0042	0.0037	0.0049
LL		-7081.52	-6993.56		

Notes: Core sample of 3,164 households. The loss is assumed to be equal to CRC 25 million.

Table 1.11 provides the results for the case of the model with state-dependent utility and probability weighting. The mean fitted value of r is 0.000000206, higher than the estimated value in Table 1.9. The mean fitted value for $\Gamma_{D\&I}$ is 0.0009 and for Γ_S is 0.0001. The state-dependent-utility interpretation of these parameters is $\alpha_{D\&I} \approx 1.7052$ and $\alpha_S \approx 0.1778$. This interpretation suggests that the death or incapacity state increases the marginal utility of consumption, and the serious-disease state reduces it. The probability-weighting interpretation suggests that the ratio between the weight assigned to death-or-incapacity state and the serious-disease state is approximately $\dfrac{\Gamma_{D\&I}}{\Gamma_S} \approx 9.59$, indicating a relative overweighting of the death or incapacity state. These results are consistent with the findings in Table 1.9.

Moreover, the results from Table 1.11 suggests that as people age, they worry

Table 1.11: Estimation of the Model Assuming State-Dependent Utility and Probability Weighting. Observed Heterogeneity

		Homogeneous preferences	Observed heterogeneity		
			Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_r$	Constant	-15.8962	-15.6860	-16.0233	-15.3505
	Female		0.6076	0.2892	0.7896
	$30 \leq$ Age ≤ 39		-0.0654	-0.1192	0.0234
	$40 \leq$ Age ≤ 49		-0.2090	-0.5095	-0.1015
	$50 \leq$ Age		-1.2049	-2.1611	-0.1480
$\boldsymbol{\beta}_{\Gamma_{D\&I}}$	Constant	0.0021	0.0016	0.0006	0.0030
	Female		-0.0016	-0.0029	-0.0006
	$30 \leq$ Age ≤ 39		0.0000	-0.0003	0.0000
	$40 \leq$ Age ≤ 49		0.0001	0.0000	0.0015
	$50 \leq$ Age		0.0028	-0.0001	0.0040
$\boldsymbol{\beta}_{\Gamma_S}$	Constant	0.0000	0.0000	0.0000	0.0003
	Female		0.0001	0.0000	0.0004
	$30 \leq$ Age ≤ 39		0.0000	0.0000	0.0003
	$40 \leq$ Age ≤ 49		0.0000	-0.0002	0.0001
	$50 \leq$ Age		0.0000	-0.0001	0.0009
Mean fitted value	$\ln(r)$	-15.8962	-15.3944	-15.8754	-14.9776
	$\Gamma_{D\&I}$	0.0021	0.0009	0.0002	0.0014
	Γ_S	0.0000	0.0001	0.0000	0.0006
σ		0.0034	0.0064	0.0032	0.0157
LL		-7079.75	-6957.39		

Notes: Core sample of 3,164 households. The loss is assumed to be equal to CRC 25 million.

relatively more about death or incapacity. According to the Costa Rican National Household Survey, the head of household's average age was 49.81 years old between 2010 and 2015. Therefore, the presence of dependents might be one reason for people increasing their concern on the death or incapacity state after their forties.

The results from Table 1.11 also indicate heterogeneity by gender. First, women are more risk averse than men, the mean fitted value of $\ln(r)$ for women is -15.3574 and for men -15.9088. Second, compared to men, women assign a lower weight to the death or incapacity state and a higher weight to the serious-disease state. The corresponding mean fitted value of $\Gamma_{D\&I}$ for women is 0.0005 and for men

0.0020, while the mean fitted value of Γ_S is 0.0001 for women and 0.00001 for men. Following a state-dependent-utility interpretation, the different decision weights translate to an asymmetric change in the marginal utility of consumption of the adverse event across genders. The increase in the marginal utility of consumption due to death or incapacity is higher for men than for women, and similarly, the drop in marginal utility due to serious disease is stronger for men. According to a probability-weighting interpretation, the distortion in probabilities operates in different forms across genders. Although both men and women inflate the actual probability of death or incapacity state with respect to the chances of the serious-disease state, the overweighting is more pronounced for men.

A likelihood ratio test rejects at the one percent level the hypothesis of homogeneous preferences for both the standard expected-utility model and for the state-dependent-utility and probability-weighting models. Therefore, similar to previous studies, I find support for the existence of heterogeneity in preferences. Between the estimates under heterogeneous preferences, the model that assumes state-dependent utility and probability weighting fits the data significantly better than the standard expected-utility model.

1.5 Robustness Checks

My main results suggest that (a) the state-dependent-utility and probability-weighting model fits significantly better the data than the standard expected-utility model, (b) accounting for heterogeneity in preferences is important, (c) according to a state-dependent utility interpretation, both states change the marginal utility of consumption, but the magnitude is larger for the death or incapacity

state, and (d) according to a probability-weighting interpretation, the death or incapacity state is overweighted relative to the serious-disease state. In this section, I highlight that these results are preserved across alternative specifications. To conserve space, a more detailed description of each robustness check and complete results are available in Appendices A.2, A.3, and A.4.

In the first robustness check (Appendix A.2), I investigate the sensitivity of my estimates to reassigning insured amounts to households that do not choose a focal point. I re-estimate the model using a restricted sample in which I drop all households with reassigned choices (6.86% of the core sample). In the second robustness check (Appendix A.3), I assess the possibility of results being affected because of using empirical frequencies estimated from the claim data provided by the insurance company. Instead, I estimate empirical frequencies using national-level data and re-estimate the model. In the last robustness check (Appendix A.4), I study how results change after assuming constant relative risk aversion (CRRA) utility. The CRRA utility requires household's prior wealth, that is unobservable from the data. Therefore, I re-estimate the model assuming different reasonable values for prior wealth according to the Costa Rican National Household Survey. All in all, the results from the robustness checks reinforce my main message.

1.6 Conclusion

In this chapter, I estimate a structural model of risky choice using data on choices in life-insurance contracts. In addition to standard risk aversion (diminishing marginal utility for wealth), the model allows for weights on all states that might differ from their actual probabilities. These weights can be interpreted as state-

dependent preferences or probability distortions. To estimate the nature of risk preferences using proprietary data is a different approach from most of the existing work that investigates life-insurance purchases. The existing work has focused on survey data to estimate how demographic and socioeconomic factors are associated with demand.

My results suggest that either state-dependent utility for wealth or decision weights that might differ from their actual probabilities play a statistically and economically significant role in explaining risky choices. Moreover, I find evidence of heterogeneity in preferences. The presence of heterogeneity is consistent with results in Cohen and Einav (2007a), Chiappori et al. (2012), Barseghyan et al. (2013), and Chiappori et al. (2014). My analysis suggests that women are more risk averse and assign different decision weights than men. Also, the decision weights vary across age categories.

According to a state-dependent-utility interpretation of my results, the adverse state changes the marginal utility of consumption. On the one hand, the serious-disease state decreases the marginal utility of consumption. That health deterioration is associated with a drop in the marginal utility of consumption is consistent with findings in Viscusi and Evans (1990), Sloan et al. (1998), and Finkelstein et al. (2013). On the other hand, the death or incapacity state increases the marginal utility of consumption, and the increase is stronger for men. According to the Costa Rican National Household Survey, between 2010 and 2015, on average males earned approximately 61.86% of the household's total income. Therefore, the higher weight on the death or incapacity state in men might reflect a concern due to the potentially high financial loss in case of death. Lillard and Weiss (1997) also find a positive state dependence for health and survival uncertainty in

retirees and explain it through bequest motives.

In contrast, income in the serious-disease state might not be as valuable from the life-insurance perspective because of the financial protection offered by the social security system in Costa Rica. In Costa Rica, the public health system provides near-universal access to a full range of health services, including complex procedures such as transplants (OECD, 2017). Moreover, workers receive a subsidy during sick leave (CCSS, 2014). Therefore, the possibility of catastrophic health expenditures due to serious disease is low.

According to a probability-weighting interpretation of my results, households evaluate risk different from what the corresponding objective probabilities indicate. In particular, my results suggest that individuals overweight the chances of death or incapacity relative to the serious-disease state when buying life insurance. That a model featuring probability weighting is a better approximation to behavior is in line with studies from experimental data (e.g., Camerer and Ho 1994; Wu and Gonzalez 1996), as well as field data (e.g., Chiappori et al. 2012; Barseghyan et al. 2013).

My results reinforce most of the previous conclusions from lab and field studies on the importance of non-expected-utility models and heterogeneity in the estimation of risk preferences. However, it is worth considering certain limitations of my work. First, the life-insurance product I analyze is not compulsory. It is conceivable that individuals with high risk aversion, high probability distortions, or high state-dependent parameters are self-selecting. Therefore, it is not possible to claim I have a representative segment of the population. A second limitation is related to the exogenous variation in premiums that the identification strategy requires. Particularly in the estimation of the heterogeneous model, the exogenous variation

is caused exclusively by the price change in July 2014, and it might be insufficient in practice for identification purposes. A third limitation is that from my analysis the state-dependent parameter is indistinguishable from the probability-weighting component. One possible way to separately identify the two elements is to find a context with probability variations and where state-dependent utility matters.

CHAPTER 2

ADVERSE SELECTION AND MORAL HAZARD IN PRIVATE HEALTHCARE WHEN UNIVERSAL HEALTHCARE IS AVAILABLE

(This chapter was written in collaboration with Jason Somerville and Matthew Thirkettle, Cornell University)

2.1 Introduction

The merits of public versus private provision of healthcare services is a complex issue that generates heated discussions worldwide. Concerns about efficiency, fairness, and fiscal sustainability of different healthcare provision systems permeate debates and reform efforts such as "Medicare for All." A sizable academic literature has contributed to the discussion, although predominantly focusing on US health programs, ignoring the rich institutional variation that exists in other countries that might shed light on diverse ways that the public and private sectors interact in providing healthcare services.

In this chapter, we focus on a middle-income country in which the government provides healthcare services to all citizens, but, in addition, there exists a private healthcare sector. Using proprietary data on household private health-insurance coverage and medical utilization, we examine selection in this institutional setting, particularly selection based on asymmetric information (adverse selection and moral hazard). The insurance company offers the product we analyze in two forms: individual and collective, the difference being whether individuals can choose an insured amount (individual) or it is exogenously assigned (collective).

Both individual and collective insurance have the same features, including copays and deductibles. One unique feature of our data is that within a policy term, the copay and deductible are fixed, but the copay and deductible regimes vary over time. We leverage these features of the insurance product to disentangle adverse selection and moral hazard.[1]

Our estimates are based on a positive correlation test and a random-effect logit model, and suggest the presence of selection based on asymmetric information, both in the form of adverse selection and moral hazard. We disentangle adverse selection and moral hazard by comparing the effect of insured amounts and copay-and-deductible regimes on claim rates in the individual and collective insurance. In the individual-insurance sample, we find that people that choose highest insured amounts also have higher claim rates. This is consistent with adverse selection and moral hazard. We also find that claim rates are decreasing with copays and deductibles, suggesting the presence of moral hazard. On the other hand, in the collective-insurance sample, the insured amount is exogenously assigned, and consequently, any effect is due to moral hazard. In the collective-insurance sample, we find little impact of insured amounts in claim rates and again a negative relationship between copay-and-deductible regimes and claim rates. Comparing the magnitude of the estimates in the individual and the collective samples, we conclude that most of the impact of insured amounts on claiming behavior in the individual sample is a product of adverse selection.

Given our results indicating the presence of asymmetric information, we develop a model of coverage choice and medical spending that embeds four sources of heterogeneity that could give rise to adverse selection and moral hazard in our

[1]Throughout the chapter, we consider adverse selection as selection based on types at the stage of choosing insured amounts, and moral hazard as how the coverage impacts on an agent's ex-post behavior at the stage of choosing medical utilization.

institutional setting: risk preferences, risk type, concern for health, and tastes for convenience provided in the private sector. In future work, we plan to structurally estimate the model to measure the relative importance of each dimension of selection, analyze the degree of heterogeneity in preferences, and to evaluate the welfare implications of a market where private and public sector provides healthcare services.

In Section 2.2, we describe in detail the institutional setting of the healthcare market that we analyze. The government provides universal healthcare financed through payroll deductions. The public system covers medical consultations, prescriptions, surgeries, and complex treatments (e.g., chemotherapy, dialysis, or organ transplants) without charging any copays or deductibles. In addition to the public healthcare system, private healthcare is available at a cost to anyone who chooses to utilize its services. The private healthcare system offers the same range of services as the public healthcare system with comparable health outcomes, but it is superior in terms of convenience services, such as shorter waiting times or private accommodation. We present evidence from a national health survey on the relevance of convenience services. In general, the number of people per room, the admission process, and the food quality are the worst-ranked features in public hospitals. Moreover, using household income and expenditure survey data, we describe how households utilize public and private healthcare services. Private healthcare is more relevant for medical consultations, medications, medical imaging, and laboratory tests. However, households switch to public healthcare for relatively expensive treatments such as hospitalizations or complex treatments. This pattern motivates some of the assumptions we make when specifying consumer preferences and behavior in our modeling approach.

In Section 2.3, we provide an overview of the data. The source of the data is an insurance company that operates in the middle-income country that we are analyzing. The company offers a product that we call "Student Health Insurance." Student Health Insurance is available for any person enrolled in an educational institution and covers health costs due to accidents, incapacity, and death. The insurance company offers Student Health Insurance in two forms. The first form is individual insurance, where individuals select the maximum amount they will receive in case of an adverse event from a menu of options. The second form is collective insurance, where individuals receive their policy from their educational institution, and the insured amount is chosen by that institution. In both cases, the insurance's features are the same, and individuals can decide the amount of ex-post medical utilization. We have individual-level data on insured amounts and medical utilization for both types of customers. For reasons we explain, we restrict attention to accident coverage. Crucially for disentangling adverse selection and moral hazard, the insurance company implemented over time various changes in copays and deductibles. During the years that cover our data, the company varied copays four times, including a 0% copay regime and a 25% copay regime.

In Section 2.4.1, we present our results for the individual-insurance sample. Our results suggest the presence of asymmetric information, both in the form of adverse selection and moral hazard. Specifically, we find a positive relationship between the insured amount chosen and claim rates. The group selecting the highest insured amount has claim rates about ten percentage points higher than the group selecting the lowest insured amount. Similarly, conditional on a positive utilization, the group selecting the highest insured amount has claim amounts on average three times higher than the group selecting the lowest insured amount. This insured-amount effect can be indicative of both adverse selection and moral

hazard. We also find a negative relationship between copays and deductibles and claim rates. An increase in copays from zero to a positive percentage reduces claim rates approximately two percentage points and reduces average claim amounts 17 percent. We interpret this copay-and-deductible effect as moral hazard. Through a random-effect logit model, we confirm that the insured amount and the copay and deductible effects are robust to the inclusion of covariates.

In Section 2.4.2, we report our results for the collective-insurance sample. Because, in this case, people cannot decide on the insured amount they want, the adverse selection channel is shut down. Consistently we do not find any positive relationship between insured amounts and claim rates (claim amounts). However, we show a negative relationship between copays and deductibles and claim rates (claim amounts). An increase in copays from zero to a positive percentage reduces claim rates nearly one percentage point and reduces average claim amounts 20 percent.

In Section 2.4.3, we compare the individual and collective samples by estimating a random-effect logit model that pools both samples. The sign and magnitude of the insured amounts and the copay-and-deductible regimes are similar to the results we obtain from analyzing each sample separately. Therefore, we argue that the positive relationship between insured amount and claim behavior in the individual sample is mostly a result of adverse selection. Similarly, the negative impact of copay and deductible effects on claims in both samples is mostly a result of moral hazard.

In Section 2.5, we develop a theoretical model of coverage choice and medical spending that incorporates four types of heterogeneity that could generate adverse selection and moral hazard in our institutional setting: risk preferences, risk type,

concern for health, and tastes for convenience. We follow a modeling approach closely related to Cardon and Hendel (2001), which is also the approach used in Einav et al. (2013). The model consists of two stages. In the first stage, the agent chooses an insurance contract based on her risk preferences and her forecast of medical utilization. In the second stage, conditional on a health-shock realization that depends on the individual's risk type, the agent decides whether to receive treatment in the public or private sector. This choice is determined by how much the agent values the convenience services provided in private healthcare, by how much the agent values health, and by the magnitude of the health shock.

In Section 2.5.1, we discuss the identification of our model. The basic intuition is that under the zero-copay regime, individuals always prefer private healthcare to public healthcare. Therefore, we can recover the full distribution of health shocks from the data. After the company introduces a positive copay, individuals compare their utility of receiving public versus private treatment, and for some range of health shocks, preferences for convenience will favor private healthcare. However, as the health shock becomes large enough, individuals switch to public healthcare. Consequently, private-medical spending is observed up to the point where individuals switch to public healthcare. The switching point depends on tastes for conveniences, and on how much agents value healthcare. If we have at least two different copays that are strictly positive, then from the two switching points, both parameters can be identified. Finally, we obtain the risk aversion parameter from the choice of insurance coverage. It is manifest that having a zero copay and at least two positive copays is a key feature of our data that enables us to identify our model.

This chapter contributes to several strands of literature. Our work adds to the

literature that disentangles adverse selection and moral hazard. While a vast body of empirical literature focuses on assessing the presence of asymmetric information in health-insurance markets, most earlier studies examine the relationship between claim rates and insurance coverage, a test that Chiappori and Salanié (2000) formally analyze. Cutler and Zeckhauser (2000) review these papers and find that adverse selection is quantitatively large. More recent papers move to the empirical challenge of disentangling selection and moral hazard and estimate the welfare implications of asymmetric information. The evidence on asymmetric information in recent papers is mixed. Some studies find no evidence of informational asymmetries (e.g., Cardon and Hendel 2001), others find significant moral hazard but no adverse selection (e.g., Keane and Stavrunova 2016), and others report substantial evidence of adverse selection and moral hazard (e.g., Einav et al. 2013; Bajari et al. 2014; Powell and Goldman 2016). More consistent is the conclusion that, even with substantial asymmetric information, welfare losses are small. Our chapter contributes to this literature by leveraging changes in copays and deductibles implemented by the insurance company to disentangle moral hazard from adverse selection.

Previous studies analyze selection in insurance markets based on risk attitudes (Cohen and Einav, 2007b), cognitive ability (Fang et al., 2008), health risk (Handel, 2013b), or anticipated medical utilization (Einav et al., 2013). Our study contributes to this literature by examining the role of selection based on taste for convenience. Taste for convenience is a dimension rarely considered in the literature, one exception being Polyakova (2016b) that studies the German Healthcare system and finds no evidence of selection based on risk attitudes but significant non-pecuniary taste heterogeneity.

Another important contribution of our study is the analysis of public and private interaction in health-insurance markets. Most existing papers focus on the US Medicare or Medicaid programs (e.g., Polyakova 2016a; Marton et al. 2017). However, US public health insurance is a subsidy system for which a considerable fraction of the population is ineligible. Aside from the US healthcare system, several papers analyze the German healthcare system, where public healthcare insurance is mandatory for the majority of the population, and private healthcare insurance is available for a selected group of people (e.g., Grunow and Nuscheler 2014; Polyakova 2016b). In our chapter, we analyze an institutional setting where the government not only subsidizes healthcare but also actively provides health services to everyone. This institutional setting is similar to the UK healthcare system, studied in Olivella and Vera-Hernández (2013), who use data from the British Household Panel Survey and compares the use of healthcare services among people who receive private insurance from their employer as a fringe benefit and those who buy it directly, finding evidence of informational asymmetries in private health insurance. They argue that their results are a result of adverse selection. In contrast, our setting allows us to disentangle adverse selection and moral hazard, finding that besides adverse selection, moral hazard plays an important role.

2.2 Institutional Setting

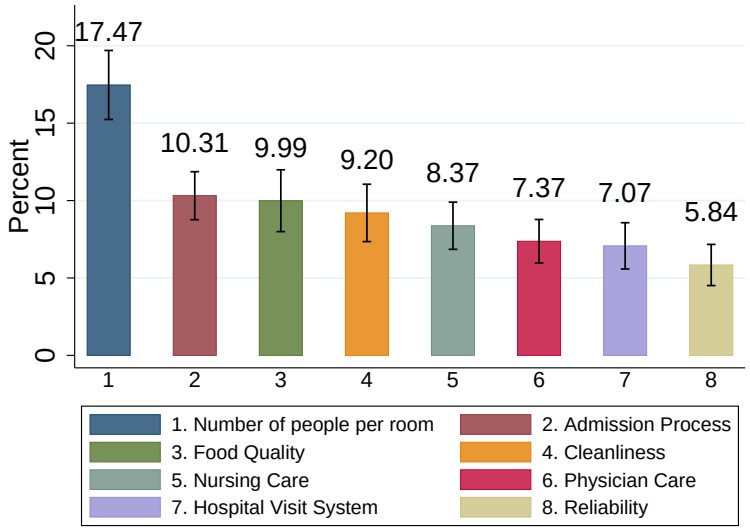
Our analysis focuses on a middle-income country that offers universal healthcare through a government-run system. The public healthcare system provides access to a broad range of services, including preventive care, medications, outpatient consultations, surgeries, medical tests, and highly complex procedures such as organ transplants. The public healthcare system is financed mainly via mandatory

payroll deductions, and users do not pay copays when using any of its services.

Despite its virtues, the public healthcare system has a series of inconveniences, particularly in the form of long waiting times. According to reports published by a unit that monitors the problem, on average, a person waits 442 days for elective surgery, 198 days for a diagnostic test, and 186 days for medical consultation. In general, cases are prioritized and the longest waiting times correspond to non-life-threating conditions. Aside from long waiting times, the public healthcare system has other types of inconveniences. In a national health survey conducted in the time window 2005-2010, respondents were asked to provide their level of satisfaction with certain features of public hospitals. Figure 2.1 displays the percentage of people who rated each feature as "bad" or "very bad." The number of people per room ranks in the first place, followed by the admission process and food quality. Notice that the features that generate more dissatisfaction with the public healthcare system correspond to discretionary services, not directly related to the provision of essential health services.

The non-monetary costs associated with the public healthcare system have given rise to a private healthcare system. In the private healthcare system, patients can have access to medical attention after paying for its services. Private healthcare reduces most of the waiting cost and provides discretionary services. For example, according to the national health survey, on the appointment day, users wait on average 55.13 minutes before receiving medical consultation in the public healthcare system, while in the private healthcare system the average waiting time is 34.44 minutes (a difference significant at the 1% level). Getting medications is also slower in the public healthcare system. While patients in private healthcare receive their medication immediately, in public healthcare they wait 161.02

Figure 2.1: Public hospital's dissatisfaction.



Source: Authors' calculations based on a national health survey conducted between 2005 to 2010. Dissatisfaction is measured by the percentage of people who rated the feature as "bad" or "very bad." Error bars show 95% confidence interval.

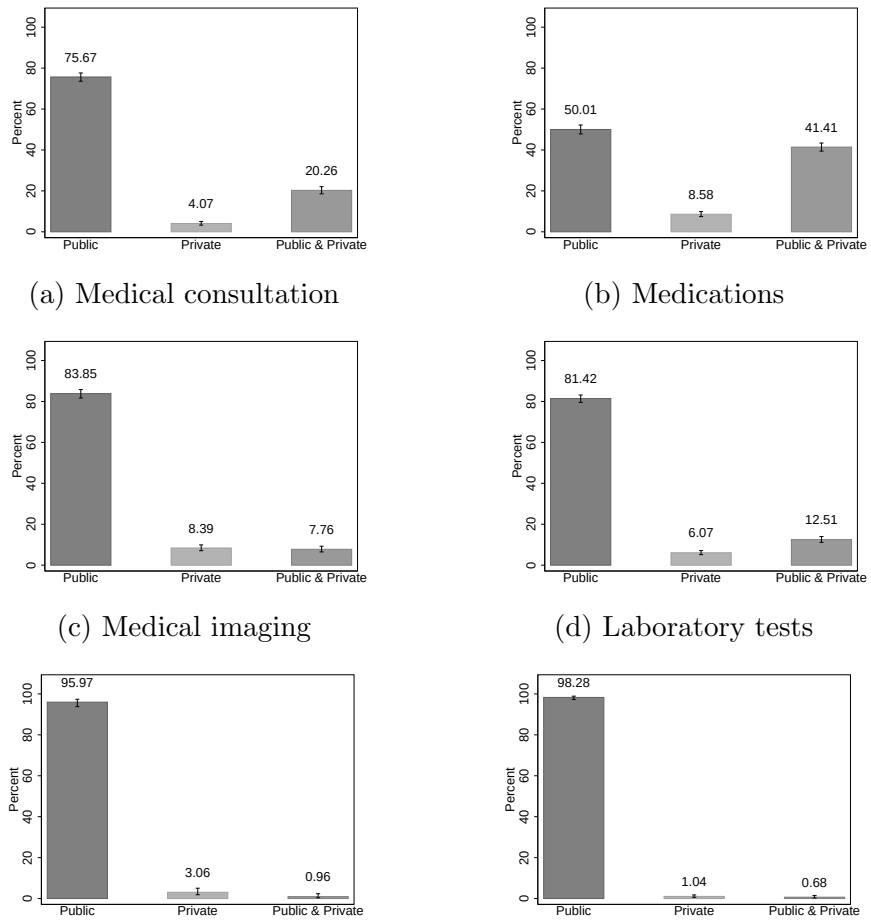
minutes on average (s.e 15.08).

Aside from the difference in monetary and non-monetary costs, the quality of both systems is similar. The public healthcare system is well-equipped to perform highly complex procedures. For instance, the public healthcare system is responsible for introducing lung and intestine transplant in the country. Although the public healthcare system is the largest employer in the country's health sector, employing nearly 74% of all physicians, it is common for physicians to have dual practice, as approximately 75% of its physicians also work in the private system. Finally, Figure 2.1 shows that respondents perceive physician care and reliability among the best-rated features of public hospitals.

According to a household income and expenditure survey conducted in the time window 2010-2015, households use private healthcare for relatively low-cost procedures and switch to public healthcare for the treatment of serious health conditions.

Figure 2.2 shows that conditional on using any healthcare system, households use the private healthcare system mostly for medications, medical consultations, medical imaging, and laboratory tests. On the other hand, more than 95% of households use public healthcare for hospitalizations and complex treatments.[2]

Figure 2.2: Conditional on receiving medical services, where households obtain it.



(a) Medical consultation (b) Medications

(c) Medical imaging (d) Laboratory tests

(e) Hospitalization (f) Complex treatments

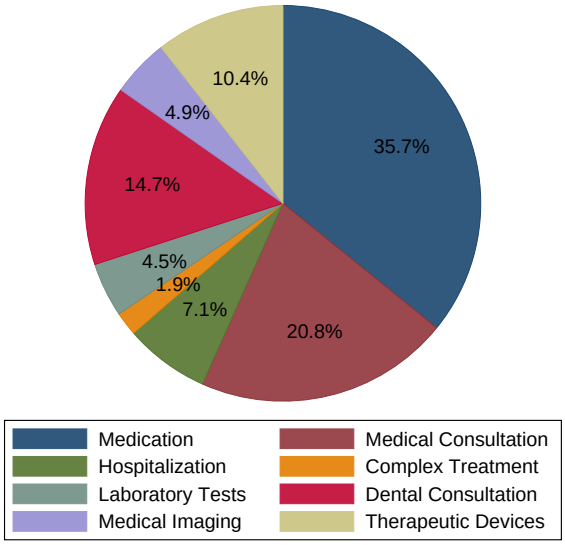
Source: Authors' calculations based on a household income and expenditure survey conducted between 2010 to 2015. Error bars show 95% confidence interval.

On average, health spending in the private healthcare system represents 2.87% of total household expenditures (median 0.92%). Figure 2.3 depicts the distribu-

[2]Complex treatments include inpatient and outpatient surgery, chemotherapy and radiotherapy.

tion of total household spending in private healthcare. About 71.18% of private health expenses correspond to medical and dental consultations and medications. On the other hand, despite being relatively more expensive procedures, hospitalizations and complex treatments account for about 8.97% of total private health expenditure. Consistent with Figure 2.2, individuals use private healthcare mostly for convenience, but for relatively large health shocks they switch to public health care where they can get relatively good quality at a lower cost. The existence of the public healthcare system to treat expensive health conditions might explain the relatively low incidence of catastrophic health spending, measured as the percentage of households with health spending exceeding 10% of the household's expenditure. In the country, approximately 6.41% of households are affected by catastrophic health spending, in contrast to a global incidence of 11.7% (Wagstaff et al., 2018).

Figure 2.3: Distribution of total household spending in private healthcare.



Source: Authors' calculations based on a household income and expenditure survey conducted between 2010 to 2015.

2.3 Data

The data come from an insurance company that offers health-insurance coverage that can be used in the private healthcare system. We focus on an insurance product that provides coverage to students. For this coverage, a student is a person of any age enrolled in an educational institution or daycare center recognized by governmental authorities. We refer to the product as "Student Health Insurance." Student Health Insurance pays for medical expenses due to an accident, permanent incapacity, and funeral expenses, up to a maximum insured amount chosen by the individual from a menu of options.

The choice set of Student Health Insurance includes four insured amount options. Each option j ($j = 1, 2, 3, 4$) in a menu corresponds to a maximum insured amount M_j in case of an adverse event, where $M_1 < M_2 < M_3 < M_4$. The premium p_j for option j does not vary across individuals. Student Health Insurance coverage is annual, and, after expiration, individuals must actively renew the policy. Each time, insurance agents present to the individuals the menu of options and the new conditions, such as any changes in premiums, copays or deductibles.

The insurance company changed the Student Health Insurance menu five times during the years that span our data, with changes to premiums, copays, and deductibles. However, the options for maximum insured amounts did not change over time. To ensure comparison of monetary amounts across the years, we convert the quantities to real amounts using a medical care price index.[3] Moreover, to protect the identity of the company, we indicate each period where a menu was available by t_i, and through the chapter, we normalize all monetary amounts such

[3]For nominal premiums and nominal insured amounts, we use the price index corresponding to the first month when each change took effect.

that the real premium for Option 1 in period t_1 is 1. Given this normalization, Table 2.1 presents the available choice menu for each period (again, in any period, the premiums are the same for all households).

Table 2.1: Student Health Insurance-Menu

Period	**Option 1** M_1 214.13	**Option 2** M_2 428.27	**Option 3** M_3 856.53	**Option 4** M_4 1284.80
t_1	1	1.88	3.68	5.47
t_2	1.82	2.22	3.92	5.82
t_3	1.97	2.57	4.94	6.52
t_4	1.91	3.24	5.53	7.24
t_5	3.09	4.36	6.90	7.26

Notes: For Option j ($j = 1, 2, 3, 4$), M_j represents the maximum insured amount in case of an adverse event. Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1.

Besides changes in premiums, the company also implemented changes in copays and deductibles. Table 2.2 summarizes the four different regimes. In periods t_1 to t_4, the copay and deductible were the same regardless of the option chosen. In period t_1, there was a copay of 5% and a zero deductible. In periods t_2 and t_3, there was no copay and no deductible. In period t_4, there was a 15% copay and a deductible of 3.81. Finally, in period t_5, there was no deductible, and the copay depended on the maximum insured amount that was chosen.

Table 2.2: Student Health Insurance-Copay-and-Deductible Regimes

Regime	Period	Deductible	Copay (%)
1	t_1	0	5
2	t_2, t_3	0	0
3	t_4	3.81	15
4	t_5	0	25 for Option 1 20 for Option 2 15 for Options 3 & 4

Notes: Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1.

The insurance company offers Student Health Insurance in two forms. The first

form is "individual insurance," and it corresponds to the standard insurance where people go to a branch and buy a policy. The second form is "collective insurance," where an educational institution signs a contract with the company to insure all of its students. Both the individual and collective insurance have the same rules and had the same copay-and-deductible regimes at the same time. A key feature of collective insurance is that 67% of the educational institutions assign the same insured amount to all their students, allowing us to shut down the adverse selection channel.

One might be concerned that in collective insurance parents are choosing educational institutions on the basis of the insurance provided by the institution, and consequently, this introduces a source of endogeneity in the insured amounts assigned to the students. However, this is unlikely to be the case. Individuals in educational institutions where parents pay tuition (private and subsidized education) account for nearly 75.3% of our collective-insurance sample. According to the household income and expenditure survey, the annual average tuition in educational institutions is 788.72 (median 448.57). Individual Student Health Insurance is available for any student under the same terms and conditions as the collective version, and in this case, the annual premium for the most comprehensive coverage is less than 10. Therefore, if parents only care about the insurance coverage, they could have the highest insured amount at lower cost using the individual modality instead of paying tuition. In general, access to health insurance does not constitute a determinant for choosing private education, surpassed by quality education, infrastructure, better technological resources, or more intensive teaching of a foreign language.

For each policy in the time window, the data record the issue date and the

insured amount chosen. Moreover, the data contain for each policyholder their date of birth, gender, and place of residence. The data also include the claims associated with each policy and specify the event that happened to the policyholder. We construct panel data for each individual based on their renewal times.

The full dataset includes more than 80,000 individuals over six consecutive years during the time window between 2010 and 2018. However, because the insurance covers students, most policyholders in the full dataset are relatively young (91.05%<18 years). For them, it is likely that a third party, such as their parents, makes the insurance and medical utilization decisions. Therefore, to avoid mixing populations with different decision environments, we drop all individuals older than 24 years old. Hence, the analysis perhaps is best interpreted as reflecting preferences of parents for their children.

The data also record the claim history for each policy and describe the reason. The occurrence of permanent incapacity and funeral expenses in our dataset is extremely low: 11 deaths and two incapacities. Therefore we ignore these possibilities and assume that, effectively, the coverage is exclusively for medical expenses due to an accident. Using claim histories, we can construct total medical spending for each policyholder during each policy term.

Moreover, for three out of the six years that we analyze, the company also provided line-item claim data. For each claim, the data includes diagnostic, procedures, and spending. The line-item claim data encompasses 75% of the claims in the individual sample and 52% in the collective sample.

After dropping individuals older than 24 years old, the total number of policyholders in our individual sample is 32,187 (53,924 individual-years). In the

collective sample, we also drop all individuals older than 24 years old, and we keep only institutions that do not allow for different insurance coverage. In the end, the collective-insurance sample consists of 27,019 policyholders (49,876 individual-years). Besides the information included in the individual data, the collective-insurance data record the policyholder's educational institution. We matched this information with educational institution information provided by the Ministry of Public Education on type of school, funding, location, and the number of students.

Table 2.3 provides descriptive statistics for the individual and collective samples, as of time of first purchase. All differences in observable characteristics between the samples are statistically significant at the 5% level. On average, the individual sample has older students, fewer females, and a higher percentage of students in primary school and college than the collective sample. To compare both samples and control for the differences in observable characteristics, in Section 2.4.3 we estimate a random-effect logit model pooling both samples.

Tables 2.4 and 2.5 summarize the distribution of choices in the individual sample and collective sample for each period respectively. The lowest-insured-amount option is more prevalent in the individual sample than in the collective sample. The difference in choices is because educational institutions are doing the choosing, not individuals.

About 98.19% of individual-year observations in the individual sample and 97.81% in the collective sample do not report any medical expenses. Figure 2.4 graphs the distribution of medical spending conditional on positive utilization. The average medical spending is 57.25 (median 33.98) in the individual sample, and in the collective sample is 67.49 (median 48.89).

Table 2.3: Student Health Insurance-Descriptive Statistics

	Individual Sample		Collective Sample	
Variable	Mean	SD	Mean	SD
Age (years)	12.245	4.47	11.737	4.69
Female (=1)	0.495		0.504	
National (=1)	0.981		0.955	
Kindergarten/Daycare (=1)	0.123		0.156	
Primary School (=1)	0.425		0.371	
High School (=1)	0.370		0.467	
Academic High School (=1)			0.304	
Technical High School (=1)			0.163	
College (=1)	0.081			
Others (=1)			0.006	
Public Education (=1)			0.247	
Private Education (=1)			0.541	
Subsidized Education (=1)			0.212	
Urban (=1)			0.934	
Average number of students			707.889	456.71
Period of first purchase				
t_1 (=1)	0.398		0.058	
t_2 (=1)	0.159		0.687	
t_3 (=1)	0.143		0.058	
t_4 (=1)	0.265		0.192	
t_5 (=1)	0.035		0.006	

Notes: Individual Student Health Insurance sample of 32,187 individuals and Collective sample of 27,019 individuals.

Table 2.4: Individual Student Health Insurance-Distribution of Choices

Period	**Option 1** M_1 214.13	**Option 2** M_2 428.27	**Option 3** M_3 856.53	**Option 4** M_4 1284.80	**N**
t_1	81.19%	15.13%	2.25%	1.42%	14,556
t_2	81.83%	14.48%	2.34%	1.35%	9,836
t_3	81.63%	14.27%	2.22%	1.89%	9,162
t_4	80.67%	14.51%	2.34%	2.48%	16,793
t_5	84.90%	10.32%	2.60%	2.18%	3,577

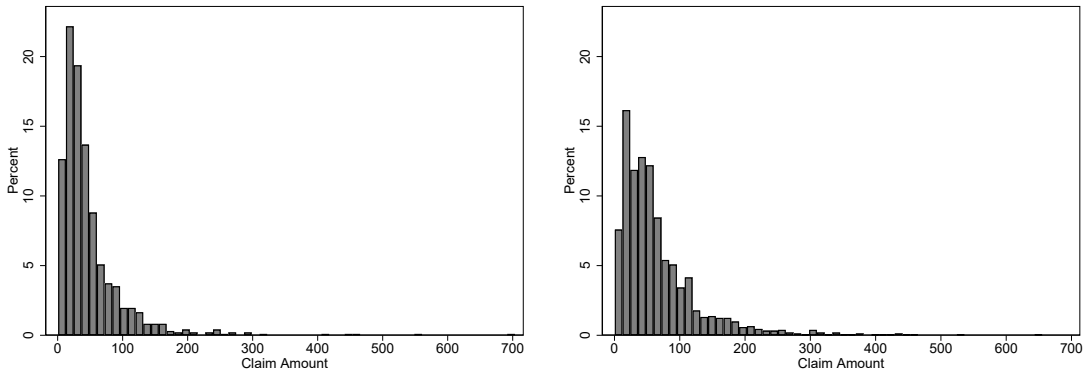
Notes: Individual Student Health Insurance sample of 32,187 individuals (53,924 individual-years). For Option j ($j = 1, 2, 3, 4$), M_j represents the maximum insured amount in case of an adverse event. Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1.

Table 2.5: Collective Student Health Insurance-Distribution of Choices

Period	**Option 1** M_1 214.13	**Option 2** M_2 428.27	**Option 3** M_3 856.53	**Option 4** M_4 1284.80	**N**
t_1	50.00%	50.00%			1,554
t_2	54.69%	41.72%	0.95%	2.64%	19,490
t_3	61.31%	33.91%	1.51%	3.27%	14,090
t_4	66.19%	30.27%	1.46%	2.08%	12,832
t_5	89.53%	1.36%	9.11%	0%	1,910

Notes: Collective sample of 27,019 individuals (49,876 individual-years). For Option j ($j = 1, 2, 3, 4$), M_j represents the maximum insured amount in case of an adverse event. Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1.

Figure 2.4: Distribution of medical expenditures, conditional on positive utilization.



(a) Individual Student Health Insurance (b) Collective Student Health Insurance

Notes: Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1. The x-axis is truncated at 700. The percentage of positive claims truncated in the individual-insurance sample is 0.92%, and in the collective-insurance sample is 0%.

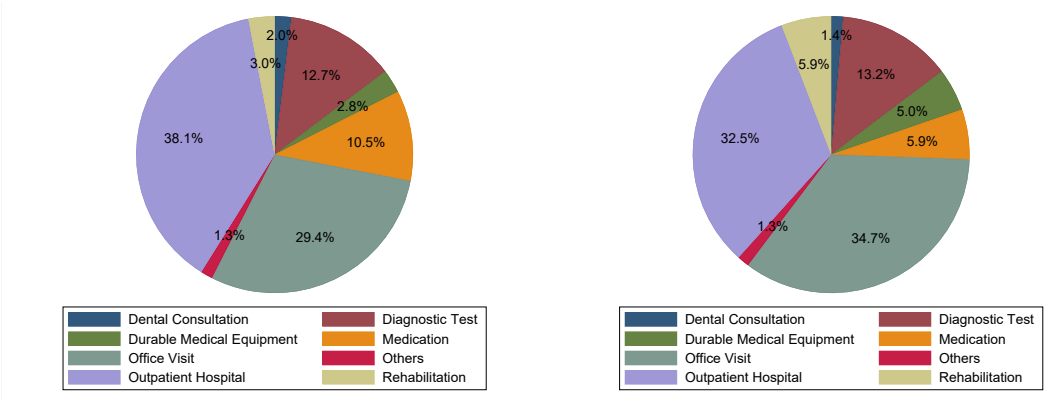
2.3.1 Revisiting the Evidence of Public versus Private Healthcare Utilization for the Student Health Insurance Sample

It is worth returning to the discussion in Section 2.2 to analyze the pattern of public versus private healthcare utilization for the Student Health Insurance policyholders. One way to approach the analysis is to use line-by-line health claims in our data. A second way to approach the analysis is to take advantage of the fact that the household income and expenditure survey specifically asks whether one has purchased Student Health Insurance. All in all, we find that Student Health Insurance customers exhibit a pattern of public versus private healthcare utilization similar to the pattern that we describe in Section 2.2, i.e., using private healthcare for relatively low-cost procedures and switch to public healthcare for serious illness.

The first approach is to use line-by-line health claims, including detailed information on procedures and spending. Figure 2.5 depicts the distribution of claims by healthcare services. Customers use Student Health Insurance mostly for outpatient hospital, office visits, diagnostic tests, and medications, representing 91% of claims in the individual-insurance sample and 86% of claims in the collective-insurance sample. Figure 2.6 presents the distribution of total claim spending. Again, outpatient hospital, office visits, diagnostic tests, and medications account for most of the spending. These categories represent approximately 79.8% of total spending in the individual-insurance sample and 85.1% of total spending in the collective-insurance sample.

The second approach is to use the household income and expenditure survey,

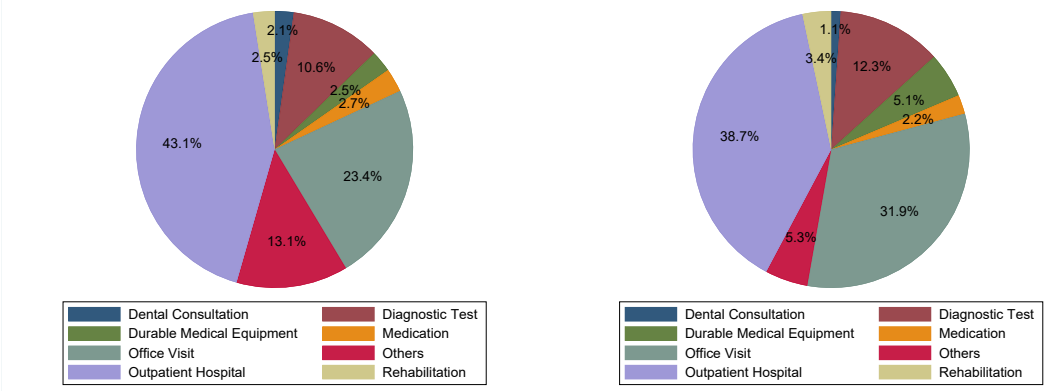
Figure 2.5: Student Health Insurance. Distribution of claims by healthcare services.



(a) Individual sample (b) Collective sample

Source: Authors' calculations based on line-by-line health claims data.

Figure 2.6: Student Health Insurance. Distribution of total claim spending.



(a) Individual sample (b) Collective sample

Source: Authors' calculations based on line-by-line health claims data.

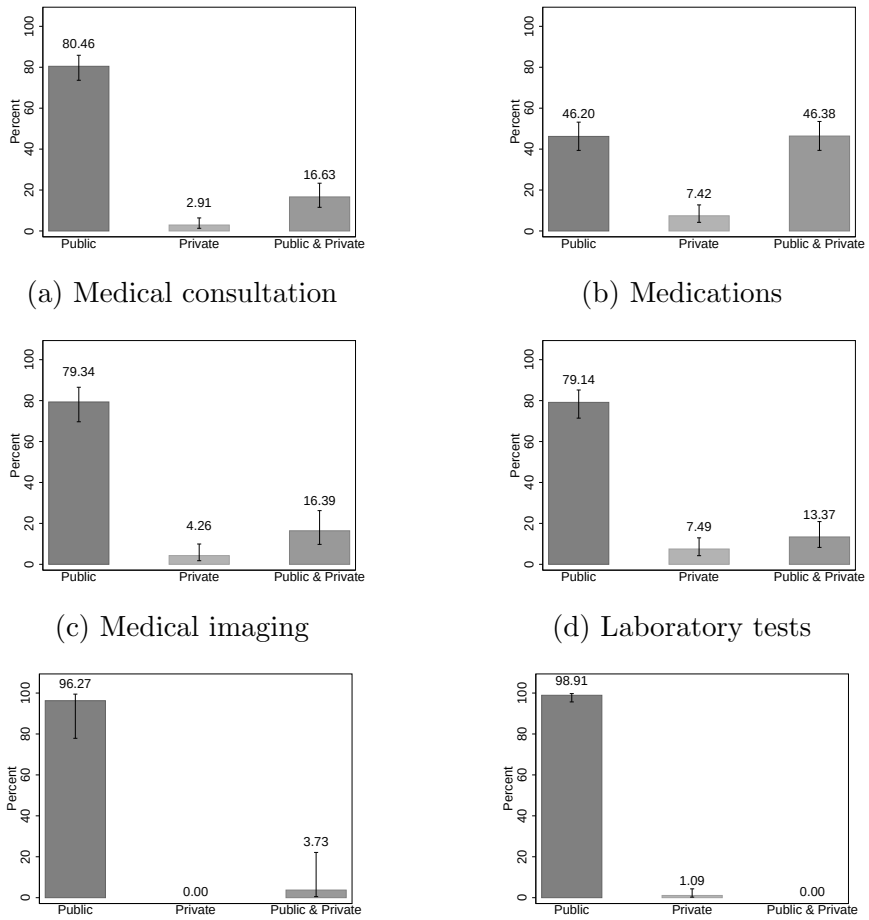
that contains information on insurance purchases, including Student Health Insurance purchases. Although it does not necessarily correspond to purchases from the same company we study, we can utilize this information to identify households that, as in our sample, have Student Health Insurance and have at least one member younger than 24 years old enrolled in an educational institution. Figures 2.7 and 2.8 replicate Figures 2.2 and 2.3 for the subset of households that report

having purchased Student Health Insurance. Much as before, conditional on using any healthcare system, households use the private healthcare system mostly for medications, medical consultations, medical imaging, and laboratory tests. For hospitalizations and complex treatments, more than 96% of households use public healthcare. Moreover, for households that bought Student Health Insurance, health spending in the private healthcare system represents on average 2.87% of total household expenditures (median 1.11%). We find that nearly 67.92% of the health expenses correspond to medical and dental consultations and medications, while more expensive procedures account for 1.88%.

2.4 Evidence of Selection Based on Asymmetric Information

In this section, we present reduced-form evidence of the presence of asymmetric information in Student Health Insurance. One of the most widely used approaches to identify asymmetric information is the positive correlation test (Chiappori and Salanié, 2000). The positive correlation test compares the claim rates of individuals with different levels of coverage and can be indicative of both adverse selection and moral hazard. Under the presence of adverse selection, high-risk individuals may self-select into higher coverage options. Consequently, claims and coverage should be positively correlated. Similarly, moral hazard can also produce the same positive correlation between claims and coverage when individuals who are identical before buying the insurance modify their behavior in response to different incentives offered by each plan, in particular, individuals with more coverage have more incentive to use the plan.

Figure 2.7: Student Health Insurance Households. Conditional on receiving medical services, where households obtain it.



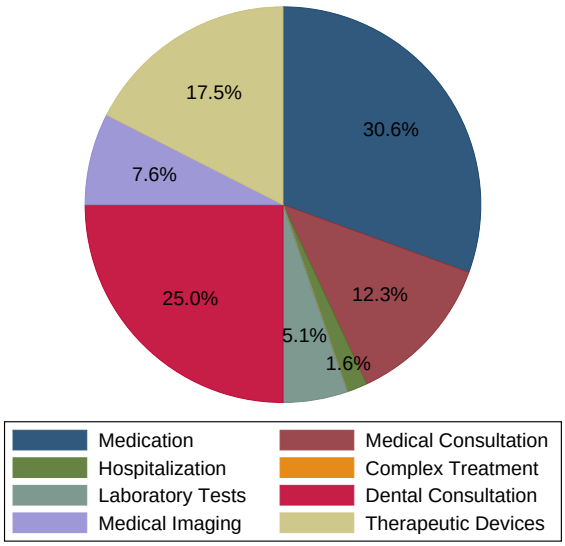
(a) Medical consultation (b) Medications

(c) Medical imaging (d) Laboratory tests

(e) Hospitalization (f) Complex treatments

Source: Authors' calculations based on a household income and expenditure survey conducted between 2010 to 2015. The sample is restricted to households that report having Student Health Insurance (not necessary form the same insurance company we analyze), and at least one member younger than 24 years-old enrolled in an educational institution. Error bars show 95% confidence interval.

Figure 2.8: Student Health Insurance Households. Distribution of total household spending in private healthcare.



Source: Authors' calculations based on a household income and expenditure survey conducted between 2010 to 2015. The sample is restricted to households that report having Student Health Insurance (not necessary form the same insurance company we analyze), and at least one member younger than 24 years-old enrolled in an educational institution.

Despite it being typically hard to disentangle moral hazard from adverse section using a positive correlation test, the existence of the individual and collective samples allow us to address this problem. First, consider the effect of insured amounts on claim rates. In the individual sample, any relationship between insured amounts and claim rates can arise either from adverse selection or moral hazard. On the other hand, in the collective insurance, the adverse selection channel is shut down; consequently, any effect purely reflects moral hazard. Therefore, after controlling for any difference in characteristics between the samples, any difference in the impact of insured amounts on claims rates is due to adverse selection, and any shared effect is due to moral hazard. Second, consider the effect of the copay-and-deductible regime on claim rates. Under the assumption that changes in copays and deductibles do not affect participation in the plan, the copay-and-deductible effect reflects moral hazard in both the individual and collective samples. In this

section, we proceed with the analysis following this reduced-form logic.

2.4.1 Individual Insurance

We begin the analysis with the individual health insurance sample. Table 2.6 shows a positive relationship between the insured amount chosen and claim rates. The group with the highest insured amount (Option 4) has a claim rate of 10.42%, about ten times higher than the group with the lowest insured amount (Option 1). The claim rates between groups are statistically different at the 1% significance level.

Not only do individuals with higher insured amounts have higher claim rates, but they also claim higher amounts. Table 2.6 presents evidence of a positive relationship between the insured amount and claim amounts. Conditional on positive utilization, the group with the highest insured amount (Option 4) has an average claim amount three times higher than the group with the lowest insured amount (Option 1). Except for Options 3 and 4, the difference in average claim amounts between groups is statistically significant at 1% level.

To study the correlation between insured amounts and claim rates after controlling for the copay-and-deductible regime, we restrict attention to regimes 1 to 3 where copays and deductibles are the same regardless of the insured amount (see Table 2.2). Rows A to C in Table 2.7 show that in each of these regimes the claim rates are increasing in the insured amount chosen. On average, the difference in claim rates between the highest and the lowest insured amount group is ten percentage points. In every regime, the differences between the lowest insured amount group (Option 1) and the highest insured amount group (Option 4) is statistically

Table 2.6: Individual Student Health Insurance-Claim Rates and Average Claim Amounts by Insurance Options

	Insured Amount			
	Option 1	**Option 2**	**Option 3**	**Option 4**
	M_1	M_2	M_3	M_4
	214.13	428.27	856.53	1284.80
Claim Rate	1.09%	3.86%	7.22%	10.42%
N	43,930	7,739	1,247	1,008
Average Claim	41.15	50.43	90.34	121.89
	(35.08)	(55.02)	(140.36)	(238.51)
N	480	299	90	105

Notes: Individual Student Health Insurance sample of 32,187 individuals (53,924 individual-years). For Option j ($j = 1, 2, 3, 4$), M_j represents the maximum insured amount in case of an adverse event. Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1. Standard deviation in parentheses. The average claim amount is conditional on positive utilization.

different at 1% level.

Rows A to C in Table 2.7 also show that in each of these regimes the average claim amount is increasing in the level of insured amount chosen. Conditional on positive utilization, the average amount claimed by the highest insured amount group (Option 4) is approximately three times higher than the average amount claimed by the lowest insured amount group (Option 1). However, except for the zero-copay regime, the difference is not statistically significant at conventional levels. Taken together, the results in Table 2.6 and Rows A to C in Table 2.7 on the positive relationship between insured amounts and claim rates (claim amounts) are indicative of the presence of asymmetric information in the individual sample, and can be considered as a joint test of adverse selection and moral hazard.

Next, we compare how different copay-and-deductible regimes impact claim rates and average claim amounts. Table 2.8 shows an inverse relationship between copay and deductibles and claim rates. While the claim rate is 2.6% under zero

Table 2.7: Individual Student Health Insurance-Claim Rates and Average Claim Amounts by Insurance Options and Copay-and-Deductible Regime

	Regime		(1)	(2)	(3)	(4)
				Insured Amount		
			Option 1	**Option 2**	**Option 3**	**Option 4**
			M_1	M_2	M_3	M_4
			214.13	428.27	856.53	1284.80
(A)	0% copay	Claim Rate	1.66%	5.71%	8.55%	13.73%
		N	15,528	2,731	433	306
		Average Claim	45.64	52.57	97.83	183.10
			(37.21)	(55.17)	(143.48)	(297.72)
		N	258	156	37	42
(B)	5% Copay	Claim Rate	1.00%	2.77%	6.10%	10.63%
		N	11,818	2,203	328	207
		Average Claim	38.11	51.00	69.02	123.22
			(29.21)	(60.86)	(76.95)	(251.54)
		N	118	61	20	22
(C)	15% copay + 3.81 Deductible	Claim Rate	0.74%	3.24%	7.89%	9.59%
		N	13,547	2,436	393	417
		Average Claim	33.78	45.31	99.51	58.91
			(34.77)	(50.75)	(170.93)	(127.20)
		N	100	79	31	40

Notes: Individual Student Health Insurance sample of 32,187 individuals (53,924 individual-years). For Option j ($j = 1, 2, 3, 4$), M_j represents the maximum insured amount in case of an adverse event. Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1. Standard deviation in parentheses. The average claim amount is conditional on positive utilization.

copay, the claim rate drops to 0.13% for 25% copay (the highest copay observed in the dataset). Conditional on positive utilization, the average claim amount is more than two times higher in the zero-copay regime than in the 25% regime. All these differences are statistically significant at the 1% level.

It is possible to control for the influence of insured amounts. Columns (1) to (4) of Table 2.7 shows that for every insured amount an increase in copays from zero to a positive amount decreases claim rates. For Options 1 and 2, the reductions in claim rates are statistically significant at 1% level. Columns (1) to (4) of Table 2.7 also show that an increase in copays from zero to a positive amount reduces

the average claimed amount conditional on positive utilization. However, it is only statistical significant at the 5% level for Option 1.

We interpret the negative relationship between copays and deductibles and claim rates as reflecting moral hazard. However, it is important to highlight two caveats to this interpretation. The first caveat is that in regime 4, the copays depend on the insured amount chosen (see Table 2.2). Therefore, the claim rates for the 15%, 20%, and 25% copays can partly reflect adverse selection.

The second caveat is that the copay-and-deductible regime can affect insurance choices or participation in the plans. For example, lower copays can attract more risky individuals. However, this is unlikely to be the case. First, according to the evidence that Table B.1 in Appendix B.1 provides, the distribution of choices is constant across copay-and-deductible regimes, suggesting small selection arising because people move between insured amounts. Second, in Table B.2 in Appendix B.2, we compare people who bought insurance under different regimes using a random-effect logit model, and we also obtain a negative relationship between copays and deductibles and claim rates.

Table 2.8: Individual Student Health Insurance-Claim Rates and Average Claim Amounts by Copay-and-Deductible Regime

	Copay-and-Deductible Regime					
	25% copay	20% copay	15% copay	5% copay	0%	15% copay + 3.81 Deductible
Claim Rate	0.13%	0.81%	1.75%	1.52%	2.60%	1.49%
N	3,037	369	171	14,556	18,998	16,793
Average claim	25.90	62.03	28.87	52.94	63.46	49.60
	(20.85)	(43.76)	(13.60)	(93.01)	(109.95)	(88.16)
N	4	3	3	221	493	250

Notes: Individual Student Health Insurance sample of 32,187 individuals (53,924 individual-years). Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1. Standard deviation in parentheses. The average claim amount is conditional on positive utilization.

Tables 2.6 to 2.8 suggest the existence of asymmetric information in the Individual Student Health Insurance. To confirm that the results are not driven by covariates, we estimate a random-effect logit model. Our dependent variable is an indicator variable equal to one if the policyholder filed a claim during each policy term and equal to zero otherwise; and we assume that:

$$y_{it} = a_i + \mathbf{X}'_{it}\beta + \delta_t + \gamma_r + \epsilon_{it} \tag{2.1}$$

where a_i is a normally distributed individual-specific intercept with standard deviation σ_a, $\mathbf{X}_{it}$ is a vector of individual characteristics (age, gender, educational stage, and nationality), δ_t are time fixed effects (year and month), γ_r are region fixed effects, and ϵ_{it} is a logistically distributed error term.

The results of the random-effect logit model are reported in Table B.3 in Appendix B.3. Table 2.9 shows the corresponding average marginal effects. Beginning with a baseline set of individual characteristics as well as time and region fixed effects, in Column (1) we add to the estimation the plan choice, in Column (2) we add the copay-and-deductible regime, and in Column (3) we add both variables. Overall, the average marginal effect is increasing in the insured amount and decreasing with the copay-and-deductible regime. All marginal effects are significant at 5% level. Note that the marginal effects in Column (3) are similar to the marginal effects in Columns (1) and (2), suggesting little interaction between plan choice and copays and deductible. From Column (3), we conclude that an increase in the insured amount from the lowest to the highest option increases in 8.3 percentage points the probability of a claim on average, and an increase in copay from zero to 25% reduces the probability of a claim in two percentage points on average. Therefore, the results of the random-effect logit support the evidence

of asymmetric information found through our more simple comparisons.

Table 2.9: Individual Student Health Insurance-Average Marginal Effects-Random-Effect Logit Model

	Plan Choice (1)	Deductible Policy (2)	Plan Choice + Deductible Policy (3)
Option			
1	vs		vs
2	0.022***		0.022***
	(0.002)		(0.002)
3	0.052***		0.053***
	(0.007)		(0.007)
4	0.084***		0.083***
	(0.010)		(0.009)
Policy			
0% Copay		vs	vs
5% Copay		-0.008***	-0.008**
		(0.003)	(0.003)
15% Copay		-0.016**	-0.020***
		(0.006)	(0.003)
20% Copay		-0.018***	-0.019***
		(0.005)	(0.004)
25% Copay		-0.022***	-0.020***
		(0.002)	(0.003)
15% copay +		-0.008***	-0.008***
3.81 Deductible		(0.003)	(0.003)
N	52,537	52,537	52,537

Notes: All regressions include demographic controls (age, age^2, gender, nationality plus interactions) as well as year, month, and region fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, in the individual-insurance sample, we find a positive relationship between insured amounts and claim rates (claim amounts). This insured-amount effect can be indicative of either moral hazard or adverse selection. Also, we find evidence of a negative relationship between copay-and-deductible regimes and claim rates (claim amounts). In this case, because changes in deductible do not affect choices or participation in the plan, the copay-and-deductible effect arguably

reflects moral hazard.

2.4.2 Collective Insurance

Table 2.10 shows that, different from our findings using the individual sample, in the collective sample we do not find a monotonic relationship between the insured amount chosen and claim rates. The group assigned to Option 1 (Option 2) do not have any significant difference in claim rates with the group assigned to Option 3 (Option 4). The positive correlation between claim amounts and insured amounts also disappears in the collective insurance. Except for Option 1, the groups do not show significant differences in average claim amounts conditional on positive utilization. Given that the students receive the insured amount exogenously, arguably the weakening of the positive correlation is a result of shutting down the adverse-selection channel.

Table 2.10: Collective Student Health Insurance-Claim Rates and Average Claim Amounts by Insurance Options

	Insured Amount			
	Option 1	**Option 2**	**Option 3**	**Option 4**
	M_1	M_2	M_3	M_4
	214.13	428.27	856.53	1284.80
Claim Rate	2.19%	4.48%	1.84%	4.10%
N	30,278	17,596	759	1,243
Average Claim	56.68	74.16	118.46	90.58
	(45.49)	(73.57)	(120.73)	(100.03)
N	662	789	14	51

Notes: Collective Student Health Insurance sample of 27,019 individuals (49,876 individual-years). For Option j ($j = 1, 2, 3, 4$), M_j represents the maximum insured amount in case of an adverse event. Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1. Standard deviation in parentheses. The average claim amount is conditional on positive utilization.

We control for the copay-and-deductible regime in Rows A to C of Table 2.11, finding that except for the 5% copay, in each regime there is not a clear pattern between claim rates and insured amounts. The average claim amount conditional on positive utilization neither displays a positive correlation with the insured amount. Different from our results in individual insurance where a positive relationship emerged for each regime, in collective insurance having higher insured amount does not translate into higher claim rates or higher average claim amount.

Table 2.11: Collective Student Health Insurance-Claim Rates and Average Claim Amounts by Insurance Options and Copay-and-Deductible Regime

	Regime		(1)	(2)	(3)	(4)
			Insured Amount			
			Option 1	**Option 2**	**Option 3**	**Option 4**
			M_1	M_2	M_3	M_4
			214.13	428.27	856.53	1284.80
(A)	0% copay	Claim Rate	2.43%	4.69%	2.76%	4.30%
		N	19,297	12,909	398	976
		Average Claim	60.72	76.82	140.84	95.86
			(46.53)	(71.94)	(127.54)	(108.29)
		N	468	606	11	42
(B)	5% Copay	Claim Rate	0.77%	2.96%		
		N	777	777		
		Average Claim	22.67	151.83		
			(11.01)	(129.17)		
		N	6	23		
(C)	15% copay + 3.81 Deductible	Claim Rate	2.14%	4.07%	1.60%	3.37%
		N	8,494	3,884	187	267
		Average Claim	48.53	53.16	36.39	65.93
			(42.12)	(59.17)	(23.43)	(40.10)
		N	182	158	3	9

Notes: Collective Student Health Insurance sample of 27,019 individuals (49,876 individual-years). For Option j ($j = 1, 2, 3, 4$), M_j represents the maximum insured amount in case of an adverse event. Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1. Standard deviation in parentheses. The average claim amount is conditional on positive utilization.

Table 2.12 shows how different copay-and-deductible regimes impact claim rates and average claim amounts. We do not consider 15% and 20% copays in the analysis because we have few observations (26 and 174 respectively). Similar to the

results using individual insurance, an inverse relationship between copay and deductibles and claim rates emerges. While the claim rate under a zero-copay regime is 3.36%, the claim rate drops to 0.35% for 25% copay. Moreover, conditional on positive utilization, the average claim amount is approximately three times higher in the zero-copay regime than in the 25% regime. All these differences are significant at the 1% level.

In Columns (1) to (4) of Table 2.11 we compare regimes conditional on the insured amount, finding that the negative relation holds for each insured amount. Except for the 5% copay in Option 2, Columns (1) to (4) of Table 2.11 also suggests that increasing copays from zero to a positive percentage reduces the average claim amount conditional on positive utilization in every insured amount.

Table 2.12: Collective Student Health Insurance-Claim Rates and Average Claim Amounts by Copay-and-Deductible Regime

	Copay-and-Deductible Regime			
	25% copay	5% copay	0%	15% copay + 3.81 Deductible
Claim Rate	0.35%	1.87%	3.36%	2.74%
N	1,710	1,554	33,580	12,832
Average claim	23.59	125.11	71.47	50.95
	(19.30)	(126.35)	(66.20)	(50.33)
N	6	29	1,127	352

Notes: Collective Student Health Insurance sample of 27,019 individuals (49,876 individual-years). Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1. Standard deviation in parentheses. The average claim amount is conditional on positive utilization.

As in our analysis of the individual insurance, we estimate equation (2.1) using a random-effect logit model. The results of the random-effect logit model are reported in Table B.4 in Appendix B.4. Table 2.13 shows the corresponding average marginal effects. The results do not show any increasing relationship between claim

rates and insured amount. In particular, the average marginal effect for Options 2 and 4 is not statistically different. On the other hand, except for the 5% copay, the copay-and-deductible regimes are significant at the 1% level and suggests that the average marginal effect is decreasing with the copay-and-deductible regime. Similar to the individual-insurance case, the marginal effects in Column (3) are roughly the same as the marginal effects in Columns (1) and (2), suggesting little interaction between plan choice and copays-and-deductible regimes. From Column (3), we conclude that an increase in copay from zero to 25% reduces the probability of a claim in three percentage points on average.

Table 2.13: Collective Student Health Insurance-Average Marginal Effects-Random-Effect Logit Model

	Plan Choice (1)	Deductible Policy (2)	Plan Choice + Deductible Policy (3)
Option			
1	vs		vs
2	0.016***		0.013***
	(0.003)		(0.003)
3	-0.007		-0.008
	(0.006)		(0.006)
4	0.025**		0.019*
	(0.011)		(0.010)
Policy			
0 Copay		vs	vs
5% Copay		0.006	0.004
		(0.008)	(0.007)
25% Copay		-0.027***	-0.026***
		(0.004)	(0.004)
15% copay +		-0.010***	-0.008***
3.81 Deductible		(0.002)	(0.002)
N	46,902	46,902	46,902

Notes: All regressions include demographic controls (age, age^2, gender, nationality plus interactions) as well as month and region fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, in the collective-insurance sample, we find no monotonic relationship

between insured amounts and claim rates (claim amounts), but we find evidence of a negative relationship between copay-and-deductible regimes and claim rates (claim amounts). Given the nature of the collective-insurance sample, arguably adverse selection is irrelevant, and all the effects are a result of moral hazard.

2.4.3 Comparison Individual and Collective Insurance

The copay-and-deductible effects in the individual and collective samples (Sections 2.4.1 and 2.4.2, respectively) suggest the presence of moral hazard. On the other hand, the insured-amount effect in the individual sample can be due to adverse selection or moral hazard. However, because the adverse selection channel is shut down in the collective sample, the lack of a positive correlation between insured amounts and claim rates in the collective sample suggests that the positive insured-amount effect in the individual sample is largely due to adverse selection.

To make a more explicit comparison of the samples, and to account for the differences in observable characteristics that Table 2.3 shows, we estimate a random-effect logit model pooling both samples. Our specification follows equation (2.1), but we also include an indicator variable equal to one if policy i at time t belongs to the individual insurance and zero otherwise, and also we include interactions between plan choice and copay-and-deductible regime and the individual-insurance indicator. The results of the random-effect logit model are reported in Table B.5 in Appendix B.5. Table 2.14 shows the corresponding average marginal effects. The main results of the analysis are unchanged, suggesting that the insured-amount effect in the individual sample is a product of adverse selection.

In the individual sample, the average marginal effect is increasing in the insured

amount. From Column (3) in Table 2.14 we conclude that an increase in insured amount from the lowest to the highest option increases in 12 percentage points the probability of a claim on average. On the other hand, in the collective sample, the pattern is not monotonic, and an increase in insured amount from the lowest to the highest option increases the probability of a claim 1.8 percentage points. The lower magnitude of the insured-amount effect in the collective sample suggests that adverse selection constitutes a higher share of the insured-amount effect in the individual sample.

The average marginal effect is decreasing with the copay-and-deductible regime in the individual and collective samples. An increase in copay from zero to 25% reduces the probability of a claim in four percentage points on average in the individual sample and two percentage points in the collective sample. Again, the copay-and-deductible effect indicates moral hazard.[4]

2.5 Model

Thus far we have provided reduced-form evidence for the existence of both adverse selection and moral hazard. In this section, we develop a model that incorporates four sources of heterogeneity that could give rise to adverse selection and moral hazard in this context: risk preferences, risk type, concern for health, and tastes for convenience. Risk aversion impacts adverse selection, while the other three sources impact both adverse selection and moral hazard. As we discuss in Section 2.2, the

[4]Recall from Section 2.4.1 that in regime 4 the copay and deductible depend on the insured amount chosen, and consequently, the copay-and-deductible effect for the 15%, 20%, and 25% copays could also be due to adverse selection. Tables B.6 and B.7 in Appendix B.6 show that the results are virtually identical if we use a restricted sample in which we only consider regimes 1 to 3.

Table 2.14: Comparison Individual and Collective Student Health Insurance-Average Marginal Effects-Random-Effect Logit Model

	Plan Choice (1)		Deductible Policy (2)		Plan Choice + Deductible Policy (3)	
	Individual	Collective	Individual	Collective	Individual	Collective
Option						
1	vs				vs	
2	0.034***	0.012***			0.034***	0.010***
	(0.003)	(0.002)			(0.003)	(0.002)
3	0.074***	-0.002			0.074***	-0.003
	(0.010)	(0.005)			(0.010)	(0.004)
4	0.114***	0.024***			0.120***	0.018***
	(0.012)	(0.008)			(0.013)	(0.007)
Policy						
0% Copay			vs		vs	
5% Copay			-0.016***	-0.002	-0.015***	-0.004
			(0.003)	(0.004)	(0.003)	(0.004)
25% Copay			-0.040***	-0.020***	-0.040***	-0.020***
			(0.003)	(0.002)	(0.003)	(0.002)
15% copay +			-0.021***	-0.007***	-0.022***	-0.005***
3.81 Deductible			(0.002)	(0.001)	(0.002)	(0.001)
N	101,567		101,567		101,567	

Notes: All regressions include demographic controls (age, age^2, gender, nationality plus interactions) as well as month and region fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

evidence suggests that tastes for convenience are a differentiating factor between the public and private sector in our healthcare market. Moreover, previous studies rarely emphasize selection based on tastes for convenience; therefore it constitutes a novel contribution of our model.

Consider a consumer deciding whether to purchase Student Health Insurance to cover against a potential monetized health shock, $s \geq 0$. While s is ex-ante unknown, we assume that the distribution of potential health shocks, $F(s)$, is known to the individual. Agents in our setting have access to public healthcare services that will provide treatment up to s at zero cost. While publically-provided healthcare entails zero financial costs, individuals incur hassle costs from seeking this treatment. Private firms also offer treatment with no hassle costs, but charge

an amount equal the health shock, s. If private treatment is sought, individuals choose the desired level of medical utilization, m.

Agents are offered the opportunity to purchase insurance policy $j \in \{1, ..., J\}$ for the premium p_j. Each policy reduces the cost of private treatment to $c_j(m)$ given medical utilization level m. We order j such that it is increasing in the level of coverage and zero denotes no insurance: $1 = c_0(m) < c_1(m) < ... < c_J(m) = 0$ and $0 = p_0 < p_1 < ....p_J$.

The decision problem of the agent can be modeled in two stages. In the first stage, conditional on her information on expected subsequent health realization, the consumer chooses the optimal private insurance policy. In the second stage, she observes a realized health state and decides her optimal level of medical utilization and whether to seek treatment in the public or private healthcare.

We now describe the consumer problem in detail. In the second stage, preferences associated with private insurance option j are assumed to be separable in health and money:

$$u_j(m; s, \beta) = h(m; s, \beta) + w_j(m) \tag{2.2}$$

where m denotes medical utilization and β indexes the degree to which the agent values health relative to money. The first term, $h(m; s, \beta)$, measures the benefit of medical expenditures. The second term, $w_j(m)$, is final wealth given expenditure m and policy j.

In the spirit of Ellis (1986) and Einav et al. (2013) we parametrize $h(m; s, \beta)$ and $w_j(m)$ as follows:

$$h(m; s, \beta) = -\frac{1}{\ln(\beta)} \left(m \left(1 - \ln \left(\frac{m}{s} \right) \right) - s \right), \quad \text{where } \beta \in (0, 1)$$

$$w_j(m) = w_0 - c_j(m) - p_j$$

where w_0 is initial wealth, p_j is the premium for policy j, and $c_j(m)$ is the out of pocket cost of expenditure m. For illustration, suppose that all options have a fixed copay such that for some $c_j \in [0, 1]$ the out-of-pocket cost is $c_j(m) = c_j m$. Then, the first order condition with respect to medical utilization is,

$$\frac{1}{\ln(\beta)} \ln \left(\frac{m}{s} \right) - c_j = 0 \tag{2.3}$$

and hence optimal utilization for policy j is $m_j^* = \beta^{c_j} s$. With no insurance ($c_j = 1$), agents will choose to consume medical services up to $m_j^* = \beta s < s$. With full insurance ($c_j = 0$) they will choose $m_j^* = s$. The difference between the full and no insurance, $(1 - \beta)s$, can be considered a measure of excess private healthcare consumption, or in the terminology of Einav et al. (2013), it measures ex-post moral hazard. Therefore, larger β, i.e. greater intrinsic taste for healthcare, implies lower moral hazard.

The indirect utility function of holding policy j and using private healthcare is given by:

$$u_j(m_j^*; s, \beta) = \frac{-(1 - \beta^{c_j})}{(-\ln(\beta))} s + w_0 - p_j.$$

In addition to private treatment, the consumer has access to public health healthcare. In this case, there is no choice of medical utilization; agents receive treatment equal to s at zero cost. However, although public health care is free of charge, individuals incur a fixed cost, $d(s, \phi)$, that represents hassle costs in the

form of longer waiting times or the absence of ancillary services. The function d is assumed to be concave in s. ϕ parameterizes individual distaste for these costs. For illustration, we impose the functional form $d(s, \phi) = \phi\sqrt{s}$, where $\phi > 0$. Using this specification of the hassle costs, the utility of public healthcare is,

$$u_j^{Public}(s) = -\phi\sqrt{s} + w_0 - p_j. \tag{2.4}$$

The consumer compares this utility to the indirect utility of receiving public treatment. In other words, private treatment is preferred if $u_j(m_j^*; s, \beta) > u_j^{Public}(s)$, which holds if,

$$s < \left(\frac{\phi(-\ln(\beta))}{-(1-\beta^{c_j})}\right)^2 \equiv \bar{s}_j(\beta, \phi). \tag{2.5}$$

Equation (2.5) defines the consumer's indifference point between public and private healthcare. Private treatment will be sought whenever the health shock, s, is below the indifference point, $\bar{s}_j(\beta, \phi)$. This captures the intuitive idea that individuals will seek more convenient, but less comprehensive, private care for relatively minor accidents, but will opt for more complete public care and incur the hassle costs for more serious injuries.

The indifference point between public and private healthcare depends on three factors. First, individuals are less likely to seek public care when the taste for convenience, ϕ, are large. Second, agents are more willing to be treated in the private sector as the marginal cost of utilization decreases. With full insurance, $c_j = 0$, private treatment will always be sought. With no insurance, $c_j = 1$, agents might still opt for private healthcare if the hassle cost of public services are sufficiently high or the taste for healthcare is sufficiently low. Finally, the larger

the disutility of the health shock, as indexed by β, the more likely agents are to seek the more comprehensive public treatment.

Private healthcare utilization for health shocks below the consumer's indifference point between public and private healthcare is m_j^*. Otherwise agents seek public treatment. Utility is therefore,

$$v_j^*(m; s, \beta, \phi) = \begin{cases} u_j(m_j^*; s, \beta) & \text{if} \quad s < \bar{s}_j(\beta, \phi) \\ u_j^{Public}(s) & \text{otherwise.} \end{cases} \tag{2.6}$$

To derive the optimal choice of coverage, we assume that individuals are expected utility maximizers. We further assume that utility takes the on the CARA parameterization, $u(x) = -\exp(-rx)$, where the argument x is assumed to be stage-two utility. The CARA parameterization allows us to abstract from initial wealth, w_0. Given F is known, the expected utility of option j is,

$$V_j(F, \beta, \phi, r) = -\int_0^\infty \exp(-rv_j^*(m_j^*; s, \beta))dF. \tag{2.7}$$

Optimal coverage is therefore given by,

$$j^* = \arg\max_j V_j(F, \beta, \phi, r). \tag{2.8}$$

In this model, the coefficient of risk aversion r capture risk preferences and can generate adverse selection. The parameter β represents the concern for health, ϕ the tastes for convenience, and the distribution of potential health shocks $F(s)$ captures risk type. The last three elements can either contribute to adverse selection or moral hazard.

2.5.1 Identification

In future work, we plan to structurally estimate the model. The parameters of the model will help us to analyze the relative importance of each dimension where selection is operating. Also, the estimates can shed light on the degree of heterogeneity in preferences. Finally, the structural model will allow us to consider the welfare implications of a market where the public and private sector interacts providing healthcare services.

We consider identification of the model parameters with ideal data. In this environment, ideal data consists of observing choice and utilization for the same individual for infinitely many periods for any $c_j \in [0, 1]$. First, note that for the full insurance case, $c_J = 0$, individuals will always seek private treatment up to s_i. Intuitively, both public and private care will offer s at zero financial cost, but public care entails hassle costs. As a result, observed private medical utilization in a given period will reveal the true health shock, s. As F_{0i} is fixed over time, observing the same individual's utilization for infinitely many periods is sufficient to reveal the distribution of health shocks, F_{0i}.[5]

With F_{0i} known, we next turn to identification of the preference parameters (β_{0i}, ϕ_{0i}). For some $c_j \in (0, 1)$ medical utilization is given by $m^*_{ij} = \beta_{0i}^{c_j} s_i$. Observing this utilization for infinitely many periods reveals the medical-utilization truncation point $m_{ij}^{\max}$, where

$$m_{ij}^{\max} = \beta_{0i}^{c_j} \bar{s}_j(\beta_{0i}, \phi_{0i}) = \frac{\beta_{0i}^{c_j}(\phi_{0i} \ln(\beta_{0i}))^2}{(1 - \beta_{0i}^{c_j})^2}. \tag{2.9}$$

For $j = 1, 2$, and without loss of generality, suppose that $0 < c_1 < c_2 < 1$. Note that m_{ij}^{max} is decreasing in c_j and hence $m_{i1}^{\max} > m_{i2}^{\max}$. Theorem 2.1 shows how we

[5]The subscript 0 denotes the true distribution.

can identify (β_{0i}, ϕ_{0i}) given (m_{ij}^{max}, c_j). Henceforth we omit the subscript i. The proof is in Appendix B.7

Theorem 2.1. *Suppose for any $0 < c_1 < c_2 < 1$ we observe maximum utilizations $m_1^{\max}$ and $m_2^{\max}$. Then the preference parameters (β_0, ϕ_0) are identified.*

Theorem 2.1 states that once the utilization cutoffs are observed for two distinct policies, those cutoffs can be used to infer β. Once β is known, ϕ can be recovered from equation (2.9). Note also that one implication of the model is that data on medical utilization for values below $m_{ij}^{\max}$ also contribute to identifying the parameter β because they take the form $m_j^* = \beta^{c_j} s$. Therefore, the estimates do not completely rely on identifying truncation points.

Conditional on F_i, β_i and ϕ_i, the coefficient of absolute risk aversion, r_i, is the only remaining influence on choice. In particular, the optimal choice of j is strictly decreasing in r_i. Observing the choice of j for various copays $c_j \in [0, 1]$ can therefore be used to recover r_i.

2.6 Conclusion

In this chapter, we use household-level data on health-insurance choices and medical utilization to analyze adverse selection and moral hazard in a healthcare market where, besides a private healthcare sector, the government provides universal healthcare. Although it is typically hard to disentangle adverse selection and moral hazard, two features of our setting allow us to address the problem. First, while the copay and deductible are fixed over the policy term, the company implemented various changes in copays and deductibles during most of the years that span our

data. Second, we have a collective version of the insurance where individuals exogenously received coverage. Our results suggest that asymmetric information, both in the form of adverse selection and moral hazard, are quite important in the private-health-insurance market. We also propose a model that embeds selection based on risk aversion, risk type, concern for health, and tastes for convenience provided in the private-healthcare sector. Our model highlights the role of selection based on tastes for convenience, a dimension that previous studies rarely consider.

The results in our chapter suggest that while both adverse selection and moral hazard are relevant, most of the impact of insured amounts on claims rates is the product of adverse selection. However, in our context, adverse selection can originate from any of the four sources of heterogeneity embedded in our model. One limitation of the reduced-form analysis we use throughout the chapter is that we cannot disentangle each source, a step that is necessary to derive welfare implications. For example, if adverse selection originates mostly from tastes for convenience, then it is not inefficient that the private-healthcare sector keeps most of the high types, because people who value the private services the most will utilize them. On the other hand, if adverse selection is a product of risk-type, then it indicates that the insurance company should adjust premiums based on individual characteristics (recall that all households pay the same premiums). The welfare analysis can also shed light on how the cost borne by the government providing universal healthcare affects the private-insurance market. For instance, optimal pricing can change if the fact that people switching to the public sector to receive more expensive treatments serve as a subsidy to the insurance companies, by reducing the cost of insuring customers. All this endeavor remains for future work.

CHAPTER 3

MULTINATIONALS AND DEVELOPMENT: EVIDENCE FROM THE UNITED FRUIT COMPANY IN COSTA RICA

(This chapter was written in collaboration with Diana Van Patten, UCLA)

"It happened once that someone at the table complained about the ruin into which the town had sunk when the banana company had abandoned it, and Aureliano contradicted him with maturity and with the vision of a grown person. His point of view, contrary to the general interpretation, was that Macondo had been a prosperous place and well on its way until it was disordered and corrupted and suppressed by the banana company..."

Gabriel García Márquez,

One Hundred Years of Solitude

3.1 Introduction

Multinational companies have been a major driver for the spread of global capitalism. As multinationals grew in number and scale, they began evoking mixed feelings among the general public and policymakers, becoming heroes or villains of the globalized economy. The attitudes towards multinational companies and the debate it triggers have direct repercussions on the welfare of millions of people through the willingness to support protectionist policies and anti-globalization agendas. However, finding a domain in which one can clearly identify the impact of a multinational has proven to be a challenge. The first difficulty is that most

multinationals do not have a well-defined area of influence. The second challenge is that multinationals do not choose their location randomly, and especially in the case of agricultural companies, they want to exploit the more appropriate land for its interests. We overcome both challenges by studying an agricultural multinational with a well-defined and extensively documented boundary, which was the product of a quasi-random allocation of land.

In particular, we study the case of the United Fruit Company (UFCo) in Costa Rica. Founded in 1899, the UFCo was engaged in the cultivation and commercialization of tropical fruits, primarily bananas. By 1930 the UFCo had become one of the largest multinational corporations in the Western Hemisphere: it owned 3 million acres in Latin America (roughly the size of Connecticut) and controlled approximately 80% of the global banana production. While some commentators have admired the UFCo for transforming jungles into centers of human activity with a well-organized plantation economy,[1] most have accused the UFCo of being responsible for creating "Banana Republics": politically unstable countries dependent on a single exportation product produced by foreign enterprises.[2]

To identify the causal effect of the UFCo on its host country, we use a regression discontinuity (RD) approach. The approach compares units located in a close distance but on different sides of the UFCo boundary where geographic characteristics are balanced. Our setting is an ideal case study as there is historical evidence

[1]Palmer (1932, p. 271) wrote that "the opening of the nineteenth century was the dawn of a new era for the countries of the Caribbean. Until that time, these countries had been sparsely populated, and the lowlands had been unoccupied except by poison snakes, ferocious animals, myriads of insects, and dreaded diseases. [...] The scene was changing. [...] American engineers are invading the jungles with steam shovels. Swamps are being drained and axes are heard ringing in the woodland. Fruitful banana plantations are appearing as if by magic."

[2]Kepner and Soothill (1935, p. 336) denounced that the United Fruit Company "has throttled competitors, dominated governments, manacled railroads, ruined planters, choked cooperatives, domineered over workers, fought organized workers, and exploited consumers. Such usage of power by a corporation of a strongly industrialized nation in relatively weak foreign countries constitute a variety of economic imperialism."

of a quasi-random assignment of some portion of the land allocated to the UFCo. In 1904 the government issued a decree to clarify differences in the interpretation of previous land concession granted to the company. The decree announced some wastelands as state property, and specified that the property rights over these lands should be sold only to nationals and restricted its use to agricultural colonization. Because at that time the Atlantic Coast of Costa Rica was hardly explored, the drawing of the boundaries was arbitrary and based on salient features of the environment. For example, the decree stipulates among the limits *"an imaginary line from the intersection between Toro Amarillo river with the old railroad up to a point in the coast located five miles northeast from the mouth of Tortuguero river."* (ANCR, 1904, p. 44)

Together with this quasi-random design of the boundary, the availability of disaggregated data is a key aspect of our identification strategy. We use restricted access census micro-data available for 1973, 1984, 2000, and 2011, geo-referenced at the census block level. The census block is the smallest territorial division of the country, and its level of disaggregation allows us to compare households in a close distance, including values as small as 1 kilometer (km). Since the multinational operated in the region of our study from 1899 to 1984, our dataset starts at a time when the company was functioning and spans up to three decades after it stopped production. We, therefore, explore both the contemporaneous effect of the UFCo's presence, as well as its short-, medium- and long-run effects on the region's economic development.

We find that the UFCo had a positive and long-lasting impact on the regions where it operated. Households living in former UFCo areas are less likely of having an unsatisfied basic need in any of the dimensions we consider (housing, sanitation,

education, and consumption capacity), and in general, are 12 percentage points less likely of being poor than households living outside. When we disentangle the dynamic of the UFCo effect, we find that it is persistent over time, but it shows some evidence of convergence. For example, the difference in the severity of poverty in the two regions has decreased. While in 1973 households within the UFCo plantations have 0.67 less unsatisfied basic needs than households outside, in 2011 the difference drops to 0.13. Further, collecting data from primary sources, we are able to speak to plausible mechanisms behind our findings. Before the company's arrival, the region was isolated and sparsely populated, and transforming these areas into plantations required a series of investments. We identify investments in physical and human capital carried out by the UFCo, such as hospitals, housing provision, roads, railroads, sanitary programs, and schools, as probable relevant mechanisms behind our results. That said, it is worth to highlight that we focus on the economic outcomes of regions nearby the company landholdings. Our analysis is not able to say whether the company had a positive impact on every aspect of life, or for the country as a whole.

A potential concern regarding our counterfactual is that our results are driven by lower levels of government spending in the vicinity of the company, i.e., that the UFCo crowded out public investment in nearby regions. Using primary public investment data we collected from the Comptroller General of the Republic of Costa Rica (*Contraloría General de la República de Costa Rica*), we rule out this hypothesis by showing that government spending per capita was not different in UFCo municipalities.

As a falsification test, we draw placebo borders by moving the real border 2 km both inwards and outwards and re-run our analysis. Because with placebo borders

we are comparing households exposed to the same treatment, we should not expect to find any effect. Consistent with this hypothesis, all the estimates using placebo borders are economically and statistically insignificant, suggesting that our results capture a specific UFCo effect.

Our results are robust to different specifications of the RD polynomial, and to varying control variables. Moreover, exploiting the disaggregation in our data at the census block level, we compare households within a distance of one kilometer from the UFCo boundary, obtaining similar results. We also verify that migration is not driving our results by obtaining the same estimates after dropping all households with any member who previously had a different place of residence; and by showing that the UFCo effect is the same between households engaged in the agricultural sector (for whom selective migration should have been more stronger) and households in other economic sectors. As an additional robustness check we use nighttime lights data, and we find a higher luminosity in the former UFCo areas, a proxy for higher levels of income and economic activity.

The results of our chapter contribute to several strands of the literature. First, our chapter is related to a growing body of literature which analyzes the long-run impact of historical institutions on economic development (e.g., Banerjee and Iyer 2005; Gennaioli and Rainer 2007; Iyer 2010; Michalopoulos and Papaioannou 2013). Our chapter complements this literature by analyzing outcomes during the company's tenure, when the institutions were being formed, along with the short-, medium- and long-run effects of these institutions. Furthermore, while most of the existing literature on the long-run impact of historical institutions on economic development have considered settings where labor was coercive (the slave trade [Nunn, 2008], the *mita* system [Dell, 2010], the forced rubber cultivation [Lowes

and Montero, 2016], or the Dutch Cultivation System [Dell and Olken, 2017]), we study an institution within the context of global free-market capitalism. In contrast to most of the literature which has found long-run negative economic impacts due to institutional persistence, we find a positive effect of the UFCo. Similar to Dell and Olken (2017), who also found a positive effect, in our case the investments carried out by the company can explain the results.

Furthermore, our chapter uses a novel dataset that contains geo-referenced information of the census block where each household is located, collected across several censuses. This feature of the dataset allows us to work at a higher level of disaggregation and gives us more flexibility to test the robustness of our results. This level of disaggregation is rarely available, especially in developing countries, and is key when performing a geographic RD to precisely identify observations at a nearby distance. On this line, we complement previous work using RD on developing countries at a higher level of aggregation (e.g. district level like Dell (2010) and Dell et al. (2015); or village level like Lowes and Montero (2016)), by conducting a finer analysis that can get very close to the boundary increasing the precision of the results, and including placebo tests that move the boundary only a few kilometers.[3]

Our chapter also speaks to the debate on the effects and spillovers of Foreign Direct Investment (FDI). Empirical studies of the effects of FDI have produced mixed evidence. While some studies find evidence of FDI being beneficial using macro- and micro-data (e.g., Blomstrom 1986, Blomstrom and Wolff 1989, Lipsey 2002, Smarzynska Javorcik 2004, Harrison and Rodríguez-Clare 2009), others are not so optimistic about these benefits, especially for developing countries (e.g.,

[3]One exception is Becker et al. (2016), that uses a level of aggregation based on electoral registers or census enumeration areas. However, their dataset is representative only at the national level.

Aitken and Harrison 1999 , Borensztein et al. 1995, Xu 2000, Alfaro et al. 2003, Alfaro and Charlton 2007). We provide novel micro-evidence of the benefits of large-scale FDI through our historical setting, arguing that, *even in the case of the multinational that inspired the term "Banana Republic,"* we find evidence of positive and persistent effects.

Finally, our chapter also contributes to the study of the legacy of the UFCo, one of the largest and most controversial multinationals in history. With arms over both economic and political spheres in many Latin American countries, the UFCo became known as *El Pulpo* (the octopus). Therefore, it is not surprising that the company inspired an extensive body of literature, ranging from fiction works to academic research.[4] Virtually all studies that rely on quantitative data consider the impact of the UFCo at the aggregate level, analyzing national or local trends in productivity, land patterns, export levels, and labor mobility (e.g., Casey 1979; Ellis 1983; Viales 1998; Royo 2009). To the best of our knowledge, our chapter is the first analysis of the legacy of the UFCo using microeconomic data to obtain quantitative estimations regarding its impact.

The rest of the chapter is organized as follows. Section 3.2 provides an overview of the historical background. Section 3.3 includes details of the data used in our analysis. We describe our estimation framework in Section 3.4. Section 3.5 includes our results and a discussion about the mechanisms behind our findings. Finally, Section 3.7 concludes.

[4]Some examples of fiction works inspired by the relationship between the UFCo and the Latin American countries are the novels: "Mamita Yunai" by Carlos Luis Fallas, the "Banana Republic Trilogy" ("Strong Wind," "Green Pope," and "The Eyes of the Interred") by Miguel Ángel Asturias, and "One Hundred Years of Solitude" by Gabriel García Márquez.

3.2 Historical Background

3.2.1 The United Fruit Company: An Octopus in Latin America

The United Fruit Company (UFCo) was founded in 1899 and was engaged in the banana business. The banana is a fruit that perishes quickly and easily. To avoid losses, the UFCo coordinated the whole banana production process from the beginning to the end.[5] Through its investments in all production stages, the company became the first vertically integrated fruit multinational (Jones, 2005, p. 51). The UFCo acquired lands in scarcely populated humid lowlands and transformed these lands from tropical forests to plantations and towns that revolved around banana activity. To increase labor efficiency in remote locations threatened by tropical diseases, the UFCo provided healthcare, housing, and sanitation to its workers. The UFCo also invested in infrastructure, such as wireless communication systems to coordinate the whole process, and railroads to carry the bananas from the plantations to the ports. Once in the ports, the bananas were shipped to the United States and Europe in the company's vessels, known as the Great White Fleet. Finally, the bananas were distributed to wholesale and retailers through its subsidiaries.[6]

In three decades, the UFCo became the world's largest banana producing and

[5]See Cutter (1926) or Palmer (1932) for a detailed account of the UFCo banana production stages.

[6]The UFCo was responsible for popularizing bananas and changing consumption patterns. While in the 1880s bananas were considered an expensive and exotic fruit in the United States, by 1910 they already were part of daily diet. Mass consumption of banana in the United States was mostly a result of the UFCo advertising bananas as convenient, delicious, and nutritious; through recipe books, films, songs, and educational materials for classroom use (Jenkins, 2000).

marketing corporation, consolidating a monopoly in the banana market (Ellis 1983, p. 42; Bucheli 2005, p. 49). By 1930, production from the UFCo represented close to 80% of the global banana production. According to the UFCo's Annual Reports to the Shareholders, by 1930, the company landholding reached 13,339.12 km^2 (roughly the size of Connecticut). The company operated in Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, Guatemala, Honduras, Jamaica, Nicaragua, and Panama (May and Lasso, 1958, p. 104). Although its principal product was bananas, the UFCo also produced sugar, cacao, oil palm, abacá (Manila hemp), and rubber. After 1930, the hegemony of the UFCo was affected by several challenges (Wiley, 2008, p. 37-38). First, a fungus, known as Panama disease, wiped out the output of entire plantations. Then, the Great Depression and Second World War reduced the demand in consumer countries.

After 1950, the challenges to the UFCo hegemony continued as a result of a different environment in its host countries, in the US, and in the banana industry. In its host countries, the UFCo faced increased labor union activity and growing nationalism (Wiley, 2008, p. 69). In the US, the Department of Justice filed a case against UFCo for violating antitrust legislation that forced the company to sell part of its holdings (Bucheli, 2005, p. 61). Finally, in 1956 the banana industry had a major technological transformation when the UFCo's main rival, Standard Fruit Company, introduced a variety of banana that was resistant to the Panama disease (Ellis, 1983, p. 176-178). More problems followed, as in 1974, Hurricane Fiji destroyed 70% of the company's plantations in Honduras (Mirabile and Derdak, 1990, p. 595). By 1984, the company made significant divestitures by selling several subsidiaries and focusing mainly on marketing bananas, buying the product from local producers.

3.2.2 Banana Plantations and the United Fruit Company in Costa Rica

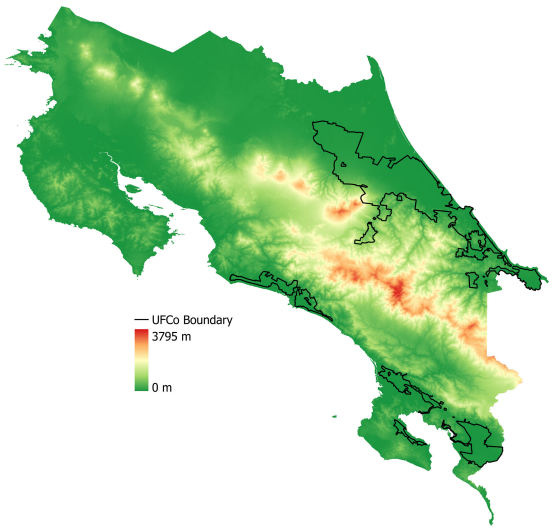
The history of banana plantations in Costa Rica goes back to the construction of a railroad to the Caribbean Coast. Since the 1840s, coffee rapidly expanded and became virtually the single Costa Rican export product, representing around 90% of total exports between 1840 and 1890 (Hall, 1978, p. 41). However, even though Europe was the main export destination, Costa Rica's only port was on the Pacific Coast. Although the Caribbean Coast provided a more efficient access to Europe, it was completely isolated and without any relevant settlements (Hall, 1978, p. 59-67). Therefore, in 1871 the government contracted the construction of a railroad to the Caribbean Coast.

The limited knowledge of the Caribbean Coast led to underestimating the costs of the construction. To pay for the unanticipatedly high costs of the railroad the government gave a concession of 3,333 km^2 of undeveloped land (roughly the size of Rhode Island, and equivalent to 6.5% of the national territory), and the lease of the railroad for 99 years (Casey, 1979, p. 26) to an American contractor, Minor C. Keith. After completing the construction, Keith experimented with exporting the bananas he had planted along the railroad tracks to feed the workers during the construction (Bucheli, 2005, p. 46). The experiment was successful, and Keith organized the Tropical Trading and Transport Company, which merged in 1899 with the Boston Fruit Company to form the UFCo. In the subsequent years, the banana business increased in volume and importance, and by 1905 bananas had reached the same place in Costa Rica's exporting value as coffee.

The banana activity and in particular the UFCo significantly transformed the

Costa Rican economy. Banana production gave rise to centers of economic activity in the lowlands, far from the regions specializing in coffee farming. By 1950 the UFCo was responsible for 58% of the country's total exports. The UFCo landholdings in the country represented roughly 4% of the national territory (as shown in Figure 3.1). Moreover, the UFCo employed approximately 7% of the country's total labor force and 12% of its agricultural labor force. The company operated three of the four ports in the country and around 89% of the railway infrastructure. The major role that the multinational had within Costa Rica reinforces its choice as a case study on how large-scale FDI can impact economic development. Finally, in 1984 the UFCo abandoned banana production in Costa Rica, in part due to a general corporate strategy that divested in the production process to focus on marketing. Appendix C.1 provides more historical details.

Figure 3.1: UFCo boundary.



Notes: Elevation is shown in the background.

3.3 Data

3.3.1 Census Data

We obtained restricted-access microdata from Costa Rican Censuses collected by the National Institute of Statistics and Census (*Instituto Nacional de Estadística y Censos [INEC]*) for years 1973, 1984, 2000, and 2011. As the UFCo stopped operations in 1984, the range covered by these censuses allows us to analyze the outcomes during and after the company's tenure.

The data is recorded at the census block level. The census block is the smallest territorial division of the country. The size and delimitation of a census block changed across censuses. For 1973, 1984, and 2000 censuses, each census block contains approximately 60 dwellings in urban areas and 40 dwellings in rural areas, and it tends to be constituted by one or two city blocks in urban areas (Bonilla and Rosero, 2008). For the 2011 census, in most cases, the census block coincides with a city block (Fallas-Paniagua, 2013). For all years, the data include each census block centroid's latitude and longitude coordinates. The level of spatial disaggregation provided by the census block data allows us to compare observations within close proximity of each other.

Except for the 1973 census, which includes information on wages, later censuses do not contain direct measures of income or consumption. Therefore, we follow the "Unsatisfied Basic Needs" (UBN) method to generate variables that measure economic outcomes. The UBN method was introduced by the Economic Commission for Latin America and the Caribbean (ECLAC) to identify households in poverty without relying on income data (Feres and Mancero, 2001). The method

requires specifying a set of basic needs and a threshold for attaining those needs (Armendáriz and Larraín B., 2017). Méndez and Trejos (2004) propose a set of unsatisfied basic needs for Costa Rica using data from the 2000 census, and it is straightforward to apply their method to the 2011 census (Méndez and Bravo, 2014). Their methodology defines four basic-needs dimensions: housing, sanitation, education, and consumption. Each dimension consists of components selected by its explanatory power for income in household surveys.

We follow the methodology in Méndez and Trejos (2004), but to be able to adapt it to the 1973 and 1984 censuses, we select the components whose information is available in all the four censuses. In the end, we also have the same four basic-needs dimensions. Appendix C.2 details the components that constitute each of our dimensions, and the specific variables from the censuses that we use. A general description of each dimension is the following: (i) housing: refers to the quality of the household dwelling's material and household overcrowding; (ii) sanitation: refers to the method for disposal of human excreta that the household uses; (iii) education: refers to school attendance and academic achievement for household members from 7 to 17 years old; and (iv) consumption: refers to the relationship between the number of income recipients (employed, pensioned, or renter), their years of schooling, and the total number of household members. We construct each dimension as an indicator variable equal to one if the household does not meet the threshold to attain a need in some component, and zero otherwise.

We consider a household as poor if it has at least one unsatisfied need. Moreover, we estimate the severity of poverty through the total number of UBN. Namely, the total number of UBN is an index that ranges from 0 to 4, where each unsatisfied

basic need adds one point to the index.

3.3.2 Historical Data

We also use data from 1864, 1892, 1927, 1950, and 1963 Costa Rican Censuses. Although these censuses do not contain enough spatial detail to be considered in our regression discontinuity (RD) design, the information allows us to analyze aggregated population patterns, such as migration before and during the UFCo apogee, or the size and occupation of the country's labor force.

Documents published by the UFCo provide valuable information to explain the mechanism behind the effect found in our chapter. From 1912 to 1931 the Medical Department of the UFCo issued an annual report describing the sanitation and health programs carried out by the company as well as the living conditions within the UFCo plantations. Moreover, the company regularly circulated reports with information about wages, number of employees, production, and investments in areas such as education, housing, and health. We obtained primary print copies of these documents from collections held by Cornell University, University of Kansas, and the Center for Central American Historical Studies (*Centro de Investigaciones Históricas de América Central [CIHAC]*).

Moreover, we gathered Costa Rican Statistic Yearbooks, which from 1907 to 1917 contain information on the number of patients and health expenses carried out by the hospitals in Costa Rica, including the ones ran by the UFCo. The Costa Rican Statistic Yearbooks from 1908 to 1948 and the Export Bulletin 1941-1947 contain data on product exports and port of export. Finally, 19 agricultural censuses taken between 1900 and 1984 provide information to track changes in

land use in the country, and the ones corresponding to 1950, 1955 and 1963 allow us to estimate worker productivity.

3.3.3 Geographic and Spatial Data

We obtained the elevation and temperature data from the Global Climate Database created by Hijmans et al. (2005). The spatial resolution is 30 arc-seconds (approximately 1 km^2 at the equator). The elevation above sea level is in meters and was constructed using NASA's Shuttle Radar Topography Mission (SRTM) data. From the elevation information, we calculate the slope data (in degrees). Hijmans et al. also compiled monthly averages of temperature measured by weather stations from 1960 to 1990. We measure temperature in Celsius and take the annual average for our analysis.

Location of the UFCo To help organize the production and keep track of land use, the UFCo Engineering Department made maps of the company's properties. The maps contain detailed information about farms, railroads, buildings, and landforms. The National Archives of Costa Rica (*Archivo Nacional de Costa Rica*) has a collection of such maps. Although the Virtual Map Library of the National University of Costa Rica (*Mapoteca Virtual de la Universidad Nacional de Costa Rica*) has digitized part of the collection, collecting all available maps required in-person visits to the archives, taking high-quality pictures of the original maps, and digitizing them. Based on these maps we obtain data on the location of the plantations. Figure C.5 in Appendix C.3 provides an example of a map showing the UFCo landholdings in the Costa Rican Pacific Coast.

Nighttime Lights Data We use nighttime lights data as a robustness check of our main results, treating satellite-recorded data on nighttime lights as a proxy for income and economic activity. A series of papers that have shown a strong correlation between nighttime lights and economic activity (Chen and Nordhaus 2011; Henderson et al. 2012; Michalopoulos and Papaioannou 2014; Hodler and Raschky (2014)). The data on nighttime light is collected by the US Air Force Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS) and is processed by the National Oceanic and Atmospheric Agency's (NOAA) National Geophysical Data Center (NGDC). The data covers the years 1992 to 2013 at a spatial resolution of 30 arc-seconds. For each grid cell, an integer between 0 (no light) and 63 represents its light intensity.

3.4 Estimation Framework

To estimate the causal effect of the UFCo, we use well-defined boundaries based on historical records and compare observations located just inside former UFCo plantations to observations located just outside them. Our estimation of the *average* UFCo effect uses the following RD specification:

$$y_{igt} = \gamma UFCo_{gt} + f(\text{location}_{gt}) + \mathbf{X}_{igt}\beta + \mathbf{X}_{gt}\Gamma + \alpha_t + \varepsilon_{igt}, \tag{3.1}$$

where y_{igt} is an outcome of individual or household unit i in census block g and year t; and $UFCo_{gt}$ is an indicator variable equal to one if the census block g's centroid was inside a UFCo plantation, and equal to zero otherwise. $f(\text{location}_{gt})$ is a RD polynomial, which controls for geographic location of census block g. $\mathbf{X}_{igt}$ is a vector of covariates for individual or household unit i. $\mathbf{X}_{gt}$ is a vector of

geographic characteristics for census block g. α_t is a year effect. The coefficient γ is our main parameter of interest because it captures the effect of being in a zone under the UFCo's control.

The RD is a polynomial in latitude and longitude. Including latitude and longitude allow us to control by the exact location of units in the study region, and consequently, we compare households on opposite sides but in the same points of the UFCo boundary. Following Gelman and Imbens (2017), and in line with recent work whose estimation framework relies on a geographical regression discontinuity design (e.g., Dell et al. 2015; Lowes and Montero 2016), we use a linear RD polynomial and test for robustness to a variety of specifications.

Furthermore, to analyze the *dynamics* of the UFCo effect over time, we allow for a different UFCo effect coefficient in every census, by estimating the following RD specification:

$$\begin{aligned} y_{igt} =& \gamma_{1973} UFCo_{g,1973} + \gamma_{1984} UFCo_{g,1984} + \gamma_{2000} UFCo_{g,2000} + \gamma_{2011} UFCo_{g,2011} + \\ & f(\text{location}_{gt}) + \mathbf{X}_{igt}\beta + \mathbf{X}_{gt}\Gamma + \alpha_t + \varepsilon_{igt} \end{aligned} \tag{3.2}$$

where $UFCo_{g,t}$ is an indicator variable equal to one if at time t individual or household unit i is in census block g, whose centroid was inside a UFCo plantation; and equal to zero otherwise.

Quasi-Random Land Assignment: One of the assumptions required by a geographical RD is that all relevant factors besides treatment must vary smoothly at the UFCo boundary. This assumption is required to claim that observations

located just across the former UFCo plantations are a good counterfactual for those located just across the non-UFCo areas.

In general, the UFCo followed a policy of continuous expansion to new fields to counter soil exhaustion and the appearance of banana diseases (Jones and Morrison, 1952). At the moment of deciding which lands to own and operate, the firm took into consideration geographic characteristics (Casey 1979, p. 46; Cerdas Albertazzi 1993, p. 122-123). This land accumulation pattern, coupled with the varied Costa Rican landscape explain why geographical features change discretely along many segments of the UFCo boundary.[7] For example, Figure 3.1 shows that in the Pacific Coast, the UFCo domains coincide with the lowlands, while nearby regions have a higher elevation. As a consequence, the areas out of the UFCo landholdings were mostly inappropriate for tropical fruit production.

However, in the Atlantic Coast, a clarification of the land concessions granted by the government lead to a zone where the land was assigned quasi-randomly. Initially, due to ambiguities in certain capsules of the contract, the UFCo and the government had some discrepancies on the land concessions. In 1904, a legislative decree resolved the differences in criterion between the government and the UFCo. As a result, the modification declared some wastelands, that the UFCo considered as part of the original concessions, as state property, establishing a region known as Astúa-Pirie (Soley, 1940, p. 90). The decree specified that the property rights over these lands should be sold only to nationals (Viales, 2012).

Because the Atlantic Coast was completely isolated from the more densely

[7] Although Costa Rica covers a small continental area of 50,980 km^2, the topographic extremes, combined with changes in slope, wind direction, and orographic precipitation result in varied micro-climates, vegetation cover, and soil types within short distances (Alvarado and Cárdenes, 2016).

populated areas and most of it was unexplored,[8] the Astúa-Pirie Region's limits were drawn using salient features of the landscape as a reference. In particular, the boundary was chosen so that it would be easy to enforce for the local authorities. The legislative decree declared that on the south the boundary *"follows the Reventazón River, from La Junta to the Caribbean Sea."*[9] On the east, the boundary adjoins the *"Atlantic Ocean"*. On the northern part of the region, the boundary *"follows an imaginary line drawn from the intersection between Toro Amarillo River with the old railroad up to a point in the coast located five miles northeast from the mouth of Tortuguero River."*[10] Finally, the western boundary *"follows the main railroad, from La Junta to the point where the railroad crosses Toro Amarillo River"*.(ANCR, 1904, p. 44)

Consistent with this quasi-random design of the boundary, we find that the geographic characteristics of the region change continuously across this segment of the UFCo boundary where we focus our attention (see Figure 3.2). Table 3.1 shows that elevation, slope, and temperature do not change discretely across this segment of the UFCo boundary. The unit of analysis to examine the geographic characteristics is a 1x1 km grid cell,[11] and we present both robust standard errors, and standard errors that account for spatial correlation (Conley, 1999).[12] Table 3.1 also shows that as we move far away from this segment of the boundary the differences in elevation, slope, and temperature become significant. Therefore,

[8]Keith wrote that when he first arrived "Limón and all the country between it and the cultivated portions of the interior was a dense wilderness. With the exception of the little village of Matina, which contained fifty or sixty inhabitants, not one individual was settled anywhere on the line. In fact, the route had not even been explored, and the rivers were first named when the engineers crossed them."(Keith, 1886, p. 8).

[9]"La Junta" was the point where the railroad from San José intersected the railroads from Limón and Guápiles.

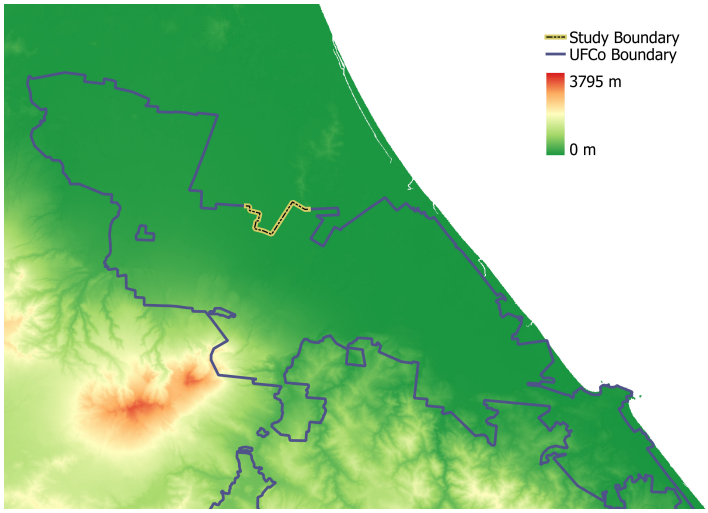
[10]The "old railroad" was the name given to the railroad to Guápiles because it was the remains of an unsuccessful previous attempt to build a railroad to the central valley.

[11]All results are similar if the census blocks are the unit of analysis.

[12]We compute Conley Standard errors at the cutoff distance of 2 km. However, the results are robust to alternative cutoffs.

exploiting our disaggregated data, and not to contaminate the analysis due to changes in the landscape, our chapter restricts attention to census block located at most within 5 km from this segment of the UFCo boundary.

Figure 3.2: Study boundary.



Notes: Elevation is shown in the background.

Table 3.1: Balance on Geographic Characteristics

	Sample falls within					
	<5 km of UFCo boundary			<10 km of UFCo boundary		
	Inside	Outside	s.e	Inside	Outside	s.e
Elevation	38.552	38.235	(1.330)	50.893	37.759	(2.273)***
			[3.530]			[6.514]**
Slope	0.256	0.312	(0.072)	0.493	0.328	(0.063)***
			[0.140]			[0.154]
Temperature	26.087	26.097	(0.006)	26.028	26.097	(0.011)***
			[0.014]			[0.031]**
N	96	85		168	141	

Notes: The unit of observation is 1x1 km grid cells. Robust standard errors for the difference in means between UFCo and non-UFCo observations are in parentheses. Conley standard errors for the difference in means are in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Preexisting social and economic characteristics: Besides the geographical characteristics, in a geographical RD preexisting social and economic character-

istics should also change smoothly at the boundary. In our case, the study area was practically uninhabited before the railroad construction and the UFCo arrival. According to the 1864 Costa Rican Census, only 545 people lived in the Atlantic Coast, a 0.45% of Costa Rican population at that time (Oficina Central de Estadística, 1868, p. 21).

Details on the Counterfactual: We document how public investment per capita in the region outside the boundary during the company's tenure was not significantly different from that on average Costa Rican rural areas. This indicates that our results are unlikely to be driven by the UFCo's presence crowding-out public investment in nearby regions. In particular, we gather data on government spending per municipality from annual reports from the Comptroller General of the Republic of Costa Rica (*Contraloría General de la República de Costa Rica*) published between 1955 and 1984. In Appendix C.4 we compare the spending per capita between UFCo municipalities and other rural municipalities and do not find significant differences.

Migration: Another assumption for an RD design is that individuals cannot precisely manipulate the assignment variable. On the one hand, differential rates of migration at the time of each census are relevant for our long-run analysis. Each census contains information about individuals' place of residence 5 years before the census took place. In census blocks located in UFCo areas, 9.35% of individuals migrated from a former non-UFCo municipality, while in the non-UFCo areas 11.90% of individuals migrated from a UFCo municipality. Table 3.2 shows that the migration rates are decreasing over time and their difference is not statistically significant. As a robustness check, we examine the influence of migration in the

estimates, with no change in our conclusions.

Table 3.2: Migration Rates in UFCo and Non-UFCo Census Blocks (Percentage)

Census	UFCo	Non-UFCo	P-value of the difference
	(1)	(2)	(3)
1973	16.83	32.74	0.37
1984	14.62	13.48	0.79
2000	7.45	10.25	0.24
2011	6.20	6.73	0.69
All	9.35	11.90	0.30

Notes: The p-values in the third column are for the test of the hypothesis that the rates of migration in the UFCo and non-UFCo areas are equal. The p-values are clustered at the census block level.

On the other hand, historical migration during UFCo operation is relevant. Table C.1 in Appendix C.1 summarizes the population dynamic from census data for Limón, the province where our area of study is located. Between 1883 and 1927 Limón had annual population growth rates higher than the national population growth rates; primarily a consequence of the railroad construction and banana activity (Casey, 1979, p. 214). Table C.2 in Appendix C.1 shows that the increase was driven by international migration. Initially, lightly populated Costa Rica combined with the hard working conditions made it difficult to recruit labor for rail construction and growing bananas. As a labor source, the company resorted to convicts from New Orleans jails (Bucheli, 2005, p. 46) and workers from the economically-depressed sugar plantations in the Antilles (Viales, 1998, p. 44-45). Both the population growth rate and the foreigner population in Limón decrease after 1927 to relatively low levels.

The fact that this migration took place, however, changes the interpretation of our estimates. Namely, we will interpret this selective migration, which is common whenever a large company like this multinational increases labor demand signifi-

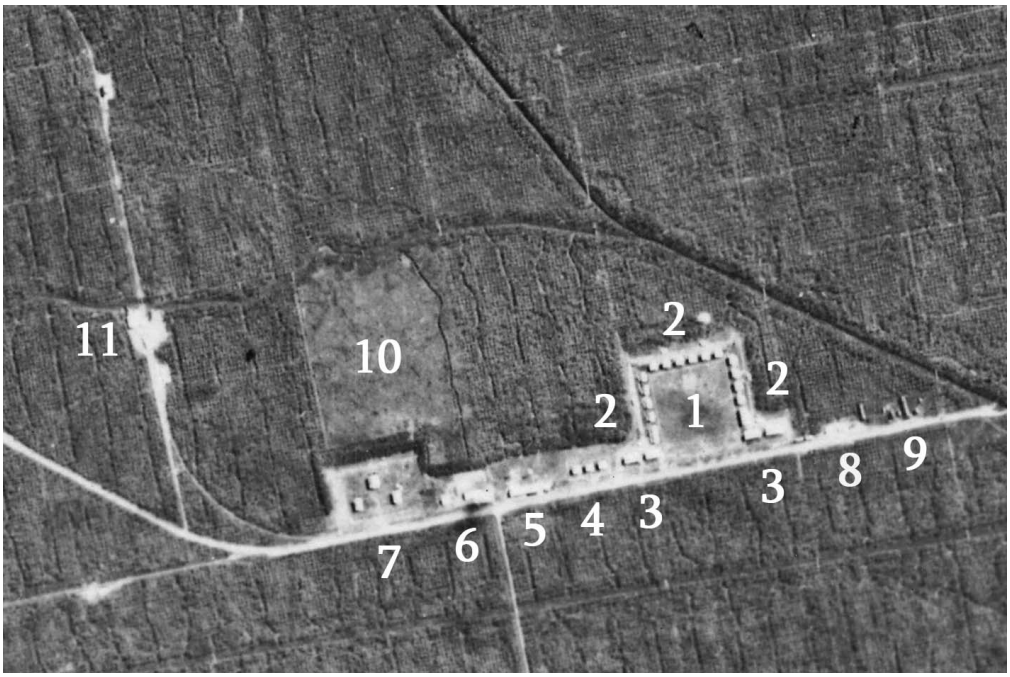
cantly, as part of the UFCo effect. Notice how, in our case, migrants seem to be negatively selected, as growing bananas requires low-skilled labor. This suggests that our positive UFCo effect might be a lower bound of the impact a multinational may have on development as compared, for instance, with a company inducing migration from high-skilled labor.

Commuting: Another concern for the identifying assumption could be that people who lived outside the UFCo plantations commuted and worked for the company or used its services. However, this is unlikely to be the case. Different from other types of agricultural activities that require labor on a temporal basis, the UFCo needed a permanent labor supply of around 150 workers per 800-acre farm. Due to the extension of the plantations and to reduce transportation costs, the UFCo created camps within their farms for its workers (Cerdas Albertazzi, 1993, p. 141). The typical farm consisted of 800 acres of land, with about 20 acres devoted to campsite and buildings, and 150 acres to pasture land (Jones and Morrison, 1952, p. 14). Besides houses and administrative buildings, special facilities were also present, such as commissaries, schools, electric plants, sewage systems, and recreational facilities (Wiley, 2008, p. 29). The wide range of services and facilities provided by the company converted plantations into communities that allowed people to live and work full time within them.[13] Figure 3.3 is an aerial photograph of an UFCo banana plantation we digitized from the National Geographic Institute (*Instituto Geográfico Nacional*) archive, and details the distribution of a typical banana farm based on descriptions in Jones and Morrison (1952, p. 14)

[13]For people within the plantations, the company was omnipresent in their lives. Harpelle (2001, p. 67) mention that typical residents "were likely born in the company hospital, educated in the company school, lived in company housing, obtained household supplies and clothing from the company commissaries, and, if they could afford it, looked forward to being carried to their final resting places in the Northern Railway's [a subsidiary of the UFCo] funeral car."

and Sandner (1962, p. 183).

Figure 3.3: Plan of a typical UFCo banana farm.



Notes: 1. Soccer field 2. Houses for married workers 3. Houses for single workers 4. Houses for foremen 5. Houses for spray laborers 6. Fungicide storage facility 7. Houses for timekeeper, overseer, and assistant overseer 8. Packing plant for wash and sort bananas 9. Stable for work animals, and shed for tractors and trailers 10. Pasture for work animals 11. Irrigation pump building, and irrigation pump operator's house. The photograph covers an area approximately equal to 1.5x1 km.
Source: National Geographic Institute. Project VV 91 PL. Line 559. Film: M1006. Photo 97. February, 14, 1948.

Moreover, people in the surrounding areas did not enjoy the services provided by the company. For example, as we describe in detail in Section 3.5, the available data on medical attention in UFCo hospitals suggests that few patients were not on the company's payroll, and for this group the average spending per patient was lower relative to patients within the company's payroll. Given that health-care provision was one of the services in which the UFCo was superior to any local provider (Casey, 1979, p. 114), this supports our claim that there was no

commuting to enjoy the amenities the company provided.

3.5 Results

Table 3.3 presents the results of estimating equation (3.1) using as dependent variables the probability of having an unsatisfied basic need (UBN) in each dimension (housing, sanitation, education, and consumption), the probability of being poor, and the total number of unsatisfied basic needs. We pooled data from the 1973, 1984, 2000, and 2011 censuses and estimate an average UFCo effect for all years. All regressions include geographic controls, demographic controls for the number of household members aged 0-4 (infants), 5-14 (children), and 15 and older (adults), census fixed effects, and a linear polynomial in latitude and longitude. We report standard errors clustered at the census block level and Conley standard errors.

Table 3.3: Contemporary Household Outcomes: Average UFCo Effect

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
UFCo	-0.095	-0.016	-0.057	-0.059	-0.124	-0.228
	(0.026)***	(0.017)	(0.022)**	(0.025)**	(0.031)***	(0.057)***
	[0.029]***	[0.015]	[0.019]***	[0.025]**	[0.026]***	[0.051]***
Adjusted R^2	0.102	0.173	0.241	0.015	0.115	0.200
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
Mean	0.176	0.060	0.235	0.200	0.481	0.670

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The estimates suggest that households located in the former UFCo region are in general better off. Columns (1) to (4) of Table 3.3 show that these households

have a lower probability of having a UBN in each of the dimensions we consider. Moreover, the UFCo effect decreases the probability of being poor by about 12.4 percentage points (column 5). Finally, the severity of poverty is lower in the former UFCo areas, where households have on average 0.228 fewer shortcomings than households in the non-UFCo region (column 6). Except for sanitation, all the magnitudes are statistically significant at the 1% or 5% level.

Table 3.4 explores how the UFCo effect changed across time. The company stopped operations in 1984, and we examine census data from 1973, 1984, 2000, and 2011. Therefore, we can disentangle the differentiated effects of the company's presence during its tenure, and also at different points in time after it stopped operating. We first discuss persistence on the probability of having a UBN (columns 1-4). For the housing dimension, the effect is very persistent across years. In 2011, approximately 30 years after the UFCo left, households within UFCo former lands are 9.3 percentage points less likely of having a UBN in housing relative to households outside. The magnitude of the UFCo effect in this dimension is high given the mean probability for the entire region (0.124). For the case of education and consumption, we find that although it maintains the negative sign on every census, its significance disappears after 2000. The effect on water infrastructure rapidly vanishes and is insignificant after 1973. Finally, columns (5) and (6) show that the overall probability of being poor and the total number of UBN are also quite persistent over time, being significant during every year of our study.

Besides persistence, the results suggest there has been a convergence in some of the dimensions over the years. We find that for sanitation and consumption the UFCo effect coefficient in 1973 is statistically different from the coefficient in 2011. More generally, the severity of poverty has decreased over time. While a household

in 1973 had 0.668 less UBN than a household outside, in 2011 the difference was reduced to 0.126, and the difference is statistically different from zero at 1% level.

Table 3.4: Contemporary Household Outcomes: Dynamics Across Years

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{1973}$	-0.202	-0.272	-0.069	-0.125	-0.229	-0.668
	(0.064)***	(0.081)***	(0.043)	(0.048)***	(0.070)***	(0.164)***
	[0.066]***	[0.081]***	[0.034]**	[0.045]***	[0.054]***	[0.149]***
$UFCo_{1984}$	-0.056	0.013	-0.086	-0.067	-0.081	-0.196
	(0.048)	(0.028)	(0.028)***	(0.049)*	(0.046)**	(0.093)**
	[0.034]*	[0.013]	[0.027]***	[0.030]**	[0.032]**	[0.063]***
$UFCo_{2000}$	-0.079	0.020	-0.057	-0.132	-0.132	-0.199
	(0.032)**	(0.017)	(0.022)**	(0.036)***	(0.036)***	(0.059)***
	[0.029]***	[0.017]	[0.019]***	[0.024]***	[0.031]***	[0.053]***
$UFCo_{2011}$	-0.093	0.021	-0.039	-0.014	-0.101	-0.126
	(0.030)***	(0.016)	(0.030)	(0.037)	(0.038)***	(0.064)**
	[0.033]***	[0.020]	[0.031]	[0.055]	[0.053]*	[0.095]
Adjusted R^2	0.103	0.199	0.241	0.017	0.116	0.206
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
$Mean_{1973}$	0.462	0.353	0.393	0.208	0.777	1.416
$Mean_{1984}$	0.209	0.060	0.362	0.201	0.579	0.832
$Mean_{2000}$	0.145	0.031	0.230	0.178	0.452	0.584
$Mean_{2011}$	0.124	0.018	0.156	0.215	0.402	0.512

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3.4 summarizes the results. Each dot corresponds to the centroid of a census block in our study area; a monochromatic color scale represents the average outcome value for the households within the census block, where lighter colors stand for better outcomes; and each dot's size represents the number of observations in the census block. The background in each sub-figure shows predicted values, for a finely spaced grid of longitude-latitude coordinates, from a regression of the outcome variable under consideration on the UFCo dummy and a linear polynomial in latitude and longitude. Panels 3.4c, 3.4d, 3.4e, and 3.4f present the

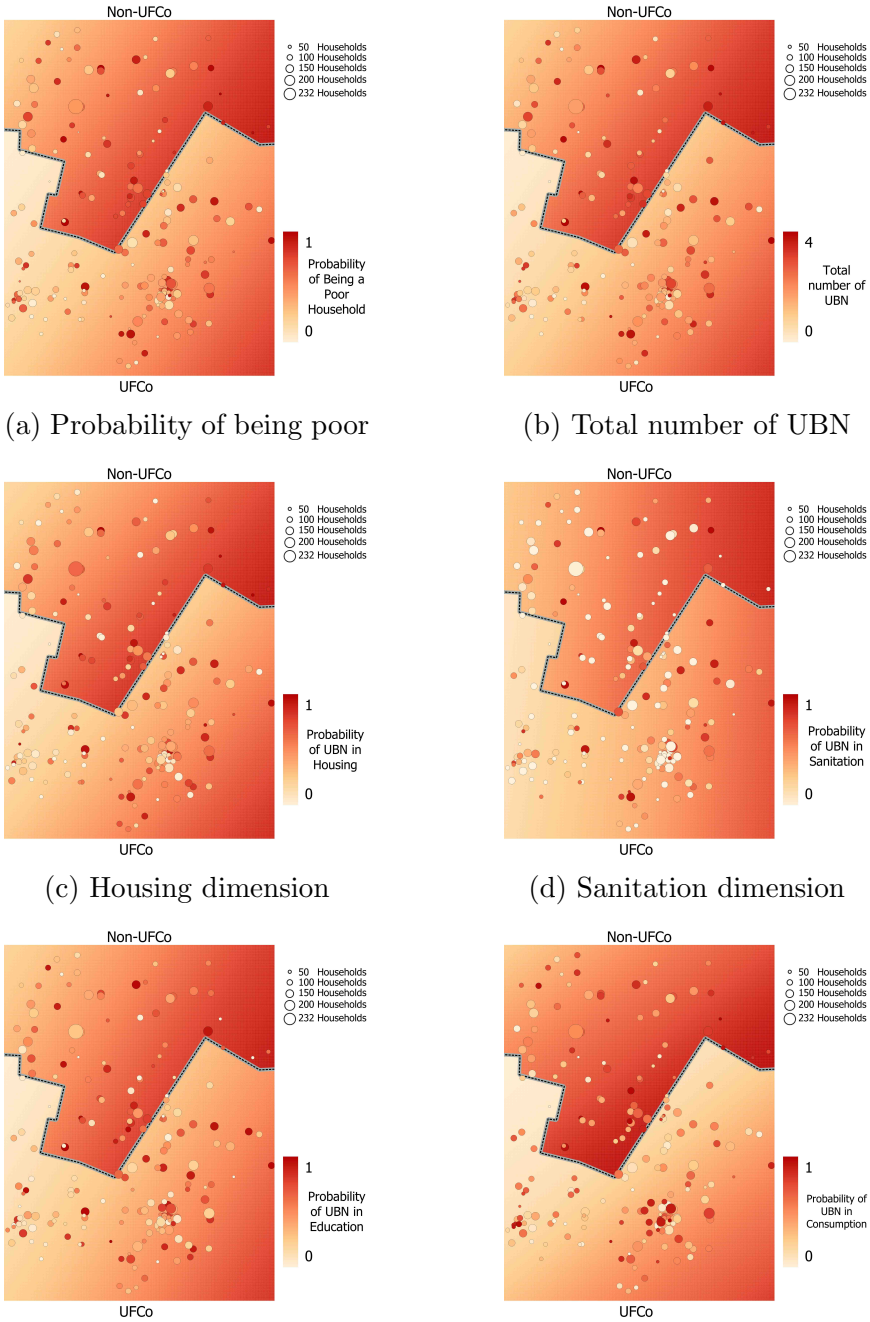
probability of having a UBN in housing, sanitation, education, and consumption respectively. Panel 3.4a shows the probability of being classified as a poor household and Panel 3.4b shows the total number of UBN. The predicted jump across the UFCo boundary is clear in all the sub-figures. Moreover, lighter dots tend to overlay lighter background areas, and the lighter areas (better outcomes) coincide with the former non-UFCo regions.

To understand why the UFCo had a positive impact on the economic development of the regions where it operated, we turn on to the analysis of the plausible mechanisms behind our results.

Investments in Sanitation Infrastructure and Healthcare: While Minor C. Keith was constructing the railroad to the Caribbean Coast in Costa Rica, he experienced the loss of around five thousand workers due to the unhealthy and dangerous conditions of the tropical forest (Bucheli, 2005, p. 46). The experience taught Keith the necessity to improve sanitation and hygiene to sustain a large enough workforce in an environment threatened by tropical diseases. As a consequence, the UFCo invested in sanitation infrastructure, launched health programs, and provided medical attention to its employees.

The UFCo was responsible for most of the sanitation programs that made possible to begin the colonization process in the Atlantic Coast of Costa Rica. While in 1884, more than half of the land in Puerto Limón was a swamp, by 1910 the area hosted the UFCo headquarters and had pipes, sewage system, street lighting, macadamized roads, and a dike (Sanou and Quesada, 1998, p. 270-272). Similarly, in the plantations areas, the UFCo drained swamps and invested in sewer and potable water systems (May and Lasso, 1958, p. 22).

Figure 3.4: Plots of the UFCo effect on contemporary household outcomes.



(a) Probability of being poor

(b) Total number of UBN

(c) Housing dimension

(d) Sanitation dimension

(e) Education dimension

(f) Consumption dimension

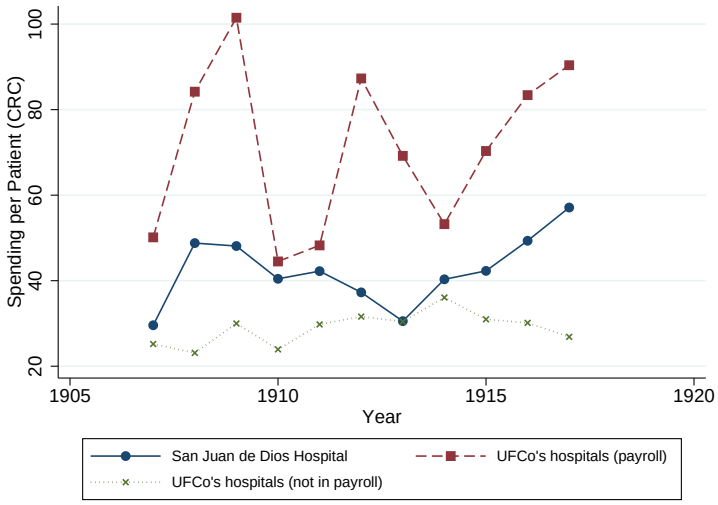
Notes: The figure shows the study boundary, with UFCo territories being south. Each dot represents a census block's centroid. Dot-color indicates the average outcome value for households, and dot-size represents the number of households in each census block. As shown, lighter colors stand for better economic outcomes.

In 1905 the UFCo established a Medical Department in Costa Rica to carry out sanitation programs and medical research on tropical diseases. By 1942 the company operated three hospitals in the country. The staff included doctors, sanitary inspectors, and nurses from the United States and other Central American countries (Morgan, 1993, p. 20). Each hospital had up-to-date surgical and X-ray equipment, laboratory, outpatient department, and steam laundry (Deeks, 1924, p. 1008).

To cover healthcare for employees and their dependents, the UFCo deducted 2% from their salary. For the employees, the deduction gave access without further charge to medical and surgical treatment, including medicines. For dependents, the coverage provided free medical supervision, but medicines were charged (UFCo, 1916, p. 76-77). Figure 3.5 shows that between 1907 and 1917, for people on the company's payroll the UFCo's hospitals spent 71.13 Costa Rican Colones (CRC) per patient, while the San Juan de Dios hospital, the largest hospital in the country at the time, spent 42.37 CRC per patient. Although a higher level of spending does not necessarily imply a higher quality of healthcare, the UFCo medical services were known of being among the best in the country (Casey, 1979, p. 114).

In general, the UFCo's healthcare programs were successful in controlling diseases in the plantation regions (Kepner 1936, p. 118; Chomsky 1996, p. 101). In 1929 a farm superintendent wrote that the "sanitary measures have helped to stabilize labor and increase their ability to perform work [...] during recent years with little or no influx of labor we have not experienced the recurrent shortages of labor that used to occur in previous years" (UFCo, 1929, p. 10). For the particular case of malaria, cataloged by the company as "our most important disease, when considered from the standpoints of morbidity and the effects of malaria on labor

Figure 3.5: Hospital spending per patient (in Costa Rican Colones), 1907-1917.



Notes: San Juan de Dios was the largest Costa Rican hospital at the time, located in the capital, San José.
Source: Authors' calculations based on 1907 to 1917 Costa Rican Statistic Yearbooks.

efficiency" (Deeks, 1928, p. 94), the UFCo carried out an intensive campaign that reduced its prevalence from 29.5 percent in 1926 to 14 percent in 1930 (Salisbury, 1930, p. 34).

Despite the positive impact of the UFCo programs, its benefits were restricted to employees and their immediate families. The general manager of the Medical Department explained that given the size of the UFCo landholdings, it was impossible from a commercial standpoint to sanitate completely all areas and therefore their efforts were "mainly directed to protecting the larger communities and camps where our employees are located" (UFCo, 1921, p. 6). In fact, to increase sanitary benefits, company doctors suggested preventing workers from traveling between plantations and surrounding villages, which were unscreened. Although non-employees could receive medical attention in the UFCo health care network, they had to pay higher fees. According to Figure 3.5, the average spending for

patients *not* on their payroll was 28.92 CRC, lower than the 71.13 CRC that on average was spent on people on the company's payroll.

Overall, the company's sanitary and health programs were hard to implement among the workers. Workers had their strong beliefs about the causes and treatment of disease and often preferred their cures (Chomsky, 1996, p. 109). According to the UFCo Medical Department, their workers had to be disciplined because "they are only children who have never grown up mentally, and their helplessness should always stimulate us to give them our best assistance" (López, 1930, p. 167). Consequently, the company's sanitary and health programs were compulsory for its workers, and there were punishments to induce the desired behavior. For example, fines were levied against managers who did not report an ill worker; or wages were withheld for workers who did not comply with hygienic instructions (Morgan, 1993, p. 23). Even preachers were instructed to preach health along with salvation (Kepner, 1936, p. 113).

Investments on Housing Infrastructure: Given the remoteness of the plantations and to reduce transportation costs, the UFCo provided the majority of its workers with free housing *within* the company's land. Each of the UFCo's divisions consisted of farms, and each farm had a camp where workers lived, as Figure 3.3 shows.

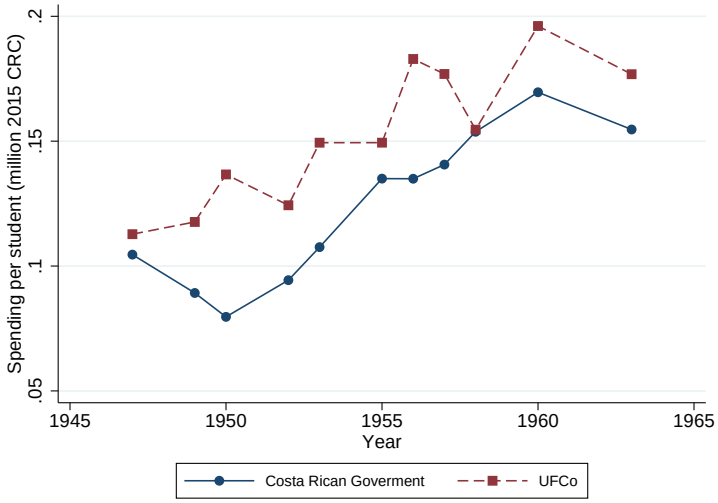
Usually, the houses for plantation laborers were laid out around a soccer field. By 1958 the majority of laborers lived in barracks-type structures. Single families occupied the majority of barracks, and there were buildings for unmarried workers (May and Lasso, 1958, p. 184-185). The barrack structures exceeded the standards of many surrounding communities (Wiley, 2008, p. 29).

Related to the sanitary programs implemented by the UFCo, a squad cleaned the grounds, collected trash, systematically sprayed with DDT to control for mosquitos and insects, and scrubbed out public toilets and bathing facilities. Moreover, the water supplied to the taps was safe for drinking. Besides housing, the UFCo provided basic services *for its employees* within each camp, such as schools, commissaries, dispensaries, and recreational facilities. May and Lasso (1958, p. 209) claim that "the places of worship, recreational facilities, and athletic fields and equipment provided for United's workers are upon a scale matched by few, if any, locally owned agricultural enterprises."

Investments in Human Capital: Although the UFCo was a multinational extracting a natural resource, it was in charge of all the production stages, from plantation to distribution. Besides agricultural workers, the company needed labor with some basic skills to perform administrative tasks, such as supervisors in its farms, or retail clerks in its commissaries. Furthermore, having employees who understand English was a valuable skill for the UFCo because the company kept its records in this language, and also it facilitated communication with the UFCo higher managers (principally United States citizens), and with immigrant workers (principally Jamaicans) (Castillo-Serrano, 1998, p. 40-46). As a result, primary education to the children of its employees was among the services that the company provided within the camps. The curriculum in the schools included vocational training and, before the 1940s, was taught mostly in English. The emphasis on primary education was significant, and child labor became uncommon in the banana regions (Viales, 1998, p. 61). By 1955, the company had constructed 62 primary schools within its landholdings in Costa Rica (May and Lasso, 1958,

p. 148). As shown in Figure 3.6,[14] spending per student in schools operated by the UFCo was consistently higher than public spending in primary education between 1947 and 1963.[15] On average, the company's yearly spending was 23% higher than government spending during this period.

Figure 3.6: Spending per student (in 2015 Costa Rican Colones), 1947-1963.



Source: Authors' calculations based on company reports ("*Compañia Bananera de Costa Rica. Algunos datos sobre sus actividades*" for 1947, 1949, and 1950; and "*Datos*" for 1952, 1953, 1955, 1956, 1957, 1958, 1960, and 1963) and Molina (2017).

By the time children completed primary education, they were old enough to work. The UFCo did not provide directly secondary education although offered some incentives. If the parents could afford the first two years of secondary education of their children in the United States, the UFCo paid for the last two years and provided free transportation to and from the United States. Moreover, if the

[14]In Figures 3.6 and 3.7 the amounts were converted to constant 2015 Costa Rican Colones (CRC) by splicing four price indexes: (i) Cost of Living Index Base 1936 = 100 (*Índice de costo de la vida Base 1936 =100*); (ii) Consumer Price Index for Middle Income and Low-Income Citizens in the Metropolitan Area Base 1964 = 100 (*Índice de precios al consumidor de ingresos medios y bajos del Área Metropolitana Base 1964=100*); (iii) Consumer Price Index Base January 1995 = 100 (*Índice de precios al consumidor Base Enero 1995 = 100*); and (iv) Consumer Price Index Base June 2015 = 100 (*Índice de precios al consumidor Base Junio 2015 = 100*).

[15]Data is only available for this subset of years.

parents organized secondary schools by themselves and paid a private tuition fee for the teachers, the UFCo provided a building and furniture (May and Lasso, 1958, p. 190). However, despite the incentives, secondary and tertiary education was costly and out of reach for most children of its employees.

To assess the impact of the UFCo educational investments on current human capital accumulation, we estimate equation (3.1) using educational attainment as the outcome variable. The results are presented in Table 3.5, restricting the sample to non-migrants. Column (1) shows a positive UFCo effect on human capital accumulation. Consistent with the emphasis on primary education by the company, column (2) shows a positive UFCo effect on primary education attainment. Individuals in the former UFCo areas are 5.3 percentage points more likely of completing primary education. On the other hand, in column (3) the effect of the UFCo presence on secondary education attainment is zero, in line with the higher costs of completing higher education levels.

Income and Consumption Capacity: On average, between 1946 and 1976, the company employed about 6.93% of the total agricultural workers in the country and 3.50% of the entire labor force, reaching a peak in 1949 when the numbers were 12.65% and 7.04%, respectively. The wages paid by the UFCo compared favorably to salaries paid in other agricultural activities in Costa Rica (Kepner 1936, p. 129; Casey 1979, p. 114; Viales 1998, p. 44). Table 3.6 shows that the difference in earnings was consistent over time. For the segment of the UFCo boundary where geographic characteristics are balanced, we find that hourly wages for agricultural workers were 6.58% higher in the UFCo plantations than in the non-UFCo area, although the magnitude is not statistically significant.

Table 3.5: Human Capital Accumulation

	Years of schooling	Primary	Secondary
	(1)	(2)	(3)
UFCo	0.269	0.053	0.003
	(0.130)**	(0.018)***	(0.009)
	[0.143]*	[0.020]**	[0.007]
Adjusted R^2	0.240	0.204	0.042
N	24,587	24,587	24,587
Clusters	198	198	198
Mean	4.595	0.462	0.056

Notes: The unit of observation is the individual. The sample is restricted to non-migrants. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; individual controls for age, age squared, and gender; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

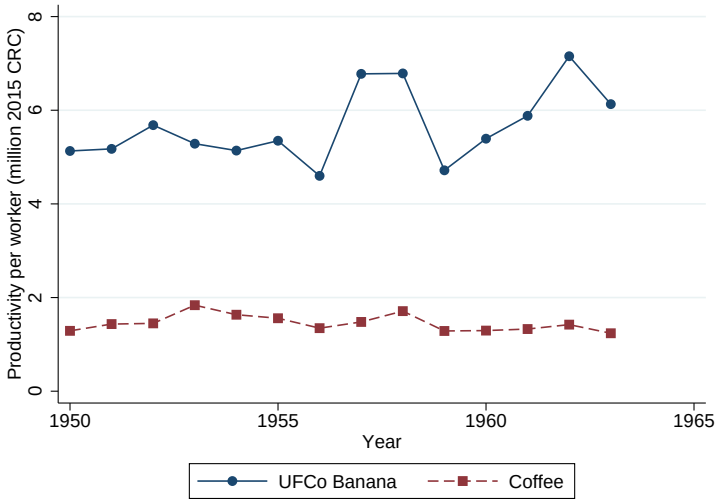
Table 3.6: Daily wages of agricultural workers in UFCo and non-UFCo municipalities

Year	UFCo Municipalities	Non-UFCo Municipalities
1916-1920	3.00	1.00-1.50
1933	2.75	1.00-1.60
1955	12.00	7.50-9.00
1973	20.85	13.05

Notes: These nominal wages are expressed in Costa Rican currency (colones). Before 1973 the data for UFCo municipalities correspond to the Limón area, while for the non-UFCo municipalities correspond to the central valley. Sources: Edelman (1992, p. 113), and authors' calculations based on 1973 Costa Rican Census.

The higher wages can be a consequence of differences in per-worker productivity in the UFCo plantations. Figure 3.7 compares return per worker between 1950 and 1963 in UFCo banana production versus coffee production (the other large Costa Rican export crop). On average, productivity per worker in the UFCo banana plantations (5.66 million 2015 CRC) was about four times productivity per worker in coffee (1.45 million 2015 CRC).

Figure 3.7: Worker productivity (in 2015 Costa Rican Colones), 1950-1963.



Source: Authors' calculations based on 1950, 1955 and 1963 Costa Rica Agricultural Censuses, and 1950 to 1963 Costa Rican Statistic Yearbooks.

Due to the remoteness of urban centers, the company established a system of commissaries in the plantations (Casey, 1979, p. 116-118). The commissaries also provided a psychological incentive to work by expanding the set of goods available to workers. The official company history explains as the justification for the commissary that "[...] the wage earners require goods for which they are willing to exchange a portion of their wages. Otherwise, there is no particular point in working" (Wilson, 1947, p. 238). Therefore, along with staple food items, shoes, and clothing, the commissaries sold durable and luxury goods.[16]

[16]In interviews conducted in Colombia to former workers of the UFCo, someone declared: "We

Regarding markups, initially, the company's stores exercised their monopolistic power in the plantation regions by charging abusive prices to the workers. The situation changed in the 1920s when the UFCo decided to provide commissary services on a break-even basis (Wilson, 1947, p. 238-239). In 1935, when the official exchange rate increased to 5.50 colones per dollar, the UFCo avoided a rise in the price of staple foods by keeping the previous exchange rate of 4.5 colones (Casey, 1979, p. 117). May and Lasso (1958, p. 196) mention that in 1956, prices in the UFCo commissaries and the state-run groceries average out about equal, even though the state-run had to sell items at price cost.

3.6 Discussion on Robustness Tests

Falsification Test: As a falsification test, we re-run the analysis using placebo borders. In particular, we draw fake borders at a distance of 2 km both inwards and outwards of the actual UFCo border. Appendix C.5 presents the results, showing that our placebo tests deliver insignificant results in every case, both economically and statistically.

Other Robustness Tests: Although in Tables 3.3 and 3.4 we use a linear polynomial in latitude and longitude, our main message is robust to alternative specifications of the RD polynomial. Appendix C.6.1 documents that using a quadratic polynomial, or a linear polynomial in latitude, longitude, and distance to the boundary leads to similar conclusions.

could find everything we needed in the company's store. From a needle to a refrigerator... There they carried imported good quality furniture, food, and clothes. Everything was much cheaper than in the other stores. Beforehand, the workers did not use silverware to eat and sleep on the floor. They dressed badly... Thanks to the company's store they were able to buy iron beds, silverware, and good imported clothes." (Bucheli, 2005, p. 127)

Besides the specification of the RD polynomial, we also analyze how the results change to varying the control variables. Appendix C.6.2 shows that results are robust to excluding demographic controls, to excluding geographic controls, and to excluding both demographic and geographic controls.

Recall from Section 3.2 that the construction of a railroad to the Caribbean Coast was fundamental in the genesis of the UFCo. To facilitate carrying bananas from the plantations to the ports in a fast and efficient way, the company owned and operated the railroad. Therefore, in Appendix C.6.3, we evaluate if the UFCo effect is mostly a result of the availability of nearby railroads and its associated benefits, such as increasing market access. We explicitly control for the presence of the railroad by including the distance of each census block centroid to a railroad. The results are unchanged, suggesting that the UFCo effect goes beyond the provision of railroads.

To exploit the availability of precisely georeferenced census block level data, we run our main specification restricting the sample to units within 1 km of the boundary. Appendix C.6.4 presents the results. Limiting the sample we ensure comparison of households located very close to each other, and we find results that are consistent with our findings within 5 km.

To consider if selective migration is generating the differences in living standards between the two regions, we reestimate equations (3.1) and (3.2) taking two approaches. In our first approach, we use a restricted sample of the full dataset in which we drop all migrant households. We classify a household as migrant if any household member lived in a different place of residence five years before the census took place.[17] Appendix C.6.5 documents that the results are similar to the

[17]Our results remain unchanged if we instead classify a household as migrant if the head of household lived in a different place of residence five years before the census took place (see Tables

estimates in Tables 3.3 and 3.4, and we cannot reject that the estimates are the same at the 10% significance level.

In our second approach, we compare the UFCo effect for households engaged in the agriculture sector versus other economic sectors. We consider a household as an agricultural household if any of its members work in agriculture.[18] Given that the UFCo was an agricultural multinational, arguably the individuals who were more likely to migrate to work with the multinational were the ones more capable of performing farm tasks. If ability in agriculture production is highly heritable and selective migration is driving our results, then the UFCo effect should be stronger for the households engaged in the agricultural sector relative to other economic activities. Nevertheless, Appendix C.6.6 shows that this is not the case, and for each outcome we consider, we cannot reject at the 10% level that the estimates are the same across both groups. In summary, the two approaches we take suggest that selective migration is unlikely to generate the differences between the regions.

Our Unsatisfied Basic Needs (UBN) are a modified version of the ones proposed by Méndez and Trejos (2004). Because Méndez and Trejos constructed the index using information from the 2000 and 2011 census, our modification consists of selecting the variables whose information is available in each of the 1973, 1984, 2000, and 2011 censuses. Therefore, as a robustness test, we re-run the estimation restricting the analysis to the 2000 and 2011 census and using the Unsatisfied Basic Needs (UBN) as proposed by Méndez and Trejos. Table C.26 shows that our main message is robust to this alternative definition of UBN.

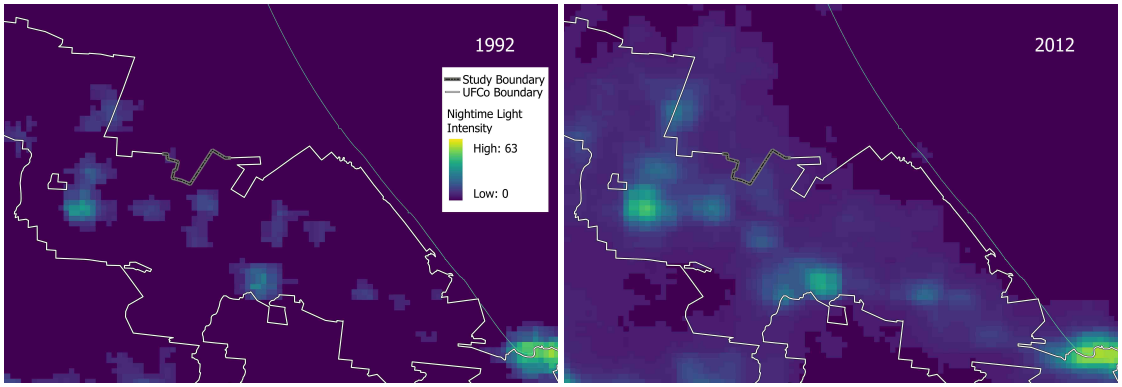
Finally, since the census data does not provide information on income, we use

C.22 and C.23 in Appendix C.6.5)

[18]Our results remain unchanged if we instead consider a household as an agricultural household if the head of household works in agriculture (see Table C.25 in Appendix C.6.6).

nighttime lights data as a proxy to confirm our findings through an alternative measure of economic development. Figure 3.8 presents the satellite image near the study boundary in 1992 and 2012 and suggests higher luminosity in areas inside the former UFCo landholdings. Column (1) in Table 3.7 confirms this difference in luminosity, by showing that nighttime light intensity is 21% (exp(0.193)-1=0.212) higher in the former UFCo plantations. To give a sense of the economic significance of this estimate, if we assume an elasticity between nighttime light intensity and GDP of 0.3 (consistent with the findings in Henderson et al. 2012 and Hodler and Raschky 2014), the 21% difference in nighttime light intensity implies that the output in the former UFCo plantations is about 6.37% higher. Column (2) shows that luminosity per capita is 18% (exp(0.165)-1=0.18) higher in the former UFCo plantations. Lastly, column (3) shows that the annual growth rate of luminosity per capita is 2.064 percentage points higher in the former UFCo areas. All estimates are significant at least at the 5% significance level.

Figure 3.8: Lights near the study boundary in 1992 and 2012.



A total of 9.2% observations in our luminosity data have a value equal to zero. The zero value can be due to a light that is too low for detection by the satellite, or because it corresponds to a sparsely populated area. Appendix C.8 presents the results after we account for the zero observations by adding 0.01 to the luminosity data (or luminosity per capita) before taking the logarithm. Our main

Table 3.7: Luminosity Data

	Log Light (1)	Log Per Capita Light (2)	Annual Growth Rate of Per Capita Light (3)
UFCo	0.193	0.165	2.064
	(0.006)***	(0.051)***	(0.781)***
	[0.017]***	[0.065]**	[0.953]**
Adjusted R^2	0.377	0.036	0.282
N	5,588	2,061	1,679

Notes: The unit of observation is 1x1 km grid cells located within 5 km of UFCo boundary. Robust standard errors are in parentheses. Conley standard errors are in brackets. All regressions include year fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

message remains unchanged. In general, the nighttime lights results are consistent with the estimates from our main specification by providing evidence that suggests significant higher levels of income and economic activity in the former UFCo areas.

3.7 Concluding Remarks

This chapter studies the impact of large-scale Foreign Direct Investment (FDI) on economic development. To do so, we consider one of the biggest agricultural multinational during the twentieth century. In particular, we examine the enclave of the United Fruit Company (UFCo) in Costa Rica. We use a regression discontinuity design that exploits a quasi-random assignment of land to the UFCo in 1904, and the availability of restricted-access census micro-data for 1973, 1984, 2000, and 2011, geo-referenced at the census block level. We find a positive and persistent effect on economic outcomes in areas where the company operated. Households in the former UFCo areas have a better satisfaction of basic needs (housing, sanitation, education, and consumption capacity), are less likely of being poor, and have a lower number of unsatisfied basic needs.

Using data we have collected from primary sources, we offer suggestive evidence that investments in physical and human capital carried out by the UFCo were relevant mechanisms behind our results. To pursue its corporate interests, the UFCo transformed tropical forests into plantation areas by implementing sanitation and health programs, improving housing and water infrastructure, establishing hospitals, and providing primary education. These investments can plausibly explain the positive effect of the multinational on the economic development of the former plantations areas.

That being said, it is worth highlighting that a limitation of our work is that our identification strategy allows us to analyze the impact of the UFCo only on lands administered by the multinational and its immediate vicinities, i.e., we are finding evidence for a micro effect. Our identification strategy does not enable us to study the macro effect of the UFCo in Costa Rica. The UFCo was a large multinational with influence on political and economic institutions that have repercussions at the national level. It is conceivable that, while the micro effect is positive, the aggregate effect is negative. Although it goes beyond the scope of the present chapter, we consider the analysis of the macro effect as a relevant question for future research.

Although anti-globalization sentiments are not new, since the onset of the Great Recession the debate has been intensified by the escalation in trade-restrictive measures and the increasing support to political parties with nationalist or protectionist agendas in different countries. Part of the discontent to globalization is due to the role of multinationals, accusing them of offering low wages and volatile jobs in host countries, avoiding taxes and exporting jobs abroad in home countries, and in general destroying local firms, creating monopolies, and having unregulated political

influence. The findings in our chapter speak to this debate by showing positive and persistent benefits from FDI even in the case of the controversial *Octopus* in Latin America.

APPENDIX A

APPENDIX OF CHAPTER 1: RISK PREFERENCES IN LIFE-INSURANCE DECISIONS

Appendix A.1. Identification

Proposition 1.1. *Take any three insured amount options:* $0 \leq m_a < m_b < m_c$, *with premium* p_a *for insured amount* m_a. *Denote by* $\tilde{p}_j$ $(j = b, c)$ *the premium that makes the household indifferent between insured amounts* m_a *and* m_j, *and assume that* $0.5(m_j - m_a) > \tilde{p}_j - p_a$. *Under standard expected utility, any* $(p_a, \tilde{p}_b, \tilde{p}_c)$ *is consistent with a unique* (r, l).

Consider a household with a coefficient of absolute risk aversion $r \neq 0$, a monetary loss l, and a vector of probabilities $(\mu_L, \mu_{D\&I}, \mu_S)$. Moreover, take any three insured amount options: $0 \leq m_a < m_b < m_c$. Let p_a be the premium for insured amount m_a.

First, Lemma A.1 shows that the household's willingness to pay is increasing in the insured amount option. Thus, for $0 \leq m_a < m_b < m_c$, $0 \leq p_a < \tilde{p}_b < \tilde{p}_c$.

Lemma A.1. *For any* $r \neq 0$, $p_a \geq 0$ *and* $m > m_a$, $\tilde{p}(m)$ *is a continuous and differentiable function with* $\dfrac{d\tilde{p}}{dm} > 0$.

Proof: Define V as:

$$V(m, l, r, \mu_L, \mu_{D\&I}, \mu_S, \tilde{p}) \equiv \mu_L(-e^{-r(-\tilde{p})}) + \mu_{D\&I}(-e^{-r(m-\tilde{p}-l)}) + \mu_S(-e^{-r(0.5m-\tilde{p}-l)}).$$

For $r \neq 0$, V is a continuously differentiable function that satisfies the conditions of the Implicit Function Theorem. Therefore, $\tilde{p}(m)$ is a continuously differentiable function such that:

$$\frac{d\tilde{p}}{dm} = \frac{-\dfrac{\partial V}{\partial m}}{\dfrac{\partial V}{\partial \tilde{p}}} = \frac{\mu_{D\&I}(e^{-r(m-l)}) + 0.5\mu_S(e^{-r(0.5m-l)})}{\mu_L + \mu_{D\&I}(e^{-r(m-l)}) + \mu_S(e^{-r(0.5m-l)})} > 0.$$

Now I proceed with the proof of Proposition 1.1:

Proof: By definition, $\tilde{p}_b$ is the premium that makes the household indifferent between insured amounts m_a and m_b:

$$\begin{aligned} &\mu_L(-e^{-r(-\tilde{p}_b)}) + \mu_{D\&I}(-e^{-r(m_b-\tilde{p}_b-l)}) + \mu_S(-e^{-r(0.5m_b-\tilde{p}_b-l)}) \quad = \\ &\mu_L(-e^{-r(-p_a)}) + \mu_{D\&I}(-e^{-r(m_a-p_a-l)}) + \mu_S(-e^{-r(0.5m_a-p_a-l)}). \end{aligned} \tag{A.1}$$

After solving equation (A.1) for l:

$$l(r) \mid_{m_a,m_b} = \frac{1}{r}\ln\left(\frac{\mu_L(e^{r\tilde{p}_b} - e^{rp_a})}{\mu_{D\&I}(e^{-r(m_a-p_a)} - e^{-r(m_b-\tilde{p}_b)}) + \mu_S(e^{-r(0.5m_a-p_a)} - e^{-r(0.5m_b-\tilde{p}_b)})}\right).$$

Note that $l(r) \mid_{m_a,m_b}$ is well behaved for $0.5(m_j - m_a) > \tilde{p}_j - p_a$.

Similarly, it can be obtained the curve $l(r) \mid_{m_a,m_c}$.

Note that the curves $l(r) \mid_{m_a,m_b}$ and $l(r) \mid_{m_a,m_c}$ intersect only once:

$$l(r)\mid_{m_a,m_c} \lesseqgtr l(r)\mid_{m_a,m_b} \iff$$

$$\frac{\mu_{D\&I}\big(e^{-r(m_a-p_a)}-e^{-r(m_b-\tilde{p}_b)}\big)+\mu_S\big(e^{-r(0.5m_a-p_a)}-e^{-r(0.5m_b-\tilde{p}_b)}\big)}{\mu_{D\&I}\big(e^{-r(m_a-p_a)}-e^{-r(m_c-\tilde{p}_c)}\big)+\mu_S\big(e^{-r(0.5m_a-p_a)}-e^{-r(0.5m_c-\tilde{p}_c)}\big)} \lesseqgtr \frac{e^{r\tilde{p}_b}-e^{rp_a}}{e^{r\tilde{p}_c}-e^{rp_a}}. \tag{A.2}$$

By condition $0.5(m_j-m_a) > \tilde{p}_j - p_a$, the left hand side term in A.2 is continuous and strictly increasing in r.

By Lemma A.1, the right hand side term in A.2 is continuous, and strictly decreasing in r.

Moreover:

- $\lim_{r\to-\infty} \dfrac{\mu_{D\&I}\big(e^{-r(m_a-p_a)}-e^{-r(m_b-\tilde{p}_b)}\big)+\mu_S\big(e^{-r(0.5m_a-p_a)}-e^{-r(0.5m_b-\tilde{p}_b)}\big)}{\mu_{D\&I}\big(e^{-r(m_a-p_a)}-e^{-r(m_c-\tilde{p}_c)}\big)+\mu_S\big(e^{-r(0.5m_a-p_a)}-e^{-r(0.5m_c-\tilde{p}_c)}\big)} = 0$
- $\lim_{r\to\infty} \dfrac{\mu_{D\&I}\big(e^{-r(m_a-p_a)}-e^{-r(m_b-\tilde{p}_b)}\big)+\mu_S\big(e^{-r(0.5m_a-p_a)}-e^{-r(0.5m_b-\tilde{p}_b)}\big)}{\mu_{D\&I}\big(e^{-r(m_a-p_a)}-e^{-r(m_c-\tilde{p}_c)}\big)+\mu_S\big(e^{-r(0.5m_a-p_a)}-e^{-r(0.5m_c-\tilde{p}_c)}\big)} = 1$
- $\lim_{r\to-\infty} \dfrac{e^{r\tilde{p}_b}-e^{rp_a}}{e^{r\tilde{p}_c}-e^{rp_a}} = 1$
- $\lim_{r\to\infty} \dfrac{e^{r\tilde{p}_b}-e^{rp_a}}{e^{r\tilde{p}_c}-e^{rp_a}} = 0$

In addition, around $r = 0$:

- $\lim_{r\to0^-} \dfrac{\mu_{D\&I}\big(e^{-r(m_a-p_a)}-e^{-r(m_b-\tilde{p}_b)}\big)+\mu_S\big(e^{-r(0.5m_a-p_a)}-e^{-r(0.5m_b-\tilde{p}_b)}\big)}{\mu_{D\&I}\big(e^{-r(m_a-p_a)}-e^{-r(m_c-\tilde{p}_c)}\big)+\mu_S\big(e^{-r(0.5m_a-p_a)}-e^{-r(0.5m_c-\tilde{p}_c)}\big)} =$
  $\lim_{r\to0^+} \dfrac{\mu_{D\&I}\big(e^{-r(m_a-p_a)}-e^{-r(m_b-\tilde{p}_b)}\big)+\mu_S\big(e^{-r(0.5m_a-p_a)}-e^{-r(0.5m_b-\tilde{p}_b)}\big)}{\mu_{D\&I}\big(e^{-r(m_a-p_a)}-e^{-r(m_c-\tilde{p}_c)}\big)+\mu_S\big(e^{-r(0.5m_a-p_a)}-e^{-r(0.5m_c-\tilde{p}_c)}\big)} =$

$$\frac{(\mu_{D\&I} + 0.5\mu_S)(m_b - m_a) - (\mu_{D\&I} + \mu_S)(\tilde{p}_b - p_a)}{(\mu_{D\&I} + 0.5\mu_S)(m_c - m_a) - (\mu_{D\&I} + \mu_S)(\tilde{p}_c - p_a)} \in (0,1)$$

By Condition $0.5(m_j - m_a) > \tilde{p}_j - p_a$

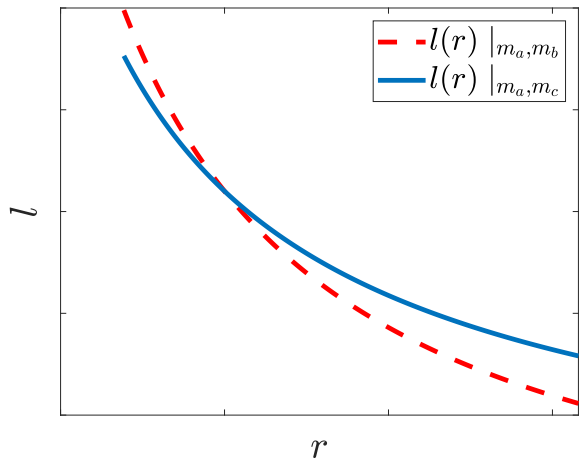
- $\lim_{r\to 0^-} \dfrac{e^{r\tilde{p}_b} - e^{rp_a}}{e^{r\tilde{p}_c} - e^{rp_a}} = \lim_{r\to 0^+} \dfrac{e^{r\tilde{p}_b} - e^{rp_a}}{e^{r\tilde{p}_c} - e^{rp_a}} = \dfrac{\tilde{p}_b - p_a}{\tilde{p}_c - p_a} \in (0,1)$ *By Lemma A.1*

Therefore, there exist a unique r^* such that:

- $l(r) \mid_{m_a,m_c} = l(r) \mid_{m_a,m_b}$ for $r = r^*$
- $l(r) \mid_{m_a,m_c} < l(r) \mid_{m_a,m_b}$ for any $r < r^*$
- $l(r) \mid_{m_a,m_c} > l(r) \mid_{m_a,m_b}$ for any $r > r^*$

Figure A.1 illustrates the intuition behind the proof. The maximum willingness to pay (WTP) for insurance increases in both the level of risk aversion r and in the level of the loss l. Therefore, as the risk aversion increases, the loss should decrease to keep the WTP unchanged. Given an initial insured amount m_a, knowing the WTP to moving to insured amount m_b is not enough to point identify the vector (r, l) because multiple combinations of the parameters generate the same WTP. However, if the WTP to moving to insured amount m_c is also known then the curves generated by each WTP will only intersect once. Stated differently, given the WTP to moving to insured amount m_b, then the lower the WTP to moving to insured amount m_c, so the larger must be r and the smaller must be l.

Figure A.1: Identification of the vector (r, l).



Appendix A.2. Analysis for the Restricted Sample (N=2,947)

In this appendix, I present the results of the analysis after restricting the sample to the 2,947 households who chose one of the 12 focal insured amounts. Tables A.1 and A.2 present the results assuming homogeneous preferences. Tables A.3 and A.4 report the results for the heterogeneous preferences case. The main message persists, suggesting that restricting the analysis to the subsample of households unaffected by the reassignment of insured amounts does not affect the conclusions.

Table A.1: Restricted Sample. Estimation of the Model Assuming Standard Expected Utility ($U(L(m))$ Specified by Equation (1.5))

	Estimate	95 percent bootstrap confidence interval	
$\ln(r)$	-15.4999	-15.5666	-15.4269
$\ln(l)$	16.9591	16.9038	17.0095
σ	0.0051	0.0043	0.0060
LL	-6533.10		

Note: Restricted sample of 2,947 households.

Table A.2: Restricted Sample. Estimation of the Model Assuming State-Dependent Utility and Probability Weighting ($U(L(m))$ Specified by Equation (1.6))

	Estimate	95 percent bootstrap confidence interval	
$\ln(r)$	-15.8030	-15.8567	-15.4354
$\Gamma_{D\&I}$	0.0017	0.0002	0.0019
Γ_S	0.0000	0.0000	0.0004
σ	0.0037	0.0033	0.0054
LL	-6532.28		

Notes: Restricted sample of 2,947 households. The loss is assumed to be equal to CRC 25 million.

Table A.3: Restricted Sample. Estimation of the Model Assuming Standard Expected Utility. Observed Heterogeneity

		Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_r$	Constant	-15.5689	-15.5884	-15.5470
	Female	-0.0182	-0.0347	-0.0045
	$30 \leq$ Age ≤ 39	-0.0241	-0.0397	-0.0105
	$40 \leq$ Age ≤ 49	-0.0396	-2.0799	-0.0255
	$50 \leq$ Age	-1.8783	-2.0606	0.0250
Mean fitted value	$\ln(r)$	-15.7416	-16.0942	-15.5937
σ		0.0038	0.0034	0.0046
LL		-6492.13		

Notes: Restricted sample of 2,947 households. The loss is assumed to be equal to CRC 25 million.

Table A.4: Restricted Sample Estimation of the Model Assuming State-Dependent Utility and Probability Weighting. Observed Heterogeneity

		Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_r$	Constant	-15.2973	-15.8142	-14.7091
	Female	0.5830	0.0755	0.7314
	$30 \leq$ Age ≤ 39	-0.0719	-0.1280	0.0744
	$40 \leq$ Age ≤ 49	-0.1863	-0.2552	-0.0140
	$50 \leq$ Age	-0.1777	-1.9347	-0.0172
$\boldsymbol{\beta}_{\Gamma_{D\&I}}$	Constant	0.0005	0.0000	0.0019
	Female	-0.0005	-0.0015	0.0000
	$30 \leq$ Age ≤ 39	0.0000	-0.0006	0.0000
	$40 \leq$ Age ≤ 49	0.0000	-0.0002	0.0002
	$50 \leq$ Age	0.0000	-0.0003	0.0015
$\boldsymbol{\beta}_{\Gamma_S}$	Constant	0.0000	0.0000	0.0002
	Female	0.0000	-0.0001	0.0001
	$30 \leq$ Age ≤ 39	0.0000	0.0000	0.0003
	$40 \leq$ Age ≤ 49	0.0000	0.0000	0.0000
	$50 \leq$ Age	0.0000	-0.0001	0.0003
Mean fitted value	$\ln(r)$	-14.9257	-15.6828	-14.6502
	$\Gamma_{D\&I}$	0.0001	0.0000	0.0008
	Γ_S	0.0000	0.0000	0.0003
σ		0.0136	0.0042	0.0211
LL		-6431.70		

Notes: Restricted sample of 2,947 households. The loss is assumed to be equal to CRC 25 million.

Appendix A.3. Alternative Probabilities

In this appendix, I present the results of the analysis using empirical frequencies estimated at the national level. An advantage of considering national level probabilities is that they should resemble actual population values. The drawback is that because the life insurance considered in this study is not compulsory, there may exist a self-selection problem and the national level probabilities might not reflect the clients' actual values.

The death probability source is the Costa Rican life table, estimated by the Central American Population Center of the University of Costa Rica (*Centro Centroamericano de Población de la Universidad de Costa Rica [CCP]*). The life table is based on population data for the years 2005-2010. The age-specific mortality is adjusted to change smoothly with age (CCP, 2012).

For the incapacity probability, I use claim data from the Labor Risk Insurance (*Seguro de Riesgos de Trabajo*). This insurance is compulsory and protects workers in case of injury, illness, or death due to circumstances related to their work. The claim data reports the number of workers that suffered a loss in earning capacity higher than 67%. I use this specific figure because the definition is similar to the one that applies to the insurance policy considered in my chapter. However, the probabilities obtained in this way are an approximation to the actual national level incapacity probabilities because incapacity due to non-work accidents are omitted (for which data is nonavailable).

To estimate the serious-disease probability, I use cancer incidence rates and hospitalization rates due to stroke, renal insufficiency, and myocardial infarction. For the cancer incidence rates, I use data from the National Tumor Registry (*Registro*

Nacional de Tumores). For the hospitalization rates due to stroke, renal insufficiency, and myocardial infarction I use data from the Bureau of Health Statistics Costa Rican Department of Social Security (*Área de Estadísticas en Salud de la Caja Costarricense de Seguro Social*).

Except for the life tables, all datasets cover the 2009-2015 period. Moreover, except for incapacity, each frequency is available at age group and gender level. To estimate the national level probabilities, I assign to each observation in the core sample the corresponding value, and then I take the mean. The empirical frequency at the national level for the death or incapacity event is 0.0014, while for the serious disease is 0.0016. For the analysis, I assume these values for each household.

I only re-estimate the standard expected-utility model because it is the only considered model that requires defining the probabilities for its estimation. Tables A.5 and A.6 present the results assuming homogeneous preferences and heterogeneous preferences, respectively.

However, in the model that includes state-dependent utility and probability distortions, the empirical frequencies are important for interpreting the results. Assuming homogeneous preferences, using the results from Table 1.9 and the empirical frequencies estimated at the national level, a state-dependent-utility interpretation of the results suggests that the change in marginal utility of consumption is $\alpha_{D\&I} = \Gamma_{D\&I}\dfrac{\mu_L}{\mu_{D\&I}} \approx 1.4831$ in case of death or incapacity, and $\alpha_S = \Gamma_S\dfrac{\mu_L}{\mu_S} \approx 0.000002$ in case of a serious disease. Assuming heterogeneous preferences, according to the results in Table 1.11, the corresponding estimates are $\alpha_{D\&I} \approx 0.6078$ and $\alpha_{D\&I} \approx 0.0554$. In both cases, the death or incapacity state

has a larger impact on the marginal utility of consumption than the serious disease state. Moreover, because the empirical frequency of a serious disease estimated at the national level is slightly higher than the empirical frequency of death or incapacity, according to a probability-weighting interpretation, the results suggest an overweighting of the death or incapacity state relative to the serious disease state. In general, the main message remains unchanged, suggesting that the results from the main analysis are not a product of using possible incorrect probabilities.

Table A.5: Alternative Probabilities. Estimation of the Model Assuming Standard Expected Utility ($U(L(m))$ Specified by Equation (1.5))

	Estimate	95 percent bootstrap confidence interval	
$\ln(r)$	-15.5787	-15.6427	-15.5092
$\ln(l)$	16.7195	16.6719	16.7637
σ	0.0047	0.0041	0.0056
LL	-7081.65		

Note: Core sample of 3,164 households.

Table A.6: Alternative Probabilities. Estimation of the Model Assuming Standard Expected Utility. Observed Heterogeneity

		Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_r$	Constant	-15.5254	-15.5493	-15.4921
	Female	-0.0227	-0.0476	-0.0021
	$30 \leq$ Age ≤ 39	-0.0423	-0.0651	-0.0209
	$40 \leq$ Age ≤ 49	-0.0814	-1.7751	-0.0.0582
	$50 \leq$ Age	-1.8073	-2.0429	-0.0041
Mean fitted value	$\ln(r)$	-15.7224	-15.8249	-15.5715
σ		0.0042	0.0037	0.0051
LL		6994.93		

Notes: Core sample of 3,164 households. The loss is assumed to be equal to CRC 18 million.

Appendix A.4. CRRA Utility

In this appendix, I present the complete results of assuming constant relative risk aversion (CRRA) utility. CRRA utility is given by $u(x) = \dfrac{x^{1-\rho}}{1-\rho}$ for $\rho \neq 1$ and $u(x) = \ln(x)$ for $\rho = 1$, where ρ is the coefficient of relative risk aversion. A disadvantage of CRRA utility is that it requires imposing assumptions on the household's prior wealth, a variable that is unobservable from the data. Moreover, there is no consensus in the literature on the wealth that is relevant for the household's decisions. Previous papers often use the monthly or annual income, although Sydnor (2010) argues that a more appropriate measure should be related to permanent income.

I assume the same level of wealth for all households. This assumption is questionable because there surely is heterogeneity in wealth across my core sample. As the relevant level of wealth, I conduct the analysis assuming both an annual level of income and a lifetime level of income. According to the Costa Rican National Household Survey, between 2010 and 2015 the annual per capita household income was CRC 3,933,668. I consider this amount as the annual income. From the same survey, I calculate the present value of the lifetime income to be approximately CRC 80 million. Using this value as a benchmark, the levels of lifetime income I consider in the analysis are CRC 40 million, CRC 80 million, and CRC 120 million.

To estimate the CRRA models, I follow the same steps as in the CARA analysis. Because the power function is defined for positive levels of final wealth, I re-estimate the loss for every level of wealth using the standard expected-utility model. Table A.7 summarizes the results assuming heterogeneity in preferences. Complete results are available in Tables A.8 through A.23, including the homo-

geneous preferences case (the conclusions are preserved). In every case the corresponding loss exceeds the value of the wealth, underlining the high-stakes nature of the lottery. As expected, the magnitude of the results is sensitive to the particular assumption on wealth. However, the main message is unchanged.

Table A.7: Summary of CRRA Utility Results (Assuming Heterogeneity)

	Estimate CARA	Estimate CRRA			
		Annual Income	Lifetime income		
			40 million	80 million	120 million
$\ln(l)$	17.0402	15.3597	17.5230	18.2069	18.6092
	Standard expected utility				
Relative risk aversion (ρ)		2.8400	1.2713	1.0370	0.9276
Absolute risk aversion (r)	$146 \cdot 10^{-9}$	$722 \cdot 10^{-9}$	$318 \cdot 10^{-10}$	$130 \cdot 10^{-10}$	$773 \cdot 10^{-11}$
σ	0.0042	$619 \cdot 10^{-10}$	$141 \cdot 10^{-7}$	$27 \cdot 10^{-5}$	0.0011
LL	-6993.56	-7604.12	-7302.67	-7188.16	-7128.77
	State-dependent utility and probability weighting				
Relative risk aversion (ρ)		0.6925	0.6838	0.6794	0.7093
Absolute risk aversion (r)	$206 \cdot 10^{-9}$	$176 \cdot 10^{-9}$	$171 \cdot 10^{-10}$	$849 \cdot 10^{-11}$	$591 \cdot 10^{-11}$
$\Gamma_{D\&I}$	0.0009	0.0151	0.0031	0.0019	0.0015
Γ_S	0.0001	0.0075	0.0015	0.0010	0.0005
σ	0.0064	0.6665	0.1463	0.0979	0.0438
LL	-6957.39	-6961.57	-6960.03	-6960.12	-6958.63

Note: Core sample of 3,164 households.

Tables A.8 through A.11 present the estimation of the standard expected-utility model under homogeneous preferences. From these estimates, I obtain the loss that I use in the subsequent analysis. Tables A.12 through A.15 present the results of the state-dependent utility and probability weighting model assuming homogeneous preferences. Tables A.16 through A.23 contain the analysis assuming observed heterogeneity, and they report the complete results that Table A.7 summarizes.

Table A.8: Estimation of the Model Assuming Standard Expected Utility. Annual Income

	Estimate	95 percent bootstrap confidence interval	
$\ln(\rho)$	-0.1664	-0.1730	0.1000
$\ln(l)$	15.3597	15.3837	15.3597
σ	0.9744	0.0178	0.9944
LL	-7770.19		

Note: Core sample of 3,164 households.

Table A.9: Estimation of the Model Assuming Standard Expected Utility Lifetime Income: 40 Million

	Estimate	95 percent bootstrap confidence interval	
$\ln(\rho)$	-0.2636	-0.2952	0.1000
$\ln(l)$	17.5230	17.5230	17.5258
σ	0.5146	0.0003	0.6955
LL	-7744.52		

Note: Core sample of 3,164 households.

Table A.10: Estimation of the Model Assuming Standard Expected Utility Lifetime Income: 80 Million

	Estimate	95 percent bootstrap confidence interval	
$\ln(\rho)$	-0.1843	-0.3082	0.0802
$\ln(l)$	18.2069	18.2020	18.2083
σ	0.0765	0.0001	0.5408
LL	-7687.24		

Note: Core sample of 3,164 households.

Table A.11: Estimation of the Model Assuming Standard Expected Utility Lifetime Income: 120 Million

	Estimate	95 percent bootstrap confidence interval	
$\ln(\rho)$	-0.1814	-0.3080	0.0404
$\ln(l)$	18.6092	18.5986	18.6102
σ	0.0530	0.0001	0.4002
LL	-7663.94		

Note: Core sample of 3,164 households.

Table A.12: Estimation of the Model Assuming State-Dependent Utility and Probability Weighting. Annual Income

	Estimate	95 percent bootstrap confidence interval	
$\ln(\rho)$	-0.2681	-0.3270	-0.2189
$\Gamma_{D\&I}$	0.0252	0.0186	0.0260
Γ_S	0.0000	0.0000	0.0067
σ	0.2430	0.1428	0.4620
LL	-7095.48		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 4.68 million.

Table A.13: Estimation of the Model Assuming State-Dependent Utility and Probability Weighting. Lifetime Income: 40 Million

	Estimate	95 percent bootstrap confidence interval	
$\ln(\rho)$	-0.2632	-0.3177	-0.2113
$\Gamma_{D\&I}$	0.0041	0.0034	0.0044
Γ_S	0.0000	0.0000	0.0010
σ	0.0368	0.0194	0.0727
LL	-7091.33		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 40.75 million.

Table A.14: Estimation of the Model Assuming State-Dependent Utility and Probability Weighting. Lifetime Income: 80 Million

	Estimate	95 percent bootstrap confidence interval	
$\ln(\rho)$	-0.2632	-0.3154	-0.2082
$\Gamma_{D\&I}$	0.0024	0.0020	0.0027
Γ_S	0.0000	0.0000	0.0005
σ	0.0216	0.0104	0.0422
LL	-7091.13		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 80.75 million.

Table A.15: Estimation of the Model Assuming State-Dependent Utility and Probability Weighting. Lifetime Income: 120 Million

	Estimate	95 percent bootstrap confidence interval	
$\ln(\rho)$	-0.2629	-0.3168	-0.2098
$\Gamma_{D\&I}$	0.0018	0.0015	0.0020
Γ_S	0.0000	0.0000	0.0004
σ	0.0157	0.0079	0.0321
LL	-7091.05		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 120.75 million.

Table A.16: Estimation of the Model Assuming Standard Expected Utility. Observed Heterogeneity. Annual Income

		Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_\rho$	Constant	0.8661	0.8424	1.0000
	Female	-0.0194	-0.0482	-0.0064
	$30 \leq$ Age ≤ 39	0.6256	-0.1498	0.9990
	$40 \leq$ Age ≤ 49	-0.2201	-0.3022	-0.1046
	$50 \leq$ Age	-0.2189	-0.3046	-0.1082
Mean fitted value	$\ln(\rho)$	1.0438	0.7505	1.3336
σ		$619 \cdot 10^{-10}$	$3 \cdot 10^{-10}$	$636 \cdot 10^{-10}$
LL		-7604.12		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 4.68 million.

Table A.17: Estimation of the Model Assuming Standard Expected Utility. Observed Heterogeneity. Lifetime Income: 40 Million

		Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_\rho$	Constant	0.2775	0.2555	0.2965
	Female	-0.0327	-0.0493	-0.0175
	$30 \leq$ Age ≤ 39	-0.0106	-0.0195	0.0031
	$40 \leq$ Age ≤ 49	-0.0241	-0.0373	-0.0122
	$50 \leq$ Age	-0.0136	-0.0329	0.0032
Mean fitted value	$\ln(\rho)$	0.2401	0.2251	0.2637
σ		$141 \cdot 10^{-7}$	$1 \cdot 10^{-8}$	$23 \cdot 10^{-6}$
LL		-7302.67		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 40.75 million.

Table A.18: Estimation of the Model Assuming Standard Expected Utility. Observed Heterogeneity. Lifetime Income: 80 Million

		Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_\rho$	Constant	0.0689	0.0505	0.0888
	Female	-0.0412	-0.0594	-0.0228
	$30 \leq$ Age ≤ 39	0.0009	-0.0092	0.0114
	$40 \leq$ Age ≤ 49	-0.0083	-0.0208	0.0049
	$50 \leq$ Age	0.0066	-0.0118	0.0249
Mean fitted value	$\ln(\rho)$	0.0364	0.0260	0.0482
σ		$27 \cdot 10^{-5}$	$22 \cdot 10^{-5}$	$34 \cdot 10^{-5}$
LL		-7188.16		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 80.75 million.

Table A.19: Estimation of the Model Assuming Standard Expected Utility. Observed Heterogeneity. Lifetime Income: 120 Million

		Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_\rho$	Constant	-0.0443	-0.0619	-0.0236
	Female	-0.0460	-0.0677	-0.0255
	$30 \leq$ Age ≤ 39	0.0059	-0.0048	0.0167
	$40 \leq$ Age ≤ 49	-0.0016	-0.0150	0.0118
	$50 \leq$ Age	0.0176	-0.0037	0.0361
Mean fitted value	$\ln(\rho)$	-0.0752	-0.0837	-0.0658
σ		0.0011	0.0010	0.0013
LL		-7128.77		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 120.75 million.

Table A.20: Estimation of the Model Assuming State-Dependent Utility and Probability Weighting. Observed Heterogeneity. Annual Income

		Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_\rho$	Constant	-0.3894	-0.5052	-0.2510
	Female	-0.0666	-0.0948	-0.0337
	$30 \leq$ Age ≤ 39	0.0504	0.0220	0.0762
	$40 \leq$ Age ≤ 49	0.1001	0.0645	0.1354
	$50 \leq$ Age	0.1695	0.1173	0.2256
$\boldsymbol{\beta}_{\Gamma_{D\&I}}$	Constant	0.0240	0.0220	0.0271
	Female	-0.0048	-0.0082	-0.0020
	$30 \leq$ Age ≤ 39	0.0002	-0.0030	0.0024
	$40 \leq$ Age ≤ 49	-0.0089	-0.0192	0.0002
	$50 \leq$ Age	-0.0181	-0.0221	-0.0140
$\boldsymbol{\beta}_{\Gamma_S}$	Constant	0.0009	0.0001	0.0043
	Female	0.0009	-0.0027	0.0047
	$30 \leq$ Age ≤ 39	0.0020	-0.0001	0.0058
	$40 \leq$ Age ≤ 49	0.0102	0.0004	0.0213
	$50 \leq$ Age	0.0159	0.0091	0.0218
Mean fitted value	$\ln(\rho)$	-0.3675	-0.4749	-0.2289
	$\Gamma_{D\&I}$	0.0151	0.0125	0.0193
	Γ_S	0.0075	0.0034	0.0106
σ		0.6665	0.1614	1.7787
LL		-6961.57		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 4.68 million.

Table A.21: Estimation of the Model Assuming State-Dependent Utility and Probability Weighting. Observed Heterogeneity. Lifetime Income: 40 Million

		Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_\rho$	Constant	-0.3991	-0.5137	-0.2736
	Female	-0.0551	-0.0786	-0.0317
	$30 \leq$ Age ≤ 39	0.0421	0.0201	0.0659
	$40 \leq$ Age ≤ 49	0.0845	0.0574	0.1130
	$50 \leq$ Age	0.1428	0.1022	0.1860
$\boldsymbol{\beta}_{\Gamma_{D\&I}}$	Constant	0.0049	0.0042	0.0055
	Female	-0.0007	-0.0015	-0.0001
	$30 \leq$ Age ≤ 39	-0.0002	-0.0007	0.0001
	$40 \leq$ Age ≤ 49	-0.0022	-0.0040	0.0005
	$50 \leq$ Age	-0.0040	-0.0047	-0.0029
$\boldsymbol{\beta}_{\Gamma_S}$	Constant	0.0002	0.0000	0.0007
	Female	0.0003	-0.0003	0.0013
	$30 \leq$ Age ≤ 39	0.0003	-0.0002	0.0014
	$40 \leq$ Age ≤ 49	0.0019	0.0001	0.0040
	$50 \leq$ Age	0.0028	0.0013	0.0041
Mean fitted value	$\ln(\rho)$	-0.3801	-0.4907	-0.2458
	$\Gamma_{D\&I}$	0.0031	0.0025	0.0037
	Γ_S	0.0015	0.0007	0.0023
σ		0.1463	0.0339	0.6166
LL		-6960.03		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 40.75 million.

Table A.22: Estimation of the Model Assuming State-Dependent Utility and Probability Weighting. Observed Heterogeneity. Lifetime Income: 80 Million

$\boldsymbol{\beta}_{\rho}$	Constant	-0.4047	-0.5214	-0.2809
	Female	-0.0532	-0.0760	-0.0300
	$30 \leq$ Age ≤ 39	0.0405	0.0187	0.0625
	$40 \leq$ Age ≤ 49	0.0816	0.0540	0.1088
	$50 \leq$ Age	0.1381	0.0964	0.1835
$\boldsymbol{\beta}_{\Gamma_{D\&I}}$	Constant	0.0031	0.0025	0.0036
	Female	-0.0004	-0.0010	0.0000
	$30 \leq$ Age ≤ 39	-0.0002	-0.0007	0.0001
	$40 \leq$ Age ≤ 49	-0.0015	-0.0026	-0.0004
	$50 \leq$ Age	-0.0026	-0.0031	-0.0018
$\boldsymbol{\beta}_{\Gamma_S}$	Constant	0.0001	0.0000	0.0006
	Female	0.0002	-0.0002	0.0010
	$30 \leq$ Age ≤ 39	0.0002	-0.0001	0.0010
	$40 \leq$ Age ≤ 49	0.0012	0.0001	0.0024
	$50 \leq$ Age	0.0017	0.0007	0.0025
Mean	$\ln(\rho)$	-0.3865	-0.4963	-0.2623
fitted	$\Gamma_{D\&I}$	0.0019	0.0015	0.0024
value	Γ_S	0.0010	0.0004	0.0016
σ		0.0979	0.0215	0.3254
LL		-6960.12		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 80.75 million.

Table A.23: Estimation of the Model Assuming State-Dependent Utility and Probability Weighting. Observed Heterogeneity. Lifetime Income: 120 Million

		Estimate	95 percent bootstrap confidence interval	
$\boldsymbol{\beta}_\rho$	Constant	-0.3590	-0.5246	-0.2825
	Female	-0.0501	-0.0746	-0.0298
	$30 \leq$ Age ≤ 39	0.0367	0.0184	0.0619
	$40 \leq$ Age ≤ 49	0.0745	0.0535	0.1067
	$50 \leq$ Age	0.1274	0.0955	0.0955
$\boldsymbol{\beta}_{\Gamma_{D\&I}}$	Constant	0.0022	0.0019	0.0028
	Female	-0.0001	-0.0007	0.0000
	$30 \leq$ Age ≤ 39	-0.0001	-0.0006	0.0001
	$40 \leq$ Age ≤ 49	-0.0011	-0.0020	-0.0003
	$50 \leq$ Age	-0.0021	-0.0024	-0.0014
$\boldsymbol{\beta}_{\Gamma_S}$	Constant	0.0000	0.0000	0.0004
	Female	0.0000	-0.0001	0.0007
	$30 \leq$ Age ≤ 39	0.0000	-0.0001	0.0007
	$40 \leq$ Age ≤ 49	0.0008	0.0000	0.0019
	$50 \leq$ Age	0.0013	0.0005	0.0020
Mean fitted value	$\ln(\rho)$	-0.3435	-0.4999	-0.2635
	$\Gamma_{D\&I}$	0.0015	0.0011	0.0018
	Γ_S	0.0005	0.0003	0.0012
σ		0.0438	0.0160	0.2620
LL		-6958.63		

Notes: Core sample of 3,164 households. The loss is assumed to be about CRC 120.75 million.

APPENDIX B

APPENDIX OF CHAPTER 2: ADVERSE SELECTION AND MORAL HAZARD IN PRIVATE HEALTHCARE WHEN UNIVERSAL HEALTHCARE IS AVAILABLE

Appendix B.1. Individual Student Health Insurance-Distribution of Choices by Copay-and-Deductible Regime

Table B.1: Individual Student Health Insurance-Distribution of Choices by Copay-and-Deductible Regime

Regime	Period	**Option 1** M_1 214.13	**Option 2** M_2 428.27	**Option 3** M_3 856.53	**Option 4** M_4 1284.80	**N**
1	t_1	81.19%	15.13%	2.25%	1.42%	14,556
2	t_2, t_3	81.73%	14.38%	2.28%	1.61%	18,998
3	t_4	80.67%	14.51%	2.34%	2.48%	16,793
4	t_5	84.90%	10.32%	2.60%	2.18%	3,577

Notes: Individual Student Health Insurance sample of 32,187 individuals (53,924 individual-years). For Option j ($j = 1, 2, 3, 4$), M_j represents the maximum insured amount in case of an adverse event. Monetary amounts normalized such that real premium for Option 1 in period t_1 is 1.

Appendix B.2. Comparison of Individuals that Bought Insurance under Different Copay-and-Deductible Regimes

Table B.2: Individual Student Health Insurance-Average Marginal Effects Claim Rates Regression-Random-Effect Logit Model-For Individuals Observed in Multiple Copay-and-Deductible Regimes

	Regimes Where the Individual is Observed					
	1 and 2	2 and 3	3 and 4	1, 2, and 3	2, 3 and 4	All
Option						
1	vs	vs	vs	vs	vs	vs
2	0.016***	0.031***	0.038	0.035***	0.046**	0.104***
	(0.006)	(0.008)	(15.591)	(0.008)	(0.022)	(0.033)
3	0.062**	0.056***	0.004	0.083***	0.200***	0.034
	(0.028)	(0.021)	(1.845)	(0.027)	(0.071)	(0.041)
4	0.113***	0.111***	0.058	0.094**	0.144***	0.059
	(0.037)	(0.028)	(20.106)	(0.037)	(0.039)	(0.055)
Policy						
0% Copay	vs	vs		vs	vs	vs
5% Copay	-0.488***			-0.011		-0.013
	(0.001)			(0.010)		(0.026)
15% Copay					-0.089***	
					(0.015)	
20% Copay			-0.041			
‘			(24.912)			
25% Copay			-0.048			
			(31.807)			
15% copay +		-0.016	vs	-0.006	-0.047***	0.014
3.81 Deductible		(0.010)		(0.440)	(0.015)	(0.090)
N	7,752	6,628	984	5,995	1,190	849

Notes: All regressions include demographic controls (age, age^2, gender, nationality plus interactions) as well as year, month and region fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B.3. Individual Student Health Insurance-Random-Effect Logit Model

Table B.3: Individual Student Health Insurance-Claim Rates Regression-Random-Effect Logit Model

	Baseline	Plan Choice	Deductible Policy	Plan Choice + Deductible Policy
Option				
1		vs		vs
2		1.339***		1.340***
		(0.103)		(0.104)
3		2.253***		2.259***
		(0.186)		(0.187)
4		2.877***		2.853***
		(0.179)		(0.181)
Policy				
0% Copay			vs	vs
5% Copay			-0.563**	-0.528**
			(0.223)	(0.222)
15% Copay			-1.372	-2.245***
			(0.918)	(0.868)
20% Copay			-1.767**	-2.005**
			(0.837)	(0.809)
25% Copay			-2.884***	-2.208***
			(0.719)	(0.715)
15% copay +			-0.556***	-0.558***
3.81 Deductible			(0.212)	(0.211)
Random Parameter Terms				
σ_a	2.046	1.871	2.032	1.867
	(0.101)	(0.105)	(0.102)	(0.105)
ρ	0.560	0.515	0.557	0.514
	(0.024)	(0.028)	(0.025)	(0.028)
N	52,537	52,537	52,537	52,537
N_i	31,319	31,319	31,319	31,319
Interactions	Yes	Yes	Yes	Yes
LL	-4249.128	-4052.357	-4234.111	-4040.425

Notes: All regressions include demographic controls (age, age^2, gender, nationality plus interactions) as well as year, month, and region fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B.4. Collective Student Health Insurance-Random-Effect Logit Model

Table B.4: Collective Student Health Insurance-Claim Rates Regression-Random-Effect Logit Model

	Baseline	Plan Choice	Deductible Policy	Plan Choice + Deductible Policy
Option				
1		vs		vs
2		0.545***		0.454
		(0.104)		(0.108)
3		-0.360		-0.390
		(0.330)		(0.335)
4		0.780***		0.620**
		(0.270)		(0.275)
Policy				
0% Copay			vs	vs
5% Copay			0.177	0.135
			(0.217)	(0.216)
25% Copay			-0.377***	-0.309***
			(0.073)	(0.074)
15% copay +			-1.573***	-1.524***
3.81 Deductible			(0.441)	(0.440)
Random Parameter Terms				
σ_a	1.238	1.229	1.269	1.254
	(0.070)	(0.070)	(0.070)	(0.071)
ρ	0.318	0.315	0.329	0.323
	(0.024)	(0.025)	(0.024)	(0.025)
N	46,902	46,902	46,902	46,902
N_i	25,173	25,173	25,173	25,173
Interactions	Yes	Yes	Yes	Yes
LL	-6293.11	-6276.71	-6270.85	-6260.17

Notes: All regressions include demographic controls (age, age^2, gender, nationality plus interactions) as well as month and region fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B.5. Comparison Individual and Collective Student Health Insurance-Random-Effect Logit Model

Table B.5: Comparison Individual and Collective Student Health Insurance-Claim Rates Regression-Random-Effect Logit Model

	Baseline	Plan Choice	Deductible Policy	Plan Choice + Deductible Policy
Individual		-0.060	0.677***	0.289**
		(0.103)	(0.106)	(0.120)
Option				
1		vs		vs
2		0.601***		0.540***
		(0.087)		(0.090)
3		-0.134		-0.223
		(0.331)		(0.335)
4		1.004***		0.832***
		(0.234)		(0.238)
*Option**Individual				
1		vs		vs
2		0.729***		0.706***
		(0.127)		(0.130)
3		2.228***		2.241***
		(0.370)		(0.374)
4		1.636***		1.798***
		(0.279)		(0.284)
Policy				
0% Copay			vs	vs
5% Copay			-0.110	-0.188
			(0.211)	(0.208)
25% Copay			-2.028***	-1.927***
			(0.434)	(0.431)
15% copay +			-0.376***	-0.288***
3.81 Deductible			(0.075)	(0.074)
*Policy**Individual				
0% Copay			vs	vs
5% Copay			-0.481**	-0.348
			(0.233)	(0.230)
25% Copay			-1.397**	-1.010
			(0.665)	(0.664)
15% copay +			-0.463***	-0.580***
3.81 Deductible			(0.117)	(0.117)
Random Parameter Terms				
σ_a	1.563	1.466	1.588	1.497
	(0.056)	(0.057)	(0.057)	(0.058)
ρ	0.426	0.395	0.434	0.405
	(0.017)	(0.019)	(0.018)	(0.019)
N	101,567	101,567	101,567	101,567
N_i	58,204	58,204	58,204	58,204
Interactions	Yes	Yes	Yes	Yes
LL	-10,758.631	-10,506.828	-10,633.737	-10,408.296

Notes: All regressions include demographic controls (age, age^2, gender, nationality plus interactions) as well as month and region fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B.6. Comparison Individual and Collective Student Health Insurance, Regimes 1 to 3-Random-Effect Logit Model

Table B.6: Comparison Individual and Collective Student Health Insurance, Regimes 1 to 3-Claim Rates Regression-Random-Effect Logit Model

	Baseline	Plan Choice	Deductible Policy	Plan Choice + Deductible Policy
Individual		-0.012	0.673***	0.287**
		(0.104)	(0.106)	(0.120)
Option				
1		vs		vs
2		0.582***		0.543***
		(0.087)		(0.090)
3		-0.169		-0.214
		(0.331)		(0.334)
4		0.948***		0.835***
		(0.234)		(0.237)
*Option**Individual				
1		vs		vs
2		0.662***		0.701***
		(0.127)		(0.130)
3		2.186***		2.228***
		(0.369)		(0.373)
4		1.611***		1.795***
		(0.279)		(0.284)
Policy				
0% Copay			vs	vs
5% Copay			-0.112	-0.191
			(0.211)	(0.208)
15% copay +			-0.378***	-0.290***
3.81 Deductible			(0.075)	(0.074)
*Policy**Individual				
0% Copay			vs	vs
5% Copay			-0.476**	-0.344
			(0.233)	(0.230)
15% copay +			-0.460***	-0.578***
3.81 Deductible			(0.117)	(0.117)
Random Parameter Terms				
σ_a	1.557	1.465	1.583	1.491
	(0.056)	(0.057)	(0.057)	(0.058)
ρ	0.424	0.395	0.432	0.403
	(0.018)	(0.019)	(0.018)	(0.019)
N	96,919	96,919	96,919	96,919
N_i	57,028	57,028	57,028	57,028
Interactions	Yes	Yes	Yes	Yes
LL	-10,633.607	-10,402.639	-10,574.072	-10,348.320

Notes: All regressions include demographic controls (age, age^2, gender, nationality plus interactions) as well as month and region fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.7: Comparison Individual and Collective Student Health Insurance, Regimes 1 to 3-Average Marginal Effects-Random-Effect Logit Model

	Plan Choice (1)		Deductible Policy (2)		Plan Choice + Deductible Policy (3)	
	Individual	Collective	Individual	Collective	Individual	Collective
Option						
1	vs				vs	
2	0.033***	0.012***			0.036***	0.011***
	(0.003)	(0.002)			(0.004)	(0.002)
3	0.073***	-0.003			0.077***	-0.003
	(0.010)	(0.005)			(0.010)	(0.005)
4	0.114***	0.023***			0.125***	0.019***
	(0.013)	(0.008)			(0.013)	(0.007)
Policy						
0% Copay			vs		vs	
5% Copay			-0.016***	-0.002	-0.015***	-0.004
			(0.003)	(0.004)	(0.003)	(0.004)
15% copay +			-0.021***	-0.007***	-0.022***	-0.005***
3.81 Deductible			(0.002)	(0.001)	(0.003)	(0.001)
N	96,919		96,919		96,919	

Notes: All regressions include demographic controls (age, age^2, gender, nationality plus interactions) as well as month and region fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B.7. Identification

Theorem 2.1. *Suppose for any $0 < c_1 < c_2 < 1$ we observe maximum utilizations $m_1^{\max}$ and $m_2^{\max}$. Then the preference parameters (β_0, ϕ_0) are identified.*

We first prove a Lemma that will be used in the proof of the theorem.

Lemma B.2. *Let $f, g : [0,1] \to \mathbb{R}_+$ be functions satisfying:*

$$\lim_{x \to 0^+} \frac{f(x)}{g(x)} = 0$$

$$f(x), g(x) \text{ finite on } (0,1].$$

Then there exists an x^* such that $f(x) < g(x)$ for all $x \in (0, x^*]$.

Proof. Fix any $\epsilon \in (0,1)$. Since $\lim_{x \to 0^+} \frac{f(x)}{g(x)} = 0$, there exists a $\delta(\epsilon)$ such that

$$\left| \frac{f(x)}{g(x)} \right| < \epsilon$$

whenever $|x| < \delta(\epsilon)$. Set $x^* = \frac{1}{2}\delta(\epsilon)$ and consider any $\tilde{x} \in (0, x^*]$. We have $|\tilde{x}| < \delta(\epsilon)$ and hence

$$f(\tilde{x}) = |f(\tilde{x})| < \epsilon |g(\tilde{x})| < |g(\tilde{x})| = g(\tilde{x}),$$

as required. □

Note that we only used $\lim_{x \to 0^+} \frac{f(x)}{g(x)} = 0$, $f(x), g(x) \geq 0$, and $x \in [0,1]$ to establish this lemma.

Now we proceed with the proof of Theorem 2.1

Proof. From equation (2.9), we know that the true value of (β_0, ϕ_0) satisfies the following relation

$$m_1^{\max}\frac{(1-\beta_0^{c_1})^2}{\beta_0^{c_1}} = m_2^{\max}\frac{(1-\beta_0^{c_2})^2}{\beta_0^{c_2}} = (\phi_0 \ln(\beta_0))^2.$$

To that end, consider an arbitrary preference parameter vector (β, ϕ). We will show that there is a unique solution to

$$m_1^{\max}\frac{(1-\beta^{c_1})^2}{\beta^{c_1}} = m_2^{\max}\frac{(1-\beta^{c_2})^2}{\beta^{c_2}} \tag{B.1}$$

for $\beta \in (0,1)$, say β^*. And since the solution is unique it will follow that $(\beta^*, \phi^*) = (\beta_0, \phi_0)$, where $\phi^* = \frac{\sqrt{m_1^{\max}}}{\ln(\beta^*)}$.

Re-write equation (B.1) as follows:

$$\underbrace{\sqrt{m_1^{\max}}\beta^{-\frac{c_1}{2}}(1-\beta^{c_1})}_{L(\beta)} = \underbrace{\sqrt{m_2^{\max}}\beta^{-\frac{c_2}{2}}(1-\beta^{c_2})}_{R(\beta)}$$

We establish the following:

i. $L(\beta)$ is strictly decreasing and continuous in β with $\lim_{\beta\to 0^+} L(\beta) = \infty$ and $\lim_{\beta\to 1^-} L(\beta) = 0$.

ii. $R(\beta)$ is strictly decreasing and continuous in β with $\lim_{\beta\to 0^+} R(\beta) = \infty$ and $\lim_{\beta\to 1^-} R(\beta) = 0$

iii. $L'(1) < R'(1) < 0$

iv. $\lim_{\beta\to 0^+} \frac{L(\beta)}{R(\beta)} = 0$.

v. $L''(\beta), R''(\beta) \geq 0$ for all $\beta \in (0,1)$.

For reference we have:

$$
\begin{aligned}
L'(\beta) &= -\frac{\sqrt{m_1^{\max}}c_1}{2}\left(\beta^{\frac{c_1}{2}-1}+\beta^{\frac{c_1}{2}-1}\right)\\
R'(\beta) &= -\frac{\sqrt{m_2^{\max}}c_2}{2}\left(\beta^{\frac{c_1 2}{2}-1}+\beta^{\frac{c_2}{2}-1}\right)\\
L''(\beta) &= -\frac{\sqrt{m_1^{\max}}c_1}{2}\beta^{\frac{c_1}{2}-2}\left[\frac{c_1}{2}-1-\left(\frac{c_1}{2}+1\right)\beta^{-c_1}\right]\\
R''(\beta) &= -\frac{\sqrt{m_2^{\max}}c_2}{2}\beta^{\frac{c_2}{2}-2}\left[\frac{c_2}{2}-1-\left(\frac{c_2}{2}+1\right)\beta^{-c_2}\right]
\end{aligned}
$$

From the definition of $L(\beta)$ and $R(\beta)$ it is straightforward to see that Part (i) and (ii) hold. Consider Part (iii). We have

$$
\begin{aligned}
L'(1) &= -c_1\sqrt{m_1^{\max}}\\
&= -c_1\frac{\beta_0^{c_1/2}\phi_0|\ln(\beta_0)|}{1-\beta_0^{c_1}}
\end{aligned}
$$

It follows that $L'(1) < R'(1)$ iff

$$
\begin{aligned}
&c_1\frac{\beta_0^{c_1/2}\phi_0|\ln(\beta_0)|}{1-\beta_0^{c_1}} > c_2\frac{\beta_0^{c_2/2}\phi_0|\ln(\beta_0)|}{1-\beta_0^{c_2}} && \iff\\
&c_1\frac{\beta_0^{c_1/2}}{1-\beta_0^{c_1}} > c_2\frac{\beta_0^{c_2/2}}{1-\beta_0^{c_2}}.
\end{aligned}
$$

Define $f(c;\beta_0) = c\frac{\beta_0^{c/2}}{1-\beta_0^c}$. We claim that for any fixed $\beta_0 \in (0,1)$ that $f'(c;\beta_0) < 0$ for all $c \in (0,1)$ and hence $L'(1) < R'(1)$. We have:

$$
f'(c;\beta_0) = \frac{2\beta_0^{c/2}(1-\beta_0^c) + c\beta_0^{c/2}\ln(\beta_0)(1-\beta_0^c) + 2c\beta_0^{3c/2}\ln(\beta_0)}{(1-\beta_0^c)^2},
$$

which is strictly less than 0 iff for all $c \in (0,1)$ and $\beta_0 \in (0,1)$

$$
\begin{aligned}
&2(1-\beta_0^c) + c\ln(\beta_0)(1-\beta_0^c) + 2c\beta_0^c\ln(\beta_0) < 0 && \iff\\
&2(1-\beta_0^c) + c\ln(\beta_0)(1+\beta_0^c) < 0 && \iff\\
&2 < 2\beta_0^c - c\ln(\beta_0)(1+\beta_0^c).
\end{aligned}
$$

For any $\beta_0 \in (0,1)$, the RHS, $g(c;\beta_0) \equiv 2\beta_0^c - c\ln(\beta_0)(1+\beta_0^c)$, is minimized at $c = 0$ with $g(0;\beta_0) = 2$. To establish this claim we show that

$$
g'(c;\beta_0) = -\ln(\beta_0)(1-\beta_0^c + c\beta^c\ln(\beta)) > 0.
$$

$g'(c;\beta_0) \geq 0$ if and only if for any $(c,\beta_0) \in (0,1)\times(0,1)$ we have $(\beta_0^c - c\beta^c\ln(\beta)) \in (0,1)$. To that end fix any $c \in (0,1)$ and define $h : (0,1) \to \mathbb{R}$ by $h(x;c) = x^c(1 - cx^c\ln(x))$. The function $h(x;c)$ is strictly increasing since $h'(x;c) = -2c^2x^{2c-1}\ln(x) > 0$ and $h(1) = 1$. Therefore $h(x;c) \leq 1$. We also have

$$\begin{aligned}
\lim_{x\to 0^+} h(x) &= \lim_{x\to 0^+} x^c(1 - cx^c\ln(x)) \\
&= \lim_{x\to 0^+} x^c \times \left(1 - c\lim_{x\to 0^+} x^c\ln(x)\right) \\
&= 0\times(1-0) && \text{(By L'Hospital's Rule)} \\
&= 0.
\end{aligned}$$

Since $h(x;c)$ is strictly increasing on $(0,1)$ and $\lim_{x\to 0^+} h(x;c) = 0$, $h(x;c) > 0$ for all $x \in (0,1)$. This holds for any $c \in (0,1)$, so that $h(x;c) \in (0,1)$ for all $(c,x) \in (0,1)\times(0,1)$. It follows that $g'(c;\beta_0) > 0$, establishing our claim. Now since $g(c;\beta_0)$ is minimized at $c = 0$ and g is strictly decreasing in c, it follows that for any $c \in (0,1)$ and any $\beta_0 \in (0,1)$ that $f'(c;\beta_0) < 0$ and hence $L'(1) < R'(1)$, as required.

Now consider Part (iv). We have:

$$\begin{aligned}
\lim_{\beta\to 0^+} \frac{L(\beta)}{R(\beta)} &= \sqrt{m_1^{\max}/m_2^{\max}} \lim_{\beta\to 0^+} \beta^{(c_2-c_1)/2}\frac{1-\beta^{c_1}}{1-\beta^{c_2}} \\
&= \sqrt{m_1^{\max}/m_2^{\max}} \times 0 \times 1 \\
&= 0.
\end{aligned}$$

Finally, Part (v) follows since $c_1/2 - 1 < 0$ and $-\left(\frac{c_2}{2}+1\right)\beta^{-c_2} < 0$ for all $\beta \in (0,1)$.

Using Parts (i) – (v) we can now establish the Theorem. By (i), (ii), and (iii) there is an interval $I_1 = (\overline{\eta}, 1)$ such that $L(\beta) > R(\beta)$ for all $\beta \in I_1$. By (i), (ii),

(iv) and Lemma B.2, there is an interval $I_0 = (0, \underline{\eta})$ such that $R(\beta) > L(\beta)$ for all $\beta \in I_0$. By the Intermediate Value Theorem there exists at least one $\beta^* \in (0, 1)$ such that $L(\beta) = R(\beta)$. Uniqueness follows from the fact that $L(\cdot)$ and $R(\cdot)$ are both strictly decreasing and convex (Part (v)). $\square$

APPENDIX C

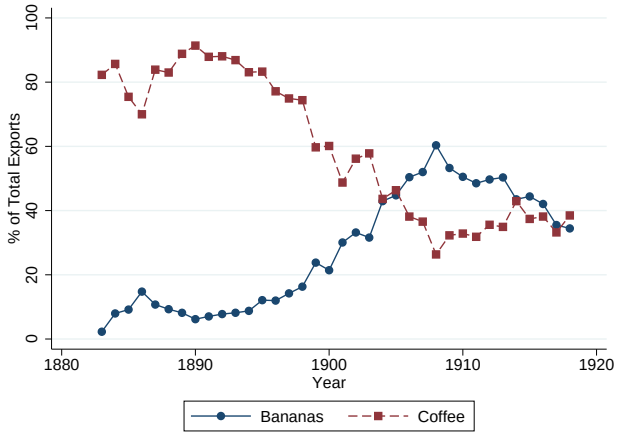
APPENDIX OF CHAPTER 3: MULTINATIONALS AND DEVELOPMENT: EVIDENCE FROM THE UNITED FRUIT COMPANY IN COSTA RICA

Appendix C.1. Historical Details

This section provides more details on the role and decay of the UFCo in Costa Rica and complements the historical background presented in Section 3.2.

Figure C.1 shows how, after 1880 banana production in Costa Rica increased in volume and importance. By 1905 bananas had reached the same place in Costa Rica's exporting value than coffee (Costa Rica's main export product at the time).

Figure C.1: Banana and coffee as percent of total exports, 1883-1918.



Source: Authors' calculations based on the "Statistical Summary, years 1883 to 1910: trade, agriculture, industry" and 1911 to 1918 Costa Rican Statistic Yearbooks.

The railroad construction and the banana activity stimulated population growth in Limón, the province where our chapter restricts attention. Table C.1

shows the dynamics of population growth in Limón using census data from 1883 to 1963, while Table C.2 shows the role of foreigners in these population dynamics.

Table C.1: Population and Growth Rates

	Census									
	1883		1892		1927		1950		1963	
	Pop.	G.R	Pop.	G.R	Pop.	G.R	Pop.	G.R	Pop.	G.R
Limón Province	1,858	-	7,484	16.74	32,278	4.26	41,360	1.08	68,385	3.94
Rest of Costa Rica	180,215	-	235,721	3.03	439,246	1.79	759,515	2.41	1,267,889	4.02

Notes: Pop= Population. G.R= Annual population growth rate (percentage).
Source: Authors' calculations based on 1883, 1892, 1927, 1950, and 1963 Costa Rican Census.

Table C.2: Percentage of Foreigners in the Population

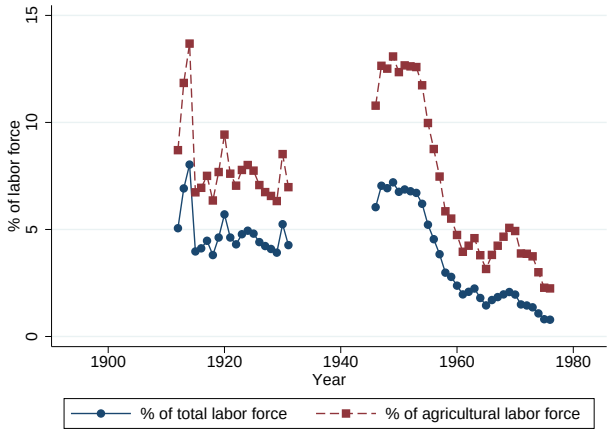
	Census				
	1883	1892	1927	1950	1963
Limón Province	68.51	14.04	68.75	26.84	7.53
Rest of Costa Rica	1.80	2.15	4.67	2.96	2.25

Source: Authors' calculations based on 1883, 1892, 1927, 1950, and 1963 Costa Rican Census.

Figure C.2 illustrates the evolution of UFCo employment in Costa Rica. On average, between 1912 and 1931 the UFCo employed around 7.96% of the total agricultural workers in the country and 4.82% of the entire labor force. Between 1946 and 1976, the numbers were 6.93% and 3.50%, respectively.

The UFCo produced bananas in the Caribbean Coast until 1938, when the Panama disease forced the company to shift operations to the Pacific Coast. Figure C.3 shows how the ports located on the Pacific Coast took a predominant role in the banana exports, while the ports in the Atlantic Coast lost relevance. However, although the enclave structure and the banana production moved to the Pacific Coast, the UFCo kept landholdings in the Caribbean Coast and continued growing alternative products such as cacao and rubber (Viales, 1998). In 1976 the UFCo, now organized under the United Brands name, returned banana production to the Caribbean Coast. By then, new entrants in the banana market prevented the UFCo

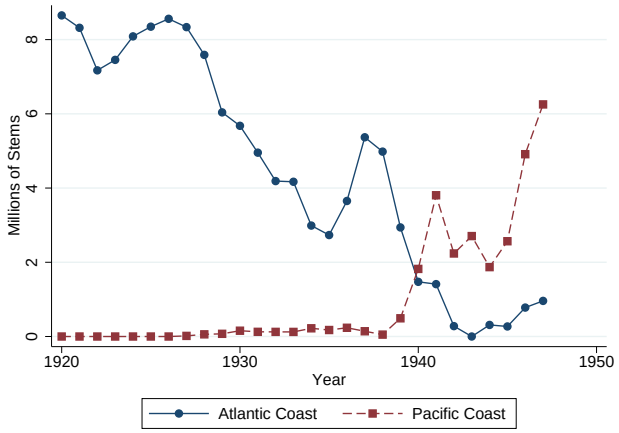
Figure C.2: UFCo employees as percentage of Costa Rican labor force, 1912-1976.



Source: Authors' calculations based on United Fruit Company Medical Department Annual Report for 1912-1931, Ellis (1983) for 1946-1976, and 1892, 1927, 1950, 1963, 1973, and 1984 Costa Rican Census.

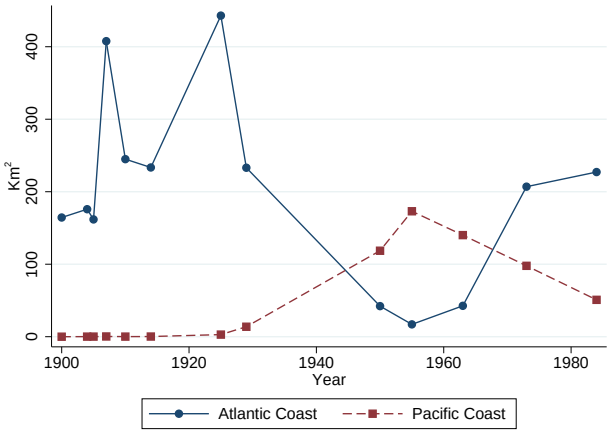
of having the protagonist role and monopoly power that it had at the beginning of the century (Viales and Montero, 2013). Finally, due to labor conflicts, soil exhaustion, increases in production costs, and a corporate strategy that divested in the production process to focus on marketing, the UFCo abandoned banana production in the Pacific Coast in 1984 (Royo, 2009, p. 37). The overall production pattern is evident in Figure C.4, which documents the total land destined to banana grow.

Figure C.3: Banana exports by coast of origin, 1920-1947.



Source: "Statistical Summary, years 1883 to 1910: trade, agriculture, industry," 1911 to 1926 Costa Rican Statistic Yearbooks, and "Export Bulletin 1941-1947."

Figure C.4: Squared kilometers of banana plantations, 1900-1984.



Source: 1900 to 1984 Costa Rica Agricultural Censuses.

Appendix C.2. Unsatisfied Basic Needs (UBN) Index Construction

To specify the set of basic needs that we consider in the chapter and the threshold for attaining those needs, we follow the methodology for Costa Rica proposed by Méndez and Trejos (2004). Méndez and Trejos constructed the index based on information from the 2000 Census. The method can be applied straightforwardly to the 2011 Census, given the similarity of the questions between the 2000 and 2011 censuses (Méndez and Bravo, 2014). To adapt the method to the 1973 and 1984 censuses, we select the components for which similar variables are available in all four censuses. Moreover, for consumption capacity, we adjusted the minimum average years of schooling for the income recipients according to the average years of schooling of the population from 25 to 65 years old. Table C.2 presents the variables from the census that constitute each basic need.

Appendix C.7 shows that if we use the index proposed by Méndez and Trejos for the census where it can be directly applied (2000 and 2011 censuses), the main results of the chapter are preserved.

Table C.3: Definition and Classification of Basic Needs

Dimension	Component	Variable from Census
Housing	House Quality	Household living in a temporary shelter or slum. Household living in a dwelling with waste material in wall, roof or dirt floor. Household living in a dwelling with bad conditions in roof, wall, and floor. simultaneously.
	Overcrowding	Household with more than two persons per room.
Sanitation		Urban household where the sanitary service is connected to ditch, trench, river, estuary, cesspit, or latrine, or without sanitary service. Rural household where the sanitary service is connected to direct connection to ditch, trench, river, estuary, or without sanitary service.

Continued on next page

Table C.3 – continued from previous page

Dimension	Component	Variable from Census
Education	School Attendance	Household with at least one member from 7 to 17 years old not attending school.
	School Achievement	Household with at least one member from 7 to 17 years old attending school regularly, but with a school backwardness higher than 2 years.
Consumption	Consumption Capacity	Household without regular income recipients (employed, pensioners or rentiers) and whose head is 50 years old or older and with: • 3.59 years of schooling or less for Census 1973. • 5 years of schooling or less for Census 1984. • 6 years of schooling or less for Census 2000. • 6.39 years of schooling or less for Census 2011.

Continued on next page

Table C.3 – continued from previous page

Dimension	Component	Variable from Census
Consumption	Consumption Capacity	Urban household with three or more dependents and one income recipient with less than: • 3.59 years of schooling for Census 1973. • 5 years of schooling for Census 1984. • 6 years of schooling for Census 2000. • 6.39 years of schooling for Census 2011.

Continued on next page

Table C.3 – continued from previous page

Dimension	Component	Variable from Census
Consumption	Consumption Capacity	Urban household with three or more dependents and two income recipients whose on average have less than: • 2.59 years of schooling for Census 1973. • 4 years of schooling for Census 1984. • 5 years of schooling for Census 2000. • 5.39 years of schooling for Census 2011.

Continued on next page

Table C.3 – continued from previous page

Dimension	Component	Variable from Census
Consumption	Consumption Capacity	Urban household with three or more dependents and three or more income recipients whose on average have less than: • 1.59 years of schooling for Census 1973. • 3 years of schooling for Census 1984. • 4 years of schooling for Census 2000. • 4.39 years of schooling for Census 2011.

Continued on next page

Table C.3 – continued from previous page

Dimension	Component	Variable from Census
Consumption	Consumption Capacity	Rural household with three or more dependents and one income recipient with less than: • 1.59 years of schooling for Census 1973. • 3 years of schooling for Census 1984. • 4 years of schooling for Census 2000. • 4.39 years of schooling for Census 2011.

Continued on next page

Table C.3 – continued from previous page

Dimension	Component	Variable from Census
Consumption	Consumption Capacity	Rural household with three or more dependents and two income recipients whose on average have less than: • 0.59 years of schooling for Census 1973. • 2 years of schooling for Census 1984. • 3 years of schooling for Census 2000. • 3.39 years of schooling for Census 2011.

Continued on next page

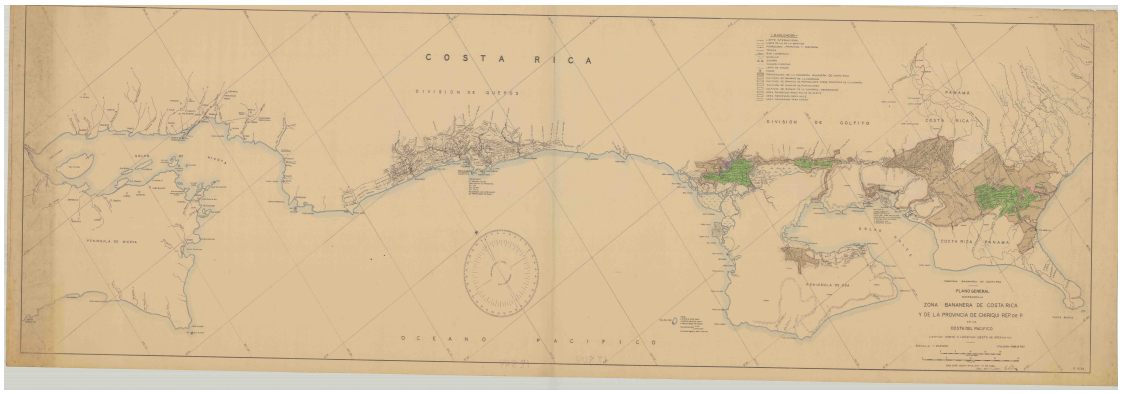
Table C.3 – continued from previous page

Dimension	Component	Variable from Census
Consumption	Consumption Capacity	Rural household with three or more dependents and three or more income recipients whose on average have: • 0 years of schooling for Census 1973. • Less than 1 years of schooling for Census 1984. • Less than 2 years of schooling for Census 2000. • Less than 2.39 years of schooling for Census 2011.

Appendix C.3. Additional Figures

Figure C.5 provides an example of one of the original maps from the National Archives of Costa Rica (*Archivo Nacional de Costa Rica*) that we collected, scanned, and digitized.

Figure C.5: One of the original maps from the National Archives of Costa Rica (*Archivo Nacional de Costa Rica*).



Source: National Archives of Costa Rica. Fondo: Mapa. Signatura: 17849.

Appendix C.4. Details on Government Expenditures

Table C.4 compares government spending per capita between UFCo municipalities and all other rural municipalities in the country, and do not find significant differences. We gathered data on government spending per canton from annual reports from the Comptroller General of the Republic of Costa Rica (*Contraloría General de la República de Costa Rica*) published between 1955 and 1984,[1] and estimate spending per capita.

Table C.4: Comparison of Government Spending per Capita

	Ln Government Spending per Capita	
	(1)	(2)
UFCo	0.004	-0.006
	(0.084)	(0.086)
Adjusted R^2	-0.001	0.349
N	669	669
Clusters	50	50
Year FE	No	Yes

Notes: The unit of observation is the municipality. Robust standard errors, adjusted for clustering by municipality, are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[1] Although the publication was annual, the records on government spending per municipality appear for 15 years between 1951 (the first publication year) and 1984 (when the UFCo ended operations).

Appendix C.5. Falsification Test

Table C.5: Contemporary Household Outcomes: Placebo Test

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Placebo at +2km					
UFCo	0.022	-0.009	0.027	-0.010	0.008	0.031
	(0.034)	(0.019)	(0.018)	(0.030)	(0.040)	(0.066)
	[0.039]	[0.017]	[0.021]	[0.020]	[0.031]	[0.067]
Adjusted R^2	0.098	0.173	0.240	0.014	0.111	0.195
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
	Panel B: Placebo at -2km					
UFCo	-0.030	0.008	-0.006	0.005	-0.008	-0.023
	(0.025)	(0.019)	(0.019)	(0.024)	(0.030)	(0.056)
	[0.031]	[0.019]	[0.019]	[0.027]	[0.029]	[0.054]
Adjusted R^2	0.098	0.173	0.239	0.014	0.111	0.195
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix C.6. Additional Robustness Checks

Our additional robustness checks presented in this section include: alternative specifications of the RD polynomial, ignoring geographic and/or demographic controls, controlling for distance to a railroad, running our regressions at different distances from the boundary, using only subsamples of non-migrants, and comparing the results of subsamples where individuals worked in agricultural versus non-agricultural activities.

C.6.1 Varying Specifications for the RD Polynomial

In our original results, we used a linear polynomial in latitude and longitude. In this section, we test the robustness of our results to different specifications for the RD polynomial. In particular, we use a quadratic polynomial and a linear polynomial in latitude, longitude, and distance to the boundary.

Quadratic Latitude-Longitude Polynomial

Table C.6: Average UFCo Effect-Quadratic Latitude-Longitude Polynomial

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
UFCo	-0.097	-0.013	-0.058	-0.059	-0.122	-0.226
	(0.028)***	(0.019)	(0.022)**	(0.025)**	(0.032)***	(0.060)***
	[0.033]***	[0.015]	[0.012]***	[0.025]**	[0.027]***	[0.055]***
Adjusted R^2	0.102	0.173	0.241	0.015	0.115	0.200
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
Mean	0.176	0.060	0.235	0.200	0.481	0.670

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a quadratic polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.7: Dynamics Across Years-Quadratic Latitude-Longitude Polynomial

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{1973}$	-0.204	-0.277	-0.064	-0.127	-0.225	-0.672
	(0.068)***	(0.080)***	(0.041)	(0.046)***	(0.070)***	(0.164)***
	[0.071]***	[0.078]***	[0.031]**	[0.050]**	[0.054]***	[0.148]***
$UFCo_{1984}$	-0.059	0.016	-0.087	-0.065	-0.079	-0.194
	(0.050)	(0.027)	(0.028)***	(0.036)*	(0.049)	(0.095)**
	[0.035]*	[0.010]*	[0.022]***	[0.030]**	[0.032]**	[0.060]***
$UFCo_{2000}$	-0.084	0.020	-0.062	-0.085	-0.136	-0.210
	(0.033)**	(0.019)	(0.022)***	(0.027)***	(0.038)***	(0.062)***
	[0.032]***	[0.019]	[0.012]***	[0.024]***	[0.032]***	[0.054]***
$UFCo_{2011}$	-0.095	0.021	-0.039	-0.013	-0.099	-0.126
	(0.031)***	(0.017)	(0.036)	(0.037)	(0.039)**	(0.064)*
	[0.034]***	[0.021]	[0.027]	[0.054]	[0.052]*	[0.093]
Adjusted R^2	0.103	0.199	0.241	0.017	0.116	0.207
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
$Mean_{1973}$	0.462	0.353	0.393	0.208	0.777	1.416
$Mean_{1984}$	0.209	0.060	0.362	0.201	0.579	0.832
$Mean_{2000}$	0.145	0.031	0.230	0.178	0.452	0.584
$Mean_{2011}$	0.124	0.018	0.156	0.215	0.402	0.512

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a quadratic polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Linear polynomial in latitude, longitude, and distance to the boundary

Table C.8: Contemporary Household Outcomes: Average UFCo Effect-Linear polynomial in latitude, longitude, and distance to the boundary

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
UFCo	-0.095	-0.016	-0.055	-0.060	-0.123	-0.226
	(0.026)***	(0.017)	(0.022)**	(0.025)**	(0.030)***	(0.056)***
	[0.029]***	[0.014]	[0.018]***	[0.026]**	[0.026]***	[0.051]***
Adjusted R^2	0.102	0.173	0.241	0.015	0.115	0.200
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
Mean	0.176	0.060	0.235	0.200	0.481	0.670

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude, longitude, and distance to the UFCo boundary.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.9: Contemporary Household Outcomes: Dynamics Across Years-Linear polynomial in latitude, longitude, and distance to the boundary

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{1973}$	-0.200	-0.275	-0.064	-0.127	-0.227	-0.666
	(0.066)***	(0.080)***	(0.041)	(0.048)***	(0.071)***	(0.165)***
	[0.069]***	[0.081]***	[0.034]*	[0.045]***	[0.057]***	[0.153]***
$UFCo_{1984}$	-0.055	0.013	-0.084	-0.068	-0.080	-0.195
	(0.048)	(0.028)	(0.028)***	(0.036)*	(0.049)	(0.093)**
	[0.033]*	[0.014]	[0.026]***	[0.030]**	[0.032]**	[0.063]***
$UFCo_{2000}$	-0.079	0.020	-0.057	-0.082	-0.132	-0.199
	(0.032)**	(0.017)	(0.058)***	(0.026)***	(0.036)***	(0.062)***
	[0.029]***	[0.017]	[0.018]***	[0.024]***	[0.031]***	[0.053]***
$UFCo_{2011}$	-0.093	0.020	-0.038	-0.015	-0.101	-0.125
	(0.030)***	(0.016)	(0.030)	(0.037)	(0.038)**	(0.063)**
	[0.033]***	[0.020]	[0.031]	[0.056]	[0.053]*	[0.095]
Adjusted R^2	0.103	0.199	0.241	0.017	0.116	0.206
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
$Mean_{1973}$	0.462	0.353	0.393	0.208	0.777	1.416
$Mean_{1984}$	0.209	0.060	0.362	0.201	0.579	0.832
$Mean_{2000}$	0.145	0.031	0.230	0.178	0.452	0.584
$Mean_{2011}$	0.124	0.018	0.156	0.215	0.402	0.512

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude, longitude, and distance to the UFCo boundary.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.6.2 Ignoring Demographic and Geographic Controls

No Demographic Controls

Table C.10: Average UFCo Effect-No Demographic Controls

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
UFCo	-0.102	-0.014	-0.086	-0.062	-0.142	-0.264
	(0.027)***	(0.017)	(0.025)***	(0.025)**	(0.033)***	(0.063)***
	[0.032]***	[0.014]	[0.014]***	[0.023]***	[0.025]***	[0.055]***
Adjusted R^2	0.071	0.166	0.044	0.003	0.057	0.111
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
Mean	0.176	0.060	0.235	0.200	0.481	0.670

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.11: Contemporary Household Outcomes: Dynamics Across Years-No Demographic Controls

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{1973}$	-0.209	-0.269	-0.098	-0.127	-0.247	-0.703
	(0.066)***	(0.081)***	(0.055)*	(0.052)**	(0.073)***	(0.175)***
	[0.067]***	[0.081]***	[0.052]*	[0.049]**	[0.058]***	[0.160]***
$UFCo_{1984}$	-0.056	0.013	-0.089	-0.068	-0.082	-0.200
	(0.051)	(0.027)	(0.034)***	(0.037)*	(0.057)	(0.109)*
	[0.040]	[0.014]	[0.027]***	[0.030]**	[0.035]**	[0.074]***
$UFCo_{2000}$	-0.089	0.023	-0.092	-0.085	-0.155	-0.244
	(0.031)***	(0.018)	(0.027)***	(0.026)***	(0.039)***	(0.062)***
	[0.032]***	[0.017]	[0.017]***	[0.022]***	[0.034]***	[0.059]***
$UFCo_{2011}$	-0.099	0.023	-0.075	-0.017	-0.123	-0.168
	(0.031)***	(0.016)	(0.030)**	(0.037)	(0.039)***	(0.064)***
	[0.035]***	[0.020]	[0.021]***	[0.053]	[0.047]***	[0.083]**
Adjusted R^2	0.072	0.192	0.044	0.005	0.059	0.117
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
$Mean_{1973}$	0.462	0.353	0.393	0.208	0.777	1.416
$Mean_{1984}$	0.209	0.060	0.362	0.201	0.579	0.832
$Mean_{2000}$	0.145	0.031	0.230	0.178	0.452	0.584
$Mean_{2011}$	0.124	0.018	0.156	0.215	0.402	0.512

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

No Geographic Controls

Table C.12: Average UFCo Effect-No Geographic Control

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
UFCo	-0.103	-0.021	-0.052	-0.062	-0.131	-0.238
	(0.026)***	(0.017)	(0.023)**	(0.024)**	(0.030)***	(0.057)***
	[0.031]***	[0.017]	[0.018]***	[0.024]***	[0.025]***	[0.052]***
Adjusted R^2	0.101	0.168	0.240	0.015	0.115	0.199
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
Mean	0.176	0.060	0.235	0.200	0.481	0.670

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.13: Contemporary Household Outcomes: Dynamics Across Years-No Geographic Controls

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{1973}$	-0.219	-0.288	-0.054	-0.132	-0.247	-0.693
	(0.062)***	(0.079)***	(0.045)	(0.047)***	(0.067)***	(0.158)***
	[0.066]***	[0.078]***	[0.035]	[0.048]***	[0.053]***	[0.146]***
$UFCo_{1984}$	-0.062	0.010	-0.083	-0.088	-0.082	-0.207
	(0.048)	(0.028)	(0.027)***	(0.035)**	(0.046)*	(0.092)**
	[0.035]*	[0.016]	[0.023]***	[0.031]**	[0.033]***	[0.068]***
$UFCo_{2000}$	-0.082	0.018	-0.055	-0.085	-0.136	-0.204
	(0.031)***	(0.018)	(0.023)**	(0.026)***	(0.036)***	(0.059)***
	[0.029]***	[0.017]	[0.018]***	[0.025]***	[0.030]***	[0.051]***
$UFCo_{2011}$	-0.101	0.017	-0.036	-0.020	-0.110	-0.140
	(0.030)***	(0.017)	(0.030)	(0.035)	(0.037)***	(0.062)**
	[0.032]***	[0.020]	[0.031]	[0.050]	[0.049]**	[0.087]
Adjusted R^2	0.103	0.198	0.240	0.017	0.116	0.206
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
$Mean_{1973}$	0.462	0.353	0.393	0.208	0.777	1.416
$Mean_{1984}$	0.209	0.060	0.362	0.201	0.579	0.832
$Mean_{2000}$	0.145	0.031	0.230	0.178	0.452	0.584
$Mean_{2011}$	0.124	0.018	0.156	0.215	0.402	0.512

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

No Demographic or Geographic Controls

Table C.14: Average UFCo Effect-No Demographic or Geographic Controls

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
UFCo	-0.108	-0.018	-0.080	-0.064	-0.148	-0.271
	(0.027)***	(0.017)	(0.025)***	(0.025)**	(0.033)***	(0.064)***
	[0.034]***	[0.016]	[0.012]***	[0.023]***	[0.025]***	[0.057]***
Adjusted R^2	0.070	0.161	0.044	0.003	0.057	0.110
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
Mean	0.176	0.060	0.235	0.200	0.481	0.670

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.15: Dynamics Across Years-No Demographic or Geographic Controls

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{1973}$	-0.225	-0.285	-0.080	-0.133	-0.263	-0.722
	(0.064)***	(0.079)***	(0.058)	(0.050)***	(0.071)***	(0.170)***
	[0.068]***	[0.078]***	[0.050]	[0.051]***	[0.059]***	[0.158]***
$UFCo_{1984}$	-0.062	0.010	-0.085	-0.072	-0.089	-0.209
	(0.051)	(0.028)	(0.035)**	(0.036)**	(0.055)	(0.108)*
	[0.042]	[0.017]	[0.026]***	[0.031]**	[0.037]**	[0.079]***
$UFCo_{2000}$	-0.092	0.022	-0.090	-0.088	-0.159	-0.248
	(0.031)***	(0.018)	(0.028)**	(0.026)***	(0.039)***	(0.062)***
	[0.032]***	[0.017]	[0.016]***	[0.023]***	[0.034]***	[0.057]***
$UFCo_{2011}$	-0.106	0.020	-0.071	-0.022	-0.131	-0.179
	(0.031)***	(0.017)	(0.030)**	(0.034)	(0.038)***	(0.062)***
	[0.034]***	[0.020]	[0.021]***	[0.048]	[0.043]***	[0.075]**
Adjusted R^2	0.072	0.191	0.043	0.005	0.058	0.117
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
$Mean_{1973}$	0.462	0.353	0.393	0.208	0.777	1.416
$Mean_{1984}$	0.209	0.060	0.362	0.201	0.579	0.832
$Mean_{2000}$	0.145	0.031	0.230	0.178	0.452	0.584
$Mean_{2011}$	0.124	0.018	0.156	0.215	0.402	0.512

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.6.3 Distance to a Railroad

In this section, we include as a control variable the nearest distance of each census block centroid to a railroad. Our results suggest that the UFCo effect is not exclusively a product of the provision of railroads.

Table C.16: Average UFCo Effect-Distance to a Railroad

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
UFCo	-0.096	-0.017	-0.057	-0.059	-0.123	-0.228
	(0.026)***	(0.017)	(0.022)**	(0.025)**	(0.031)***	(0.057)***
	[0.029]***	[0.014]	[0.019]***	[0.025]**	[0.027]***	[0.052]***
Adjusted R^2	0.101	0.173	0.240	0.015	0.115	0.200
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
Mean	0.176	0.060	0.235	0.200	0.481	0.670

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include a control for distance to a railroad; geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.17: Dynamics Across Years-Distance to Railroad

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{1973}$	-0.206	-0.277	-0.067	-0.126	-0.227	-0.676
	(0.064)***	(0.080)***	(0.043)	(0.048)***	(0.070)***	(0.163)***
	[0.069]***	[0.079]***	[0.034]**	[0.047]***	[0.055]***	[0.148]***
$UFCo_{1984}$	-0.055	0.014	-0.086	-0.067	-0.081	-0.194
	(0.048)	(0.027)	(0.028)***	(0.036)*	(0.049)*	(0.093)**
	[0.033]*	[0.012]	[0.028]***	[0.029]**	[0.033]**	[0.060]***
$UFCo_{2000}$	-0.080	0.019	-0.057	-0.082	-0.132	-0.200
	(0.032)**	(0.018)	(0.022)**	(0.026)***	(0.036)***	(0.059)***
	[0.028]***	[0.018]	[0.019]***	[0.024]***	[0.031]***	[0.052]***
$UFCo_{2011}$	-0.092	0.021	-0.039	-0.014	-0.102	-0.125
	(0.030)***	(0.017)	(0.030)	(0.037)	(0.038)***	(0.064)*
	[0.031]***	[0.021]	[0.031]	[0.055]	[0.053]*	[0.096]
Adjusted R^2	0.103	0.199	0.241	0.017	0.116	0.206
N	8,786	8,786	8,786	8,786	8,786	8,786
Clusters	200	200	200	200	200	200
$Mean_{1973}$	0.462	0.353	0.393	0.208	0.777	1.416
$Mean_{1984}$	0.209	0.060	0.362	0.201	0.579	0.832
$Mean_{2000}$	0.145	0.031	0.230	0.178	0.452	0.584
$Mean_{2011}$	0.124	0.018	0.156	0.215	0.402	0.512

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include a control for distance to a railroad; geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.6.4 Varying the Maximum Distance to the Boundary

In this section, we present our results restricting the sample to households located within 1 km of the boundary. Our results hold even within these narrow neighborhoods.

Table C.18: Average UFCo Effect-Restricted 1 km

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
UFCo	-0.100	-0.014	-0.085	-0.084	-0.149	-0.284
	(0.034)***	(0.030)	(0.030)***	(0.024)***	(0.046)***	(0.074)***
	[0.022]***	[0.010]	[0.018]***	[0.019]***	[0.024]***	[0.027]***
Adjusted R^2	0.144	0.224	0.274	0.031	0.157	0.269
N	1,937	1,937	1,937	1,937	1,937	1,937
Clusters	44	44	44	44	44	44
Mean	0.176	0.060	0.235	0.200	0.481	0.670

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. The sample is restricted to census block located within 1 km of the UFCo boundary. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.19: Dynamics of the UFCo-Effect Across Years-Restricted 1 km

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{1973}$	-0.123	-0.226	-0.058	-0.089	-0.132	-0.496
	(0.066)*	(0.059)***	(0.053)	(0.033)***	(0.069)*	(0.103)***
	[0.047]***	[0.061]***	[0.048]	[0.029]***	[0.054]**	[0.084]***
$UFCo_{1984}$	0.027	0.025	-0.092	-0.103	-0.063	-0.142
	(0.082)	(0.038)	(0.061)	(0.042)**	(0.072)	(0.129)
	[0.080]	[0.025]	[0.065]	[0.038]***	[0.054]	[0.109]
$UFCo_{2000}$	-0.103	0.002	-0.085	-0.042	-0.121	-0.229
	(0.044)**	(0.030)	(0.029)***	(0.027)	(0.059)**	(0.089)**
	[0.030]***	[0.025]	[0.017]***	[0.034]	[0.043]***	[0.059]***
$UFCo_{2011}$	-0.104	-0.000	-0.089	-0.117	-0.181	-0.310
	(0.039)**	(0.028)	(0.042)**	(0.032)***	(0.054)***	(0.086)***
	[0.023]***	[0.013]	[0.042]**	[0.020]***	[0.052]***	[0.061]***
Adjusted R^2	0.146	0.238	0.273	0.030	0.157	0.270
N	1,937	1,937	1,937	1,937	1,937	1,937
Clusters	44	44	44	44	44	44
$Mean_{1973}$	0.491	0.396	0.455	0.252	0.829	1.595
$Mean_{1984}$	0.265	0.053	0.357	0.186	0.563	0.861
$Mean_{2000}$	0.150	0.037	0.255	0.208	0.497	0.650
$Mean_{2011}$	0.134	0.018	0.164	0.197	0.405	0.513

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. The sample is restricted to census block located within 1 km of the UFCo boundary. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.6.5 Assessing the Impact of Migration

In this section, we run our regressions on subsamples of households where (i) nobody migrated, and (ii) the head of household did not migrate; both within 5 years of each census. Our results persist, indicating that migration is not driving our estimations.

No member migrated within 5 years of the census.

Table C.20: Average UFCo Effect-Any Migrant

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
UFCo	-0.104	-0.004	-0.062	-0.055	-0.135	-0.225
	(0.027)***	(0.015)	(0.025)**	(0.025)**	(0.030)***	(0.052)***
	[0.031]***	[0.015]	[0.023]***	[0.028]**	[0.027]***	[0.049]***
Adjusted R^2	0.077	0.145	0.226	0.012	0.102	0.165
N	6,451	6,451	6,451	6,451	6,451	6,451
Clusters	198	198	198	198	198	198
Mean	0.158	0.050	0.220	0.205	0.466	0.632
P-value for difference	0.49	0.19	0.64	0.78	0.43	0.94

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. The sample is restricted to households whose any of its members is non-migrant. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude. The p-values in the last row are for the test of the hypothesis that the UFCo coefficient is the same than the corresponding in Table 3.3. The p-values are clustered at the census block level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.21: Dynamics of the UFCo-Effect Across Years-Any Migrant

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{1973}$	-0.252	-0.301	-0.070	-0.144	-0.285	-0.767
	(0.067)***	(0.100)***	(0.042)*	(0.035)***	(0.093)***	(0.191)***
	[0.080]***	[0.102]***	[0.031]**	[0.040]***	[0.080]***	[0.183]***
$UFCo_{1984}$	-0.084	-0.000	-0.107	-0.084	-0.131	-0.275
	(0.048)*	(0.029)	(0.033)***	(0.043)*	(0.050)***	(0.094)***
	[0.044]**	[0.019]	[0.026]***	[0.036]**	[0.031]***	[0.062]***
$UFCo_{2000}$	-0.085	0.008	-0.052	-0.098	-0.144	-0.226
	(0.031)***	(0.017)	(0.026)**	(0.030)***	(0.036)***	(0.057)***
	[0.029]***	[0.017]	[0.026]**	[0.028]***	[0.031]***	[0.051]***
$UFCo_{2011}$	-0.110	0.019	-0.053	0.001	-0.113	-0.143
	(0.031)***	(0.016)	(0.033)	(0.035)	(0.037)***	(0.061)**
	[0.036]***	[0.016]	[0.033]	[0.051]	[0.044]**	[0.077]*
Adjusted R^2	0.079	0.168	0.227	0.016	0.102	0.171
N	6,451	6,451	6,451	6,451	6,451	6,451
Clusters	198	198	198	198	198	198
$Mean_{1973}$	0.434	0.360	0.342	0.204	0.758	1.339
$Mean_{1984}$	0.212	0.061	0.369	0.232	0.604	0.875
$Mean_{2000}$	0.135	0.033	0.224	0.179	0.446	0.571
$Mean_{2011}$	0.121	0.018	0.154	0.216	0.400	0.509

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. The sample is restricted to households whose any of its members is non-migrant. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Head-of-household did not migrate within 5 years of the census

Table C.22: Average UFCo Effect-Head Migrant

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
UFCo	-0.107	-0.006	-0.066	-0.062	-0.142	-0.241
	(0.026)***	(0.015)	(0.025)***	(0.025)**	(0.029)***	(0.050)***
	[0.028]***	[0.014]	[0.025]***	[0.031]**	[0.028]***	[0.051]***
Adjusted R^2	0.082	0.157	0.224	0.013	0.104	0.168
N	7,102	7,102	7,102	7,102	7,102	7,102
Clusters	198	198	198	198	198	198
Mean	0.163	0.050	0.227	0.201	0.472	0.641
P-value for difference	0.25	0.22	0.37	0.86	0.18	0.69

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. The sample is restricted to households whose head of household is non-migrant. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude. The p-values in the last row are for the test of the hypothesis that the UFCo coefficient is the same than the corresponding in Table 3.3. The p-values are clustered at the census block level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.23: Dynamics of the UFCo-Effect Across Years-Head Migrant

	Probability of UBN in				Probability	Total number
	Housing	Sanitation	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{1973}$	-0.250	-0.315	-0.076	-0.141	-0.308	-0.782
	(0.075)***	(0.102)***	(0.036)**	(0.041)***	(0.086)***	(0.180)***
	[0.087]***	[0.104]***	[0.026]***	[0.048]***	[0.075]***	[0.177]***
$UFCo_{1984}$	-0.087	-0.002	-0.106	-0.094	-0.133	-0.290
	(0.048)*	(0.027)	(0.033)***	(0.041)**	(0.047)***	(0.092)***
	[0.038]**	[0.018]	[0.024]***	[0.038]**	[0.031]***	[0.062]***
$UFCo_{2000}$	-0.089	0.010	-0.060	-0.104	-0.150	-0.242
	(0.030)***	(0.017)	(0.025)**	(0.028)***	(0.035)***	(0.055)***
	[0.028]***	[0.017]	[0.025]**	[0.027]***	[0.030]***	[0.052]***
$UFCo_{2011}$	-0.112	0.018	-0.055	-0.005	-0.118	-0.155
	(0.030)***	(0.015)	(0.033)*	(0.035)	(0.036)***	(0.061)**
	[0.032]***	[0.015]	[0.036]	[0.055]	[0.047]**	[0.082]*
Adjusted R^2	0.084	0.183	0.224	0.017	0.106	0.174
N	7,102	7,102	7,102	7,102	7,102	7,102
Clusters	198	198	198	198	198	198
$Mean_{1973}$	0.440	0.360	0.351	0.185	0.770	1.336
$Mean_{1984}$	0.213	0.057	0.379	0.219	0.603	0.868
$Mean_{2000}$	0.141	0.031	0.231	0.176	0.451	0.579
$Mean_{2011}$	0.124	0.018	0.158	0.216	0.404	0.515

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. The sample is restricted to households whose head of household is non-migrant. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.6.6 Verifying that results are not driven by persistence of better abilities in agricultural activities

A concern might be that the higher productivity and better infrastructure in the UFCo attracted people who were ex-ante better at growing crops; and that what we are capturing is the persistence of these abilities across generations. Therefore, in this section, we compare the UFCo effect in households that worked in agricultural activities with the effect on households devoted to other non-agricultural activities and find no significant difference in the UFCo effect.

Table C.24 compares our results for households where any member was employed in agricultural activities against all other households, and Table C.25 shows how households whose head works in agricultural activities deliver equivalent estimates to households where the head is employed in other activities.

Table C.24: Average UFCo Effect-Comparison of households where any member is engaged in the agriculture sector versus other economic sectors

		Probability of UBN in				Probability	Total number
		Housing	Sanitation	Education	Consumption	of being poor	of UBN
		(1)	(2)	(3)	(4)	(5)	(6)
Agricultural Sector	UFCo	-0.097	-0.022	-0.052	-0.055	-0.123	-0.225
		(0.028)***	(0.018)	(0.024)**	(0.027)**	(0.033)***	(0.059)***
		[0.027]***	[0.014]	[0.023]**	[0.025]**	[0.024]***	[0.048]***
	Adjusted R^2	0.122	0.192	0.248	0.045	0.152	0.247
	N	6,190	6,190	6,190	6,190	6,190	6,190
	Clusters	200	200	200	200	200	200
	Mean	0.185	0.070	0.267	0.187	0.495	0.709
Non-Agricultural Sector	UFCo	-0.094	0.002	-0.076	-0.065	-0.122	-0.233
		(0.037)**	(0.024)	(0.031)**	(0.049)	(0.052)**	(0.091)**
		[0.044]**	[0.026]	[0.023]***	[0.018]***	[0.034]***	[0.072]***
	Adjusted R^2	0.052	0.091	0.171	0.020	0.043	0.069
	N	2,596	2,596	2,596	2,596	2,596	2,596
	Clusters	193	193	193	193	193	193
	Mean	0.153	0.037	0.159	0.229	0.449	0.578
P-value for difference		0.94	0.32	0.48	0.85	0.98	0.93

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude. The p-values in the last row are for the test of the hypothesis that the UFCo coefficient is the same between the two groups. The p-values are clustered at the census block level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.25: Contemporary Household Outcomes: Average UFCo Effect-Comparison of households where head of household is engaged in the agriculture sector versus other economic sectors

		Probability of UBN in				Probability	Total number
		Housing	Sanitation	Education	Consumption	of being poor	of UBN
		(1)	(2)	(3)	(4)	(5)	(6)
Agricultural Sector	UFCo	-0.083	-0.025	-0.043	-0.039	-0.103	-0.191
		(0.030)***	(0.021)	(0.027)	(0.030)	(0.036)***	(0.065)***
		[0.025]***	[0.015]*	[0.029]	[0.025]	[0.030]***	[0.061]***
	Adjusted R^2	0.128	0.200	0.255	0.045	0.065	0.255
	N	5,337	5,337	5,337	5,337	5,337	5,337
	Clusters	200	200	200	200	200	200
	Mean	0.182	0.073	0.258	0.194	0.490	0.708
Non-Agricultural Sector	UFCo	-0.120	0.000	-0.086	-0.092	-0.161	-0.299
		(0.033)***	(0.017)	(0.029)***	(0.040)**	(0.039)***	(0.064)***
		[0.044]***	[0.020]	[0.021]***	[0.025]***	[0.019]***	[0.054]***
	Adjusted R^2	0.066	0.091	0.209	0.013	0.066	0.104
	N	3,449	3,449	3,449	3,449	3,449	3,449
	Clusters	197	197	197	197	197	197
	Mean	0.166	0.039	0.200	0.208	0.467	0.612
P-value for difference		0.31	0.21	0.24	0.27	0.23	0.15

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude. The p-values in the last row are for the test of the hypothesis that the UFCo coefficient is the same between the two groups. The p-values are clustered at the census block level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix C.7. Méndez & Trejos Index

In this section, we re-estimate equations (3.1) and (3.2) using the Unsatisfied Basic Needs (UBN) originally proposed by Méndez and Trejos (2004) for the 2000 and 2011 censuses. We find that our main message is unchanged.

Table C.26: Average UFCo Effect-Méndez & Trejos Index

	Probability of UBN in				Probability	Total number
	Housing	Health	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
UFCo	-0.088	-0.031	-0.057	-0.020	-0.109	-0.197
	(0.030)***	(0.051)	(0.026)**	(0.019)	(0.043)**	(0.077)**
	[0.033]***	[0.034]	[0.028]**	[0.014]	[0.034]***	[0.069]***
Adjusted R^2	0.020	0.025	0.044	0.025	0.075	0.090
N	6,623	6,623	6,623	6,623	6,623	6,623
Clusters	160	160	160	160	160	160
Mean	0.178	0.132	0.180	0.132	0.433	0.622

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.27: Dynamics Across Years-Méndez & Trejos Index

	Probability of UBN in				Probability	Total number
	Housing	Health	Education	Consumption	of being poor	of UBN
	(1)	(2)	(3)	(4)	(5)	(6)
$UFCo_{2000}$	-0.081	-0.022	-0.069	-0.038	-0.110	-0.210
	(0.036)**	(0.067)	(0.025)***	(0.022)*	(0.052)**	(0.102)**
	[0.035]**	[0.053]	[0.025]***	[0.016]**	[0.044]**	[0.084]**
$UFCo_{2011}$	-0.094	-0.039	-0.047	-0.005	-0.109	-0.186
	(0.032)***	(0.052)	(0.033)	(0.022)	(0.045)**	(0.074)**
	[0.037]***	[0.035]	[0.035]	[0.020]	[0.039]***	[0.076]**
Adjusted R^2	0.020	0.025	0.146	0.025	0.075	0.090
N	6,623	6,623	6,623	6,623	6,623	6,623
Clusters	160	160	160	160	160	160
$Mean_{2000}$	0.164	0.172	0.230	0.178	0.511	0.744
$Mean_{2011}$	0.128	0.101	0.156	0.099	0.365	0.484

Notes: UBN= Unsatisfied Basic Need. The unit of observation is the household. Robust standard errors, adjusted for clustering by census block, are in parentheses. Conley standard errors are in brackets. All regressions include geographic controls for slope, elevation, and temperature; demographic controls for the number of adults, children, and infants in the household; census fixed effects, and a linear polynomial in latitude and longitude.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix C.8. Luminosity Data

Table C.28: Luminosity data

	Log (0.01 + Light)	Log (0.01 + Per Capita Light)
	(1)	(2)
UFCo	0.342	0.215
	(0.035)***	(0.046)***
	[0.072]***	[0.059]***
Adjusted R^2	0.463	0.122
N	6,154	2,210

Notes: The unit of observation is 1x1 km grid cells located within 5 km of UFCo boundary. Robust standard errors are in parentheses. Conley standard errors are in brackets. All regressions include year fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

BIBLIOGRAPHY

Aitken, B. J. and Harrison, A. E. (1999). Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela. *American Economic Review*, 89(3):605–618.

Alfaro, L., Chanda, A., Kalemli-Ozcan, S., and Sayek, S. (2003). FDI Spillovers, Financial Markets, and Economic Development .

Alfaro, L. and Charlton, A. (2007). Growth and the Quality of Foreign Direct Investment: Is All FDI Equal?

Alvarado, G. and Cárdenes, G. (2016). Geology and Tectonic Geomorphology of Costa Rica: A Natural History Approach. In Kappelle, M., editor, *Costa Rican Ecosystems*, pages 30–63. University of Chicago Press.

ANCR (1904). National Archives of Costa Rica (*Archivo Nacional de Costa Rica*). *Serie Congreso, No. 3068, f.1.*

Armendáriz, B. and Larraín B., F. (2017). *The Economics of Contemporary Latin America*. Economics/Latin American Studies. MIT Press.

Bajari, P., Dalton, C., Hong, H., and Khwaja, A. (2014). Moral Hazard, Adverse Selection, and Health Expenditures: A Semiparametric Analysis. *The RAND Journal of Economics*, 45(4):747–763.

Banerjee, A. and Iyer, L. (2005). History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. *American Economic Review*, 95(4):1190–1213.

Barseghyan, L., Molinari, F., O'Donoghue, T., and Teitelbaum, J. C. (2013). The Nature of Risk Preferences: Evidence from Insurance Choices. *American Economic Review*, 103(6):2499–2529.

Barseghyan, L., Molinari, F., O'Donoghue, T., and Teitelbaum, J. C. (2018). Estimating Risk Preferences in the Field. *Journal of Economic Literature*, 56(2):501–64.

Becker, S. O., Boeckh, K., Hainz, C., and Woessmann, L. (2016). The Empire Is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy. *Economic Journal*, 126(590):40–74.

Blomstrom, M. (1986). Foreign Investment and Productive Efficiency: The Case of Mexico. *Journal of Industrial Economics*, 35(1):97–110.

Blomstrom, M. and Wolff, E. N. (1989). Multinational Corporations and Productivity Convergence in Mexico. Working Paper 3141, National Bureau of Economic Research.

Bonilla, R. and Rosero, L. (2008). Rasterización de los censos de población y vivienda de 1973, 1984 y 2000. In Rodríguez, A. and Saborío, M., editors, *Lo rural es diverso: evidencia para el caso de Costa Rica*, pages 135–151. San José, Costa Rica.

Borensztein, E., De Gregorio, J., and Lee, J.-W. (1995). How Does Foreign Direct Investment Affect Economic Growth? Working Paper 5057, National Bureau of Economic Research.

Bucheli, M. (2005). *Bananas and Business: The United Fruit Company in Colombia, 1899-2000*. New York University Press.

Caja Costarricense de Seguro Social (CCSS) (2014). *Reglamento para el otorgamiento de incapacidades y licencias a los beneficiarios del Seguro de Salud*. Caja Costarricense de Seguro Social. Gerencia Médica.

Camerer, C. F. and Ho, T.-H. (1994). Violations of the Betweenness Axiom and Nonlinearity in Probability. *Journal of Risk and Uncertainty*, 8(2):167–196.

Cardon, J. H. and Hendel, I. (2001). Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey. *The RAND Journal of Economics*, 32(3):408–427.

Casey, J. (1979). *Limón 1880-1940*. Editorial Costa Rica, San José.

Castillo-Serrano, D. (1998). *Afro-Caribbean Schools in Costa Rica, 1934-1948*. PhD thesis, The University of Texas at Austin.

Centro Centroamericano de Población de la Universidad de Costa Rica (CCP) (2012). *Producto 2. Tablas de vida de Costa Rica 2005-2010. Y estimación revisada de las tablas de vida de 1950 a 2005.* San José, Costa Rica.

Cerdas Albertazzi, A. L. (1993). El surgimiento del enclave bananero en el Pacífico Sur. *Revista de Historia*, (28):117–159.

Chen, X. and Nordhaus, W. D. (2011). Using Luminosity Data as a Proxy for Economic Statistics. *Proceedings of the National Academy of Sciences*.

Chetty, R. (2006). A New Method of Estimating Risk Aversion. *American Economic Review*, 96(5):1821–1834.

Chiappori, P. and Salanié, B. (2000). Testing for Asymmetric Information in Insurance Markets. *Journal of Political Economy*, 108(1):56–78.

Chiappori, P.-A., Gandhi, A., Salanié, B., and Salanié, F. (2012). From Aggregate Betting Data to Individual Risk Preferences. Working paper.

Chiappori, P.-A., Samphantharak, K., Schulhofer-Wohl, S., and Townsend, R. M. (2014). Heterogeneity and Risk Sharing in Village Economies. *Quantitative Economics*, 5(1):1–27.

Chomsky, A. (1996). *West Indian Workers and the United Fruit Company in Costa Rica, 1870-1940*. Louisiana State University Press.

Cicchetti, C. J. and Dubin, J. A. (1994). A Microeconometric Analysis of Risk Aversion and the Decision to Self-Insure. *Journal of Political Economy*, 102(1):169–186.

Cohen, A. and Einav, L. (2007a). Estimating Risk Preferences from Deductible Choice. *American Economic Review*, 97(3):745–788.

Cohen, A. and Einav, L. (2007b). Estimating Risk Preferences from Deductible Choice. *American Economic Review*, 97(3):745–788.

Conley, T. (1999). GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics*, 92(1):1–45.

Cutler, D. M. and Zeckhauser, R. J. (2000). The Anatomy of Health Insurance. In Culyer, A. J. and Newhouse, J. P., editors, *Handbook of Health Economics*, volume 1A, pages 563 – 643. Amsterdam: North-Holland.

Cutter, V. M. (1926). Caribbean Tropics in Commercial Transition. *Economic Geography*, 2(4):494–507.

Deeks, W. E. (1924). Activities of the Medical Department of the United Fruit Company. In *Proceedings of the International Conference on Health Problems in Tropical America Held at Kingston, Jamaica, B.W.I., July 22 to August 1, 1924: By Invitation of the Medical Department, United Fruit Company*, Boston, pages 1006–1010.

Deeks, W. E. (1928). Program of Malaria Control. In *United Fruit Company. Medical Department Annual Report*, Boston, pages 94–104.

Dell, M. (2010). The Persistent Effects of Peru's Mining Mita. *Econometrica*, 78(6):1863–1903.

Dell, M., Lane, N., and Querubin, P. (2015). State Capacity, Local Governance, and Economic Development in Vietnam.

Dell, M. and Olken, B. (2017). The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java.

Edelman, M. (1992). *The Logic of the Latifundio: The Large Estates of Northwestern Costa Rica Since the Late Nineteenth Century.* Stanford University Press.

Edwards, R. D. (2008). Health Risk and Portfolio Choice. *Journal of Business & Economic Statistics*, 26(4):472–485.

Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., and Cullen, M. R. (2013). Selection on Moral Hazard in Health Insurance. *American Economic Review*, 103(1):178–219.

Ellis, F. (1983). *Las transnacionales del banano en Centroamérica.* Editorial Universitaria Centroamericana (EDUCA).

Ellis, R. P. (1986). Rational Behavior in the Presence of Coverage Ceilings and Deductibles. *The RAND Journal of Economics*, 17(2):158–175.

Evans, W. N. and Viscusi, W. K. (1991). Estimation of State-Dependent Utility Functions Using Survey Data. *The Review of Economics and Statistics*, 73(1):94–104.

Fallas-Paniagua, M. (2013). Cartografía censal en Costa Rica: de lo análogo a lo digital. *Revista Geográfica de América Central*, 1(50):113–140.

Fang, H., Keane, M. P., and Silverman, D. (2008). Sources of Advantageous Selection: Evidence from the Medigap Insurance Market. *Journal of Political Economy*, 116(2):303–350.

Feres, J. C. and Mancero, X. (2001). El método de las necesidades básicas insatisfechas (NBI) y sus aplicaciones en América Latina. Series Estudios Estadísticos y Prospectivos, CEPAL, UN.

Finkelstein, A., Luttmer, E. F. P., and Notowidigdo, M. J. (2009). Approaches to Estimating the Health State Dependence of the Utility Function. *American Economic Review*, 99(2):116–21.

Finkelstein, A., Luttmer, E. F. P., and Notowidigdo, M. J. (2013). What Good Is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption. *Journal of the European Economic Association*, 11:221–258.

Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, 80(1):27–38.

Gelman, A. and Imbens, G. (2017). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*.

Gennaioli, N. and Rainer, I. (2007). The Modern Impact of Precolonial Centralization in Africa. *Journal of Economic Growth*, 12(3):185–234.

Grunow, M. and Nuscheler, R. (2014). Public And Private Health Insurance In Germany: The Ignored Risk Selection Problem. *Health Economics*, 23(6):670–687.

Halek, M. and Eisenhauer, J. G. (2001). Demography of Risk Aversion. *The Journal of Risk and Insurance*, 68(1):1–24.

Hall, C. (1978). *El café y el desarrollo histórico-geográfico de Costa Rica*. Editorial Costa Rica., San José.

Handel, B. R. (2013a). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *American Economic Review*, 103(7):2643–82.

Handel, B. R. (2013b). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *American Economic Review*, 103(7):2643–82.

Harpelle, R. (2001). *West Indians of Costa Rica: Race, Class, and the Integration of an Ethnic Minority*. Mcgill-Queen's Studies in Ethnic History. McGill-Queen's University Press.

Harrison, A. and Rodríguez-Clare, A. (2009). Trade, Foreign Investment, and Industrial Policy for Developing Countries. Working Paper 15261, National Bureau of Economic Research.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review*, 102(2):994–1028.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very High Resolution Interpolated Climate Surfaces For Global Land Areas. *International Journal of Climatology*, 25(15):1965–1978.

Hodler, R. and Raschky, P. A. (2014). Regional Favoritism. *The Quarterly Journal of Economics*, 129(2):995–1033.

Iyer, L. (2010). Direct versus Indirect Colonial Rule in India: Long-Term Consequences. *The Review of Economics and Statistics*, 92(4):693–713.

Jenkins, V. S. (2000). *Bananas: An American History*. Smithsonian Institute.

Jones, C. F. and Morrison, P. C. (1952). Evolution of the Banana Industry of Costa Rica. *Economic Geography*, 28(1):1–19.

Jones, G. (2005). *Multinationals and Global Capitalism: From the Nineteenth to the Twenty First Century*. Oxford University Press, Oxford.

Jullien, B. and Salanie, B. (2000). Estimating Preferences under Risk: The Case of Racetrack Bettors. *Journal of Political Economy*, 108(3):503–530.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–91.

Keane, M. and Stavrunova, O. (2016). Adverse selection, moral hazard and the demand for medigap insurance. *Journal of Econometrics*, 190(1):62 – 78.

Keith, M. C. (1886). *Mr. Keith's letter to the Costa Rica bondholders. June, 1886*. London.

Kelsey, D. (1992). Risk and Risk Aversion for State-Dependent Utility. *Theory and Decision*, 33(1):71–82.

Kepner, C. D. (1936). *Social Aspects of the Banana Industry*. Studies in History, Economics, and Public Law. Columbia University Press.

Kepner, C. D. and Soothill, J. H. (1935). *The Banana Empire: A Case Study of Economic Imperialism*. New York. Vanguard Press.

King, G. and Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, 9:137–163.

Liebenberg, A. P., Carson, J. M., and Dumm, R. E. (2012). A Dynamic Analysis of the Demand for Life Insurance. *Journal of Risk and Insurance*, 79(3):619–644.

Lillard, L. A. and Weiss, Y. (1997). Uncertain Health and Survival: Effects on End-of-Life Consumption. *Journal of Business & Economic Statistics*, 15(2):254–268.

Lipsey, R. E. (2002). Home and Host Country Effects of FDI. Working Paper 9293, National Bureau of Economic Research.

Lowes, S. and Montero, E. (2016). Blood Rubber: The Effects of Labor Coercion on Institutions and Culture in the Democratic Republic of Congo. Job Market Paper 2 (Economic History; Political Economy).

López, J. A. (1930). The Lower Class of Tropical American Patients: A Psychological Sketch. In *United Fruit Company. Medical Department Annual Report*, Boston, pages 163–167.

Madrian, B. C. and Shea, D. F. (2001). The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. *The Quarterly Journal of Economics*, 116(4):1149–1187.

Marton, J., Yelowitz, A., and Talbert, J. C. (2017). Medicaid Program Choice, Inertia and Adverse Selection. *Journal of Health Economics*, 56:292 – 316.

May, S. and Lasso, G. (1958). *The United Fruit Company in Latin America*. Number no. 7 in American business abroad. National Planning Association.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-Colonial Ethnic Institutions and Contemporary c Development. *Econometrica*, 81(1):113–152.

Michalopoulos, S. and Papaioannou, E. (2014). National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics*, 129(1):151–213.

Mirabile, L. and Derdak, T. (1990). United Brands Company. In Mirabile, L. and Derdak, T., editors, *International Directory of Company Histories.*, volume 2, pages 595–597. Chicago: St. James Press.

Méndez, F. and Bravo, O. (2014). Costa Rica: Mapas de pobreza 2011. In Instituto Nacional de Estadística y Censos (INEC), editor, *Costa Rica a la luz del Censo 2011*, page 9–39.

Méndez, F. and Trejos, J. D. (2004). Costa Rica: Un mapa de carencias críticas para el año 2000. In Instituto Nacional de Estadística y Censos (INEC), editor, *Costa Rica a la luz del Censo 2000*, page 205–233.

Molina, I. (2017). *Estadísticas de financiamiento, salarios docentes, matrícula, cobertura y graduación en la educación costarricense: Una contribución documental (1827-2016).* Vicerrectoría de Investigación : Centro de Investigaciones Históricas de América Central, San José, Costa Rica.

Morgan, L. (1993). *Community Participation in Health: The Politics of Primary Care in Costa Rica.* Cambridge Studies in Medical Anthropology. Cambridge University Press.

Nunn, N. (2008). The Long-term Effects of Africa's Slave Trades. *The Quarterly Journal of Economics*, 123(1):139–176.

Oficina Central de Estadística (1868). *Censo de Población 1864.* Imprenta Nacional.

Olivella, P. and Vera-Hernández, M. (2013). Testing for Asymmetric Information in Private Health Insurance. *The Economic Journal*, 123(567):96–130.

Organisation for Economic Co-operation and Development (OECD) (2017). *OECD Reviews of Health Systems: Costa Rica 2017*. OECD Publishing, Paris.

Outreville, J. F. (2013). The Relationship Between Insurance and Economic Development: 85 Empirical Papers for a Review of the Literature. *Risk Management and Insurance Review*, 16(1):71–122.

Outreville, J. F. (2014). Risk Aversion, Risk Behavior, and Demand for Insurance: A Survey. *Journal of Insurance Issues*, 37(2):158–186.

Palmer, J. T. (1932). The Banana in Caribbean Trade. *Economic Geography*, 8(3):262–273.

Polyakova, M. (2016a). Regulation of insurance with adverse selection and switching costs: Evidence from medicare part d. *American Economic Journal: Applied Economics*, 8(3):165–95.

Polyakova, M. (2016b). Risk Selection and Heterogeneous Preferences in Health Insurance Markets with a Public Option. *Journal of Health Economics*, 49:153 – 168.

Powell, D. and Goldman, D. (2016). Disentangling moral hazard and adverse selection in private health insurance. Working Paper 21858, National Bureau of Economic Research.

Royo, A. (2009). *Crisis de dependencia en la Zona Sur: Desarrollo agrario y migraciones internas en el Cantón de Osa 1973-2000*. Editorial UCR, Universidad de Costa Rica, San José.

Salisbury, E. I. (1930). The Malaria Index in the Costa Rican Division as Affected by Control Measures. In *United Fruit Company. Medical Department Annual Report*, Boston, pages 33–39.

Sandner, G. (1962). *La colonización agrícola de Costa Rica*, volume 1. Instituto Geográfico de Costa Rica.

Sanou, O. and Quesada, F. (1998). Orden, progreso y civilización (1871-1914). Transformaciones urbanas y arquitectónicas. In Fonseca, E. and Garnier, J., editors, *Historia de la Arquitectura en Costa Rica*, pages 219–317. San José, Fundación Museos del Banco Central de Costa Rica.

Sloan, F. A., Viscusi, W. K., Chesson, H. W., Conover, C. J., and Whetten-Goldstein, K. (1998). Alternative Approaches to Valuing Intangible Health Losses: The Evidence for Multiple Sclerosis. *Journal of Health Economics*, 17(4):475 – 497.

Smarzynska Javorcik, B. (2004). Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages. *American Economic Review*, 94(3):605–627.

Snowberg, E. and Wolfers, J. (2010). Explaining the Favorite-Long Shot Bias: Is it Risk-Love or Misperceptions? *Journal of Political Economy*, 118(4):723–746.

Soley, T. (1940). *Compendio de historia económica y hacendaria de Costa Rica.* Serie Escolar "Costa Rica" N° 2. Editorial Soley y Valverde.

Sydnor, J. (2010). (Over)insuring Modest Risks. *American Economic Journal: Applied Economics*, 2(4):177–99.

UFCo (1916). *United Fruit Company. Medical Department Annual Report.* Boston.

UFCo (1921). *United Fruit Company. Medical Department Annual Report*. Boston.

UFCo (1929). *United Fruit Company. Medical Department Annual Report*. Boston.

Viales, R. (1998). *Después del enclave 1927-1950: Un estudio de la región atlántica costarricense*. Colección Nueva historia. Editorial de la Universidad de Costa Rica.

Viales, R. (2012). La colonización agrícola del Atlántico (Caribe) costarricense entre 1870 y 1930. El peso de la política agraria Liberal y de las diversas formas de apropiación territorial. *Anuario de Estudios Centroamericanos*, 27(2):57–100.

Viales, R. and Montero, A. (2013). El impacto ambiental del cultivo del banano en la región Atlántico/Caribe de Costa Rica. Limón durante su segundo ciclo bananero (1960-2010). In Viales, R., editor, *La conformación histórica de la región Atlántico/Caribe costarricense: (re)interpretaciones sobre su trayectoria entre el siglo XVI y el siglo XXI*, pages 477–521. San José: Editorial Nuevas Perspectivas.

Viscusi, W. K. and Evans, W. N. (1990). Utility Functions That Depend on Health Status: Estimates and Economic Implications. *The American Economic Review*, 80(3):353–374.

Wagstaff, A., Flores, G., Hsu, J., Smitz, M.-F., Chepynoga, K., Buisman, L. R., van Wilgenburg, K., and Eozenou, P. (2018). Progress on Catastrophic Health Spending in 133 Countries: a Retrospective Observational Study. *The Lancet Global Health*, 6(2):e169 – e179.

Wakker, P. P. (2008). Explaining the Characteristics of the Power (CRRA) Utility Family. *Health Economics*, 17(12):1329–1344.

Wiley, J. (2008). *The Banana: Empires, Trade Wars, and Globalization*. At Table. University of Nebraska Press.

Wilson, C. (1947). *Empire in Green and Gold: The Story of the American Banana Trade*. Henry Holt and Company.

Wu, G. and Gonzalez, R. (1996). Curvature of the Probability Weighting Function. *Management Science*, 42(12):1676–1690.

Xu, B. (2000). Multinational Enterprises, Technology Diffusion, and Host Country Productivity Growth. *Journal of Development Economics*, 62(2):477 – 493.

Zietz, E. N. (2003). An Examination of the Demand for Life Insurance. *Risk Management and Insurance Review*, 6(2):159–191.