Three Essays in Development Economics

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

DEISY JOHANNA FAJARDO-GONZALEZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

MARC F. BELLEMARE AND PAUL W. GLEWWE, ADVISERS

JUNE, 2017

Acknowledgements

I am deeply grateful for valuable guidance and support from Marc Bellemare and Paul Glewwe, and for several years of excellent academic advising. I am indebted to the members of my thesis committee, Deborah Levison and Jason Kerwin, who provided excellent comments on this work throughout the entire process. I also thank Joe Ritter for his feedback and encouragement.

For the published work on inequality of opportunity in adult health in Colombia, I specially thank Stephen Jenkins and two anonymous reviewers, whose very detailed and helpful comments led to a substantially improved manuscript. I am also grateful to Elizabeth Davis and Florian Wendelspiess for their valuable input.

The research in this thesis was made possible with financial support from *Colciencias*, the Colombian Department of Science, Technology and Innovation, and the Graduate School at the University of Minnesota.

No words can express my immense gratitude to Juan C. Chaparro for the immeasurable ways he has contributed to my life, and for his support, care and encouragement.

I am grateful for my parents and sisters and their unyielding support in my desire to pursue a doctoral degree.

Finally, a special thanks to my friends and fellow classmates.

Dedication

To my parents, Carmen and Daniel, and my sisters, Sandra and Andrea.

Abstract

This dissertation is comprised of three independent essays that address, respectively, spousal violence and female employment in Colombia, inequality of opportunity in adult health in Colombia, and welfare of rural Peruvian households. The evidence presented in the first essay, "Intimate Partner Violence and Women's Employment: Evidence from Colombia," suggests that victims of intimate partner violence are more likely to work. This relationship is likely mediated by a wife's decision-making power: women seem to engage in paid work to escape violent situations at home by enhancing their decision-making power. The second essay, "Inequality of Opportunity in Adult Health in Colombia," suggests that differences in parental educational attainment and household socioeconomic status during childhood are the most important dimensions of inequality of opportunity in adult health. The third essay, "Foods and Fads: The Welfare Impacts of Rising Quinoa Prices in Peru," shows that increases in the international price of quinoa, which have been driven by a high international demand of quinoa, are associated with a significant yet modest increase in the welfare of households in areas where quinoa is consumed and produced in Peru.

Contents

iv

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation contains three independent essays in development economics that address, respectively, gender-biased violence and female employment in Colombia, inequality of opportunity in adult health in Colombia, and the welfare effects of rising quinoa prices in Peru.

The first essay, "Intimate Partner Violence and Women's Employment: Evidence from Colombia," studies the relationship between intimate partner violence (IPV) and women's employment using data from the Colombian Demographic and Health Survey. There is a positive relationship between IPV and employment, which persists when husband's childhood experience of domestic violence is exploited as a source of plausibly exogenous variation for the incidence of IPV. The incidence of IPV increases the likelihood of female employment by about 16 percentage points. This result is robust to small departures from the exclusion restriction. To explain these findings, this chapter explores the role of women's decision-making power. Women may enter or increase their participation in the labor force to escape violent situations at home by enhancing their decision-making power. In particular, the effect of IPV on employment appears to be lower among abused women with higher initial decision-making power.

The second essay, "Inequality of Opportunity in Adult Health in Colombia," uses the 2010 Colombian Living Standards and Social Mobility Survey, a rich household survey that provides unique information about individual childhood circumstances in the country. This chapter provides calculations for a dissimilarity index and a Gini-opportunity index, two measures of inequality of opportunity, using a self-reported variable for health status. To obtain the relative contribution of various circumstances, such as parental education and household socioeconomic status in childhood, to the variation in the dissimilarity index, this chapter uses the Shapley-value decomposition. In addition to a national-level analysis, separate estimations for residents in urban and rural areas are provided. The findings suggest that 8% to 10% of the initial opportunities enjoyed by those who are healthier should be redistributed among those who are less healthy in

order to achieve equality of opportunity. Differences in household socioeconomic status during childhood and parental educational attainment appear to be the most important dimensions of inequality of opportunity in adult health.

The third essay, "Foods and Fads: The Welfare Impacts of Rising Quinoa Prices in Peru," explores the effects of increasing quinoa prices on changes in consumption of rural households using data from the Peruvian *Encuesta Nacional de Hogares*.[1] Riding on a wave of interest in "superfoods" in rich countries, quinoa went in less than a decade from being largely unknown outside of South America to being an upper-class staple in the United States. Because of that rapid rise in the popularity of quinoa, the price of quinoa more than tripled between 2006 and 2014. This chapter studies the impact of rising quinoa prices on the welfare of Peruvian households. Using 11 years of a large-scale, nationally representative household survey, and pseudo-panel methods, this chapter examines the relationships between: (i) the purchase price of quinoa and the value of real household consumption, which proxy for household welfare; and (ii) household quinoa production and household welfare. The findings suggest that increases in the purchase price of quinoa are associated with a significant increase in the welfare of the average household in areas where quinoa is consumed, which suggests that the quinoa price increase has had general equilibrium effects extending to non-producers. The results in this chapter also suggest a significant increase in the welfare of quinoa–producing households.

This dissertation is organized as follows. Chapter 2 estimates the effects of intimate partner violence against women on women's employment in Colombia. Chapter 3 provides the theoretical and empirical framework for the measurement of inequality of opportunity in adult health in Colombia. Chapter 4 presents analysis of the effects of rising quinoa prices on the welfare of rural households in Peru. Chapter 5 concludes.

---

[1]This essay is co-authored with Marc F. Bellemare and Seth Gitter

Chapter 2

Intimate Partner Violence and Women's Employment: Evidence from Colombia

2.1 Introduction

The World Health Organization reports striking findings on the prevalence and effects of violence against women. Almost one third of all women worldwide who have been in a marital relationship have experienced physical or sexual violence perpetrated by their male partners (World Health Organization, 2013). Most of these women report serious physical and mental health consequences, which include permanent injuries, pregnancy-related complications and impaired social functioning. In Latin America and the Caribbean, according to the World Health Organization estimates, about 24% of ever-partnered women report some exposure to physical intimate partner violence. Colombia is one of the countries in the region where violence against women is highly prevalent; in 2010, 37% of Colombian women reported physical or sexual spousal abuse over their lifetime, as well as several physical and psychological consequences associated with it (Profamilia, 2011). Intimate partner violence also affects labor market outcomes: victims reported that spousal violence affected their performance in daily activities and their labor productivity.

Most empirical studies focus on the determinants of spousal violence, including women's employment, with mixed results. Aizer (2010) exploits variation in industry-specific labor demand and finds that decreases in the male-female wage gap reduce violence perpetrated by domestic partners. Bhattacharyya, Bedi, and Chhachhi (2011) suggest that boosting a wife's economic status generates struggle within the household and leads to *more* violence. Heath (2014) focuses on access to factory jobs and finds that women with low bargaining power face increased risk of domestic violence upon entering the labor force. Other studies investigate the consequences of IPV and show that violence against women is related to higher rates of female unemployment (Lloyd, 1997; Lloyd and Taluc, 1999) and women working less hours (Meisel, Chandler, and Rienzi, 2003; Swanberg and Logan, 2005; Tolman and Wang, 2005). Other studies, on the contrary, find that spousal violence appears to lead to increased labor market participation (Farmer and Tiefenthaler,

2004) and more hours of work (Staggs and Riger, 2005). Studies in Latin America and the Caribbean are similarly inconclusive; some find that abused wives are more likely to work (Morrison and Orlando, 1999; Agüero, 2013), while others find that they more likely to exit the labor force (Rios-Avila and Canavire-Bacarreza, 2017).

In this paper, I estimate the relationship between reporting having experienced intimate partner violence (IPV) and women's employment. Further, I explore the role of women's decision-making power in mediating this relationship.

The main contributions of this study are twofold. First, using household survey data, I show that the effect of IPV on woman's employment is positive in Colombia and that this result persists after using the plausibly exogenous variation in the husband's childhood exposure to domestic violence as an instrumental variable for IPV. Women victims of intimate partner violence may decide to spend more time away from home and seek employment more actively to reduce their vulnerability by improving their economic situation. My findings support this notion as reported spousal violence does not prevent women from being active in the labor force: Women who experience IPV are 16 percentage points more likely to work than women who do not.

This paper also contributes to the literature on the economics of the family and on women's empowerment by exploring the role of women's decision-making power. Wives may need to increase their power within the relationship and gain control of their decisions to increase their ability to escape domestic violence or, at least, lessen its intensity. To provide an exploratory assessment of the role of bargaining power, I conduct three analyses. First, I study the relationship between IPV, employment and initial bargaining power of the wife to assess whether the effect of IPV on employment differs by her education or age at marriage, which proxy for initial bargaining power. Second, I examine the relationship between employment and whether a woman can make spending decisions for herself and participate in household decision-making. The last exercise consists of a mediation analysis using *sequential g-estimation*, a method recently proposed by Acharya,

Blackwell, and Sen (2016). With this method, I am able to calculate the *controlled direct effect* of IPV on employment, if I were to fix a woman's decision-making power at a particular level. I do not find strong evidence that decision-making power, as proxied in this paper, is the mechanism at work. Nonetheless, a woman's outside option and her decision-making power are so highly correlated, that I explore the mediating role of her willingness to divorce. I find suggestive evidence that willingness to divorce is mediating the positive relationship between IPV and employment.

The rest of this paper is organized as follows: Section 2.2 describes the 2010 Colombian Demographic and Health Survey data. Section 2.3 explains the empirical methods and discusses the identification strategy. I present results in Section 2.4, and conduct robustness checks in Section 2.5. I explore the role of women's bargaining power and willingness to divorce in Section 2.6. Section 2.7 provides concluding remarks.

2.2   Data and Descriptive Statistics

The 2010 Colombian Demographic and Health Survey (DHS) provides demographic, socio-economic and health information for women and children and is representative of the population at the national level and for urban and rural areas in all regions and departments.[1] The DHS is a three-stage stratified cluster sample that covers all but the two most sparsely populated departments in Colombia. The DHS also provides detailed information on intimate partner violence for the female population aged 15 to 49 years who are currently married or living in a consensual union. The DHS selected 52,952 women for the domestic violence module, but women who had never been married or in a de facto union, as well as divorced and widowed women, were all excluded by the DHS team during this part of the survey.[2] Of the 33,728 women finally interviewed, 8,200

---

[1] The DHS program is funded by the U.S. Agency for International Development (USAID).

[2] Another 1.06% of women were also excluded from the DHS because they could not be safely interviewed in private. Not being able to characterize this excluded part of the sample may be of concern if these women

were married more than once and 25,528 were married only once. Given that the domestic violence module of the questionnaire refers to abuse by the current or previous male partner without distinction, I focus on the sub-sample of women who have been married or in a consensual union only once. This is because, in the data released by the DHS, it is not possible to obtain any information on previous marriages or consensual unions. The final sample includes $25,528$ partnered women (8,180 are married and 17,348 are in a consensual union) who responded to the domestic violence module in the 2010 DHS. Although most of the women in the sample are not married, I refer to them as husband and wife, for convenience.

Intimate partner violence (IPV) is measured using the modified Conflict Tactics Scale (CTS) (Straus, 1979; Straus et al., 1996). The DHS team elicits information on domestic violence by administering this set of questions to one randomly selected woman in each household. The DHS team also obtains informed consent from the respondent at the beginning of the interview. The respondents are also reminded throughout the interview of the confidentiality of their responses.

I use three dummy variables for reported IPV. The first dummy, "Physical IPV", indicates whether the woman reported any experience of physical abuse in the past 12 months. That is, whether a husband or male partner: (1) Pushed or shook or threw something at her; (2) Slapped her; (3) Punched her with fist or something harmful; (4) Kicked or dragged her; (5) Tried to strangle or burn her; (6) Threatened her with knife/gun or other weapon; (7) Attacked her with knife/gun or other weapon; (8) Physically forced sex when not wanted; or (9) Bit her. The second dummy, "Emotional IPV", indicates whether. in the past 12 months, a husband or male partner: (1) Was jealous if the wife was talking with other men;(2) Accused her of unfaithfulness; (3) Did not permit her to meet her girl-friends; (4) Tried to limit her contact with her family; (5) Insisted on knowing where she was; or (6) Did not trust her with money. The third dummy, "Any IPV", indicates

---

are affected the most by IPV.

whether the woman reported any experience of physical and/or emotional IPV in the past 12 months.

In the sample, about 44.6% of women were victims of intimate partner violence. By type of incident, 14.6% of women reported physical/sexual abuse[3] and 41.5% reported emotional abuse in the past 12 months. This survey is also informative of intergenerational events of domestic violence. About 34.7% of these women report that their fathers had beat their mothers at least once during their childhood. Although no information is reported for whether the husband's father beat his mother, about 30% of wives in the sample report their male partners were mistreated during childhood.

Table 2.1 reports summary statistics for the dependent and explanatory variables used in this study. Although the DHS is not a comprehensive labor force survey, it collects data on the labor market status of women by inquiring about the following: (1) Current work status (including work in own and family-owned businesses); (2) Work status in the past 12 months if not currently working; and (3) Whether the woman has ever worked if she did not work in the past 12 months. In this sample, over 69% of wives are currently working or worked at least one month in the 12 months prior to the survey, and about 12% had never worked. For this study, I focus on the woman's *work status now and in the past 12 months.* Unfortunately, it is not possible to determine from the DHS data the timing of employment and violence: it is unknown whether the woman was working before the first event of IPV or whether she started to work after being abused.

---

[3]About 5% of the women in the sample were sexually abused.

9

2.3 Empirical Framework

2.3.1 The Equation of Interest

One contribution of this paper lies in the estimation of the impact of intimate partner violence on women's employment. This section discusses the equations to be estimated and the identification strategy used here in an attempt to provide an unbiased estimate of the relationship between women's employment and IPV.

Let L_{ir} be a dummy variable that indicates whether a woman is currently working or worked in the past 12 months. The first equation to be estimated in this paper is:

$$L_{ir} = \alpha + \mathbf{X'_{ir}}\mathbf{\Phi} + \beta \cdot IPV_{ir} + \theta_r + \epsilon_{ir} \tag{2.1}$$

where the subscripts denote individual i in department r. IPV_{ir} is a dummy variable that indicates whether the woman reported being a victim of IPV in the past 12 months; $\mathbf{X_{ir}}$ is a vector of individual and spousal characteristics including wife's and husband's age and educational attainment, wife's ethnicity, husband's work status, quantiles from a wealth index, and a dummy for urban residence. Other variables in $\mathbf{X_{ir}}$ include an indicator for whether the husband consumes alcohol, a dummy for current pregnancy, and dummies for presence of young and old children in the household. The θ_r term denotes department[4] fixed effects that are included to address potential bias due to unobserved heterogeneity across departments. The ϵ_{ir} term is an error term with mean zero. If IPV is exogenous with respect to employment, the estimate of β represents the average treatment effect (ATE) of IPV on women's employment status.

Because some of the husband's information could be missing, I also include two dummy variables indicating whether his education or his work status are unknown to

---

[4]Departments are the first administrative division in Colombia. There are 32 departments, including the capital city of Bogota.

the wife. It cannot be assumed that the missing information on the husband is unrelated to his wife's employment status. Therefore, I include these missing indicators as regular controls in both the first and second stage equations.

I estimate Equation (2.1), weighting each observation with the associated probability weights provided in the data. Given the binary nature of the dependent variable, my use of OLS means that every equation estimated in this paper is a linear probability model (LPM). In estimating an LPM rather than a logit or a probit model, I follow the recommendations of Angrist and Pischke (2008). The primary benefits of using a LPM are: (i) LPM does not rely on distributional assumptions required by the logit and probit specifications; and (ii) LPM does a much better job than probit models at handling a large number of fixed effects. The primary drawback to using a LPM is that it produces errors that are heteroscedastic. I use robust Huber-White standard errors in all estimations in order to address this concern. These standard errors are further clustered at the primary sampling unit level,[5] given the sampling scheme, to account for further sources of heteroscedasticity within sampling units.

The primary objective of this paper is to assess whether IPV has an impact on women's employment, as discussed in the introduction. Since IPV is likely endogenous to a woman's employment, the next section discusses the identification strategy used in this paper.

2.3.2 Identification Strategy

IPV is unlikely to be exogenous in Equation (2.1). Three sources of endogeneity are of particular concern. The first source is the potential for *reverse causality* or *simultaneity*: an improvement in a wife's employment opportunities or an increase in her labor in-

---

[5]Primary sampling units (PSU) are the first stage of selection in a multi-stage sampling procedure. In the DHS data, these units typically correspond to an enumeration area or a segment of an enumeration area. In this sample, there are 3,965 PSUs.

come may lead her husband to inflict violence on her. The second source is *unobserved heterogeneity* or *non-random selection* into violent relationships based on unobservable characteristics. Unobserved variables such as social norms or characteristics of the wife and her partner can influence both intimate partner violence and female employment, so that IPV and employment can be correlated even if the former does not have a causal effect on the latter. For example, husbands' characteristics such as drug or alcohol use or involvement in crime may directly affect the wife's decisions to work and directly lead to IPV. The third source of endogeneity is *measurement error*, which is particularly driven by under-reporting of incidents of domestic violence in survey data. Any of these sources of endogeneity will cause IPV to be correlated with the error term in Equation (2.1).

The identification strategy used in this paper relies on the use of an instrumental variable (IV). To produce consistent estimates, this variable must be conditionally correlated with reported IPV, but uncorrelated with the error term in Equation (2.1). The first assumption, that the IV is correlated with IPV, can be ascertained using a test of the null hypothesis that the instrument has no explanatory power with respect to the endogenous variable. The result of this test is presented in section 2.4. The second assumption, or the exclusion restriction, requires that the IV affects women's employment only through IPV. This restriction is not directly testable but this section discusses its validity in this context.

The instrumental variable I use for reported IPV is a dichotomous variable that indicates whether a woman reports that her husband was mistreated or regularly beaten by his parents or stepparents as a child. The identifying assumption is thus that husband's childhood experience of domestic violence is uncorrelated with ϵ_{ir} in Equation (2.1). The second-stage equation is:

$$L_{ir} = \alpha + \mathbf{X}_{\mathbf{ir}}'\mathbf{\Phi} + \beta \cdot \widehat{IPV_{ir}} + \theta_r + \epsilon_{ir} \tag{2.2}$$

where $\widehat{IPV_{ir}}$ denotes the predicted probability of IPV conditional on the instrument Z_{ir}

and $\mathbf{X_{ir}}$, obtained from the first-stage regression of IPV on the husband's childhood experience of domestic violence and the control variables included in Equation (2.2), which is given by:

$$IPV_{ir} = \alpha_1 + \mathbf{X'_{ir}}\mathbf{\Pi} + \rho \cdot Z_{ir} + \varphi_r + \mu_{ir} \qquad (2.3)$$

where Z_{ir} is a dichotomous variable for the husband's childhood experience of domestic violence, μ_{ir} is an error term with mean zero, and all other variables are defined as above.

If the instrument has conditional predictive power for IPV and satisfies the exclusion restriction and the monotonicity assumption (which are discussed below), the IV estimate of the coefficient β is a local average treatment effect (LATE) of reported IPV on women's employment, i.e., the increase in the probability of work (as measured by the dependent variable) due to IPV for those couples for whom a husband being abused by his parents during childhood induces a change in IPV. This is the treatment effect on the group of "compliers". In this application, compliers are couples in which the husband's IPV propensity is affected by his exposure to violence as a child. The compliers group is a subset of all couples, and it is impossible to determine whether the effect of IPV estimated for this group is the same as that for the population as a whole.

The husband's childhood experience of domestic violence has predictive power for IPV, and this satisfies the "relevance" assumption in this setting, for various reasons. Children who are exposed to domestic violence have higher levels of internalizing (depression, anxiety) and externalizing (physical aggression) behaviors and post-traumatic stress disorder (Evans, Davies, and DiLillo, 2008; Graham-Bermann et al., 2012). Further, some studies suggest that childhood exposure to domestic violence becomes a risk factor for being a victim and/or perpetrator of violence later in life, both in developed (Whitfield et al., 2003) and developing countries (Martin et al., 2002). Previous studies for Colombia (Assaad, Friedemann-Sanchez, and Levison, 2016; Friedemann-Sánchez

and Lovatón, 2012) show that a partner's experience of violence against him as a child is highly associated with the incidence of intimate partner violence in adulthood.

One argument for why the instrumental variable proposed in this paper is likely to satisfy the exclusion restriction is that it affects a husband's potential engagement in violent behavior long before the couple's formation, as supported by the studies on intergenerational transmission mentioned in the previous paragraph. Therefore, with the inclusion of appropriate controls for household socioeconomic characteristics, it is plausible that a husband's childhood experience of violence is uncorrelated with unobserved variables affecting the wife's current employment status. It is still possible that the correlation between Z_{ir} and ϵ_{ir} is non-zero due to the effect of assortative, endogenous matching, i.e., husbands and wives choose each other on the marriage market (Ackerberg and Botticini, 2002). I include in the regression various controls for the wife's and husband's characteristics that are variables on which the matching may occur such as their education, their age and the occupation of the husband. The inclusion of these variables increases the likelihood that the exclusion restriction holds.

The estimated treatment effect could be different from the effect for the couples where the husband would be violent either way (the "always takers") or the couples where the husband does not commit IPV whether exposed to violence as a child or not (the "never takers"). "Defiers" would be cases where the man turns out violent if he was not exposed to violence as a child, but if he were exposed he would be peaceful in his marriage. Perhaps being exposed to violence makes him commit to never being violent. The empirical evidence, however, suggests that the potential for a husband consciously choosing to avoid perpetuating violence as an adult, despite being abused as a child, may be ruled out in most cases (Whitfield et al., 2003; Kishor and Johnson, 2004; Flake and Forste, 2006; Friedemann-Sánchez and Lovatón, 2012). This evidence further suggests that I can rule out the existence of "defiers" and that the monotonicity assumption is

likely to be satisfied.[6] If, however, the effect of the IV on the endogenous regressor is non-monotonic, one must assume homogenous treatment effects (i.e., the treatment effect is the same for everyone) and the LATE interpretation of the IV estimate on β may no longer be valid. In such a case, one cannot guarantee that IV estimates a weighted average of the underlying causal effects of the affected group.

Though instrumenting for IPV using the husband's childhood experience of domestic violence can mitigate simultaneity as a source of endogeneity, it does not fully address endogeneity coming from measurement error in reports of domestic violence. In this regard, note that throughout my analysis, I estimate the relationship between *reported* IPV, as opposed to actual IPV, and employment status.

2.4   Results

The main empirical results are reported in Tables 2.2 to 2.5. Demographic characteristics such as age, ethnicity, educational attainment, household wealth,[7] department of residence and urban residence are included to capture earnings potential that may affect a woman's decision to work. I include fixed effects for the department of residence and a dummy for urban residence to control for different labor demand conditions. Fertility characteristics (presence of children between 6 and 18 years old, presence of children 5 years old or less, and dummy variables for having had a child in the past 12 months) and husband's characteristics (age, educational attainment, work status, and whether he drinks alcohol) are also included to control for other potential factors that impact employment by affecting the costs and benefits of working relative to not working.

---

[6]The instrument may have no effect on some individuals, but all those who are affected are affected in the same way, so that all individuals who change their treatment status as a result of a change in the instrument either get all shifted into treatment, or get all shifted out of treatment.

[7]Household wealth is measured with the DHS wealth index readily available in the dataset and calculated using the methodology of Filmer and Pritchett (2001)

Recall that three IPV variables are used in this analysis: (1) Whether the wife experienced any physical and/or emotional IPV in the past 12 months; (2) Whether the wife experienced physical IPV in the past 12 months; and (3) Whether the wife experienced emotional IPV in the past 12 months.

As a baseline, I first estimate the probability that a woman works, treating IPV as fully exogenous, using a linear probability model (LPM) to estimate Equation (2.1). The OLS results from the linear probability model are presented in Table 2.2. I find that being physically and/or emotionally abused by a husband in the past 12 months appears to increase the likelihood that a wife currently works or has worked in the past 12 months by between 3.4 and 4.4 percentage points, when the full set of controls are included in the regressions (see columns 2, 4 and 6). Because of the potential endogeneity problems discussed in section 2.3, these results should be considered to be (conditional) associations between women's employment and IPV, and so they cannot be given a causal interpretation.

In an effort to provide an unbiased estimate of the impact of IPV on women's employment, I rely on two-stage least squares estimation. For this, husband's childhood experience of domestic abuse is used to instrument IPV. As with the LPM, I control for wife's and husband's characteristics, and cluster the standard errors at the primary sampling unit level. This estimation strategy allows me to conduct a number of tests on the validity of the instrument. The first test is a diagnostic regression of the dependent variable on the IV as the only regressor, suggested by Angrist and Pischke (2008), to provide evidence in favor of a relationship flowing from the husband's childhood experience of domestic abuse to women's employment. Table 2.3 presents the results from such a reduced-form regression and suggests that the relationship is positive and statistically significant. The second test is whether the instrument has sufficient explanatory power in the first stage equation. The F-statistics for the instrument in the first stage for any experience of IPV, shown in Table 2.4, as well as for physical and emotional IPV, are all

well above the threshold level of 10 for an instrument not to be considered weak.

For an additional test the strength of the instrument, I use a test proposed by Montiel Olea and Pflueger (2013), which is appropriate to test for weak instruments with one endogenous regressor. This test also allows for errors that are not conditionally homoscedastic and not serially uncorrelated. Upon testing the instrument in the regression where any experience of IPV is the endogenous variable, I obtain an effective F-stat of 368.9 with a bandwidth threshold of 10% and a 2SLS critical value of 23.1. These test results suggest rejection of the null hypothesis of weak instruments. Similarly, I reject the null hypothesis of weak instruments when the regressions for physical and emotional IPV are studied separately (the effective F-stats are at 159.3 and 260.4, respectively.)

Results from regressing the indicator for IPV on the instrument and various controls, the first stage of the 2SLS analysis, are shown in Table 2.4. Having a male partner who was abused by his parents as a child increases the probability of experiencing IPV by 14.6 percentage points. The effects are similar when physical and emotional violence are examined separately, with estimated increases of 8.2 and 12.7 percentage points, respectively.

Instrumental variable results from the estimation of Equation (2.2) suggest that IPV is significant and positively associated with women's employment (see Table 2.5). The experience of any event of spousal violence increases the likelihood of work by 16.1 percentage points, and this estimate is significant at the 1% level. Considering physical violence alone the increase is of 28.7 percentage points, whereas for emotional violence it is 18.5 percentage points, and both of these estimates are significant at the 1% level.

These IV estimates of the effect of IPV on women's employment are much higher than the OLS estimates in Table 2.2. When considering the magnitude of these results, it is important to keep in mind that IV estimates a local average treatment effect (LATE). This is the effect of IPV on the likelihood of employment for wives in couples in which the

husband's IPV propensity is affected by his exposure to violence as a child.

2.5 Sensitivity to Potential Violations of the Exogeneity of Husband's Childhood Exposure to Domestic Violence

2.5.1 Plausibly Exogenous Instrument

The instrument, husband's childhood experience of violence, may fail to satisfy the exclusion restriction. That is, it is possible that the husband's experience with violence in childhood is directly correlated with the wife's labor status in other ways, mainly via *assortative matching*. Assortative matching does not have to work through a direct impact of husband's childhood experiences on his wife's employment. It could work through his choice of wife, i.e., a man with certain childhood experiences chooses a wife who has certain personality traits that have an effect on her employment. This constitutes a potential threat to the exclusion restriction assumption upon which the validity of the instrument depends. The regression, however, includes a variety of husband's and wife's observable characteristics that will partly control for assortative matching. These variables are: age; ethnicity; educational attainment; husband's occupation; and wife's own childhood exposure to violence. The inclusion of controls for household socioeconomic status also support the exclusion of the husband's childhood experience of violence from the wife's labor status equation.

To estimate the sensitivity of the two-stage least squares estimates to violations of the exclusion restriction, I follow Conley, Hansen, and Rossi (2012). The effects of the failure of the exclusion restriction can be seen by re-estimating Equation (2.2) as follows:

$$L_{ir} = \alpha + \mathbf{X_{ir}}'\Phi + \beta^{IV}IPV_{ir} + \gamma Z_{ir} + \theta_r + \epsilon_{ir} \tag{2.4}$$

The exclusion restriction in the usual IV model holds when $\gamma = 0$. In this section, I consider two methods for inference about β^{IV} without assuming γ is exactly zero. In the first method, the union of confidence intervals (UCI) method, I assume only that the support of γ is known. In the second method, the local-to-zero (LTZ) approximation, I assume that γ is as a random parameter that can be described by a prior distribution.

2.5.1.1 The Union of Confidence Intervals Method

If the true value of γ is a value γ_0 in the bounded support for γ, Γ, then one could estimate via two stage least squares, using Z as instruments, the following equation:

$$L_{ihr} - \gamma_0 Z_{ir} = \alpha + \mathbf{X_{ir}}'\Phi + \beta^{IV} IPV_{ir} + \theta_r + \epsilon_{ir} \tag{2.5}$$

Further, using all points in Γ and based on the asymptotic variance of the two-stage least squares estimator, one could obtain a $(1 - \alpha)\%$ confidence interval for β^{IV}. Conley, Hansen, and Rossi (2012) explain that, by construction, the union of confidence intervals, for all values of γ in Γ, will cover the true parameter value of β^{IV} with at least a probability $(1 - \alpha)\%$, asymptotically.

To implement the UCI approach, one needs to make some assumptions about the interval for Γ. I assume that γ is close to zero since, a priori, I do not expect the direct effect of IPV on employment to be very large. Moreover, I assume a symmetric support centered at zero, so that $\Gamma = \{-\delta, \delta\}$ for different values of δ, a parameter in the $[0, \alpha]$ interval.

Graphical results are shown in Figure 2.1. I compute 95% confidence bounds for the coefficient on IPV. The figure shows how large the exclusion restriction violation would need to be to invalidate my results in the previous section. I provide results for the IPV variable that includes both physical and emotional violence as well as for the physical and

emotional variables taken separately. For the first IPV variable, I find that the exclusion restriction violation is small (i.e., δ is small) since the UCI excludes zero up to a delta value of about 0.012. The figure also shows that the true value of the coefficient on IPV is positive, which is consonant with my main results. Similar conclusions can be derived for physical and emotional IPV.

2.5.1.2 The Local-to-Zero (LTZ) Approximation

Under this approach, the exclusion restriction requirement is relaxed by allowing for uncertainty in the priors about γ. Conley, Hansen, and Rossi (2012) explain that treating γ as local-to-zero, produces the following approximation to the distribution of β^{IV}:

$$\hat{\beta} \sim_{approx} N(\beta, V_{2SLS}) + A\gamma \tag{2.6}$$

$$A = (X'Z(Z'Z)^{-1}(Z'X))^{-1}(X'Z) \tag{2.7}$$

$$\gamma \sim F \tag{2.8}$$

where V_{2SLS} is the variance-covariance matrix from the two-stage least squares estimation and F is the specified prior distribution. The A term reflects the influence of exogeneity error.

If one further assumes that the prior for γ is Gaussian, say, $N(\mu_\gamma, \Omega_\gamma)$, then

$$\hat{\beta} \sim_{approx} N(\beta + A\mu_\gamma, V_{2SLS} + A\Omega_\gamma A') \tag{2.9}$$

In the application, I assume a prior Gaussian distribution for γ centered at zero so that $\gamma \sim N(0, \delta^2)$ and compute 95% confidence bounds, with δ defined as in the previous section. Once the 95% upper limit crosses the zero-line, that is, once confidence bounds include zero, the 2SLS estimates are no longer significant at the 5% significance level. Figure 2.2 summarizes the estimated results for each type of IPV. This figure shows how

adjusted confidence intervals vary with δ and show how much uncertainty the 2SLS estimates can handle and still remain statistically different from zero. In all cases, the 95% LTZ confidence intervals include a zero value up to a δ value of $0.009 - 0.01$. For the physical and/or emotional IPV variable, the 2SLS estimate of the relationship between IPV and women's employment is found to be in the $[0.064, 0.299]$ 95% confidence interval. For physical and emotional IPV, the intervals are $[0.095, 0.458]$ and $[0.076, 0.356]$, respectively. It thus looks as though my 2SLS results are robust to small departures from the exclusion restriction assumption.

2.5.2 IPV as an Imperfect Instrumental Variable

In this section, I implement the Imperfect Instrumental Variable approach developed by Nevo and Rosen (2012) to relax the exclusion restriction assumption and bound the estimates for the parameter of interest.

The method of Nevo and Rosen (2012) relies on two critical assumptions. First, the correlations between the endogenous regressor and the error term in Equation (2.2) and between the instrument and the error term in the same equation have the same sign. This implies that

$$\rho_{Z,\epsilon} \cdot \rho_{IPV,\epsilon} \geq 0 \tag{2.10}$$

where $\rho_{Z,\epsilon}$ denotes the correlation between the instrument, Z, and the error term ϵ, and $\rho_{IPV,\epsilon}$ the correlation between IPV and the error term.

The second assumption is that the correlation between the instrument and the error term is less strong in absolute terms than the correlation between the endogenous variable and the error term. This last assumption weakens the usual assumptions for instrumental variables which would require the correlation between the instrument and the error term

to be zero. This is,

$$|\rho_{IPV,\epsilon}| \geq |\rho_{Z,\epsilon}| \qquad (2.11)$$

In my application, the above assumptions are likely to be satisfied. Unobserved characteristics of the husband and his wife influence women's employment through positive assortative matching. The assortative matching literature suggests that when choosing a spouse individuals look for partners who share common productivity traits, work status or earnings potential (Lam, 1988; Kalmijn, 1994; Jepsen and Jepsen, 2002). Some of the studies mentioned in section 2.3 suggest that individuals abused in childhood may be more likely to exhibit adverse psychosocial outcomes in adulthood, which are unobserved but may be negatively correlated with the woman's labor productivity or propensity to work. The potential for positive assortative matching then suggests that the correlation between the endogenous regressor (IPV) and the error term is likely negative, as well as the correlation between the instrument and the error term.

Since I am controlling for a large number of factors affecting employment, I expect the correlation between the error and the instrument to be negligible and at least smaller than the correlation between the endogenous regressor and the error term. When this condition is satisfied, one can estimate the lower bound of the parameter estimate using a generated instrumental variable suggested in Nevo and Rosen (2012). In my case, the generated instrumental variable is defined as

$$V(\lambda) = \sigma_{IPV} \times Z - \lambda \times \sigma_Z \times IPV \qquad (2.12)$$

where σ_{IPV} and σ_Z denote standard deviations. The λ term denotes the ratio between the correlations between the instrument and the endogenous regressor with the error term: $\lambda = \frac{\rho_{Z,\epsilon}}{\rho_{IPV,\epsilon}}$. This term is in principle unknown; however, Nevo and Rosen (2012) show that, under the above assumptions, its value lies between 0 and 1. With $\lambda = 1$ one has

the worst case in which the IV is as endogenous as the endogenous regressor. In contrast, when $\lambda = 0$, one has the IV valid case.

The first stage estimates show that the instrument has a positive effect on the endogenous regressor, IPV. If the first and second assumptions hold and, as in my case, the covariance between the endogenous variable and the instruments is positive, Nevo and Rosen (2012) show that the bound for the true parameter value is one-sided.[8] The estimate obtained using the imperfect instrumental variable proposed in Equation (2.12) will be a lower bound of the true parameter estimate of the effect of IPV on female employment. That is,

$$\beta \geq \max \left\{ \beta_{V(\lambda=1)}^{IV}, \beta_Z^{IV} \right\} \tag{2.13}$$

Results without covariates are displayed in Table 2.6. I estimate that the lower bound of the true parameter is 0.15, when any event of IPV is the endogenous regressor of interest. The lower bounds for physical and emotional IPV are 0.21 and 0.18, respectively. These results suggest that the 2SLS estimation using husband's childhood experience of domestic violence as an imperfect instrument is robust, since the estimated coefficients are not substantially modified when high levels of correlation between this imperfect instrument and unobservables in the main equation are allowed.

Controlling for other covariates in IV regressions is often important because the assumption of exogeneity may hold only after conditioning on all exogenous variables. In the Nevo and Rosen approach, the assumptions on the correlation structure do not change for the more general version of the model where there are additional covariates. To estimate the lower bound of the IPV effect using covariates, I also use 2SLS. Results with covariates, which are displayed in Table 2.7, do not change drastically. The lower bounds for any type, physical and emotional IPV are 0.16, 0.29 and 0.19, respectively. That is,

---

[8]If the correlation between the instrument and the endogenous regressor were negative, then the true parameter would be bounded between the IV estimate with the Nevo-Rosen instrument and the original IV estimate.

the Nevo-Rosen bounds for the effect of physical and/or emotional IPV on employment are $[0.16, \infty)$. The bounds for the effect of physical IPV and emotional IPV are $[0.29, \infty)$ and $[0.19, \infty)$, respectively. When I relax the exogeneity assumption, the effect of IPV on women's employment is still positive and larger than the effect estimated with OLS.

2.6 A Possible Explanation for the IPV effect on Women's Employment

In this section, I study the role of women's decision-making power in explaining the positive effect of IPV on women's work.

2.6.1 Women's Decision-making Power and Autonomy

Women's decision-making power may mediate the positive relationship between IPV and employment. In order to increase their ability to escape domestic violence, wives may need to increase their power within the relationship and gain (more) control of their decisions and earnings. This behavior is consistent with the game-theoretic model of Farmer and Tiefenthaler (2004), which includes a threat point that is increasing in a woman's income and other outside opportunities.[9] To achieve this, abused women may be more likely to work.

In order to provide an exploratory assessment of the role of women's decision-making power, I do three things. First, I study the relationship between IPV, employment and initial bargaining power of the wife to assess whether the effect of IPV on employment differs by her age at marriage or education, which proxy for initial bargaining power. Second, I examine the relationship between employment and whether a woman can make

---

[9]If an increase in a woman's threat point increases her chances of leaving and lowers the violence when she stays, then she would seek employment to improve her alternatives.

spending decisions for herself and participate more in household decision-making. Last, I use mediation analysis to assess whether decision-making power is a possible mechanism.

2.6.1.1 IPV, Employment and Initial Bargaining Power of the Wife

The relationship between IPV and employment may vary with the wife's age, education and age at marriage, which are variables that suggest the initial bargaining power of the wife upon entering the labor force, as shown in the study of Heath (2014). In order to assess whether the effect of IPV on employment differs by the initial bargaining power of the wife, I estimate OLS regressions, because of the potential endogeneity of the initial bargaining power variables. In these regressions, I include interactions of the IPV variable (which includes physical or emotional abuse) with the variables for age at marriage and years of education, along with the covariates used in previous sections.

The first column of Table 2.8 show that a one year increase in the age at marriage is associated with a statistically significant 0.3 percentage-point increase in the probability of employment. Among abused women, compared to those who have not been abused, the increase in the probability of employment is negligible. Similarly, column 5 suggests that an additional year of education is associated with a statistically significant 2.2 percentage-point increase in the probability that a woman has worked in the past 12 months, and it is almost the same for abused women. These results suggest that the correlation between IPV and employment is positive although this effect may be lessened among abused women with higher initial bargaining power.

The association between IPV and employment may be lessened among women with higher initial bargaining power due to marriage matching. To investigate this channel, I include husband's education or his age relative to his wife as controls. If marriage matching changes the relationship between IPV, employment and decision-making power, then the inclusion of the variables for difference in age or difference in years of education

25

will decrease the magnitude of the interaction term between the wife's education or age at marriage and IPV. Results in columns 2 and 5 in Table 2.8, however, show that conditional on the husband's characteristics, the relationship between a wife's initial bargaining power, IPV and employment remain almost identical. That is, the effect of a woman's age at marriage and education on bargaining power may not depend on her husband's age or education. This result is also observed when variables for intergenerational domestic violence (whether the husband was mistreated during childhood or whether the wife's mother was abused by her husband) are included as controls, as displayed in columns 3 and 6 in the same table.

2.6.1.2   IPV, Employment and Wife's Decision-Making Power

In order to increase their ability to escape domestic violence, wives may need to increase their power to make decisions within the household. Instrumental theories of domestic violence suggest that men use violence to counteract the increase in decision-making power that women get upon working (Eswaran and Malhotra, 2011), but women with sufficiently high decision-making power are more able to escape abusive marriages and thus do not face such increase in violence (Heath, 2014).[10] Although I am not able to observe transitions in and out of the labor force and the timing of violence relative to labor market decisions, I provide suggestive evidence on how employment may have affected a woman's decision-making power.

The measures of decision-making power that I use are: (i) whether a wife has the final

---

[10]Heath (2014) provides a brief summary of the predominant economic and social theories of domestic violence. These theories are broadly categorized between theories of expressive violence and of instrumental violence. In expressive violence theories, "male backlash" occurs in response to improvements in a woman's economic opportunities. A husband who feels less economically empowered than his wife may resort to violence to reassert his identity as the most powerful member of the household. In instrumental violence theories, domestic violence is a tool used by husbands to control household resources or the behavior of their wives. These theories usually employ household bargaining models, wherein a woman's outside option is a key determinant of bargaining power and the actions taken by the household members. Thus, in situations where the outside option improves sufficiently, an abused woman may be better able to leave the abusive relationship.

say in own health care; (ii) whether the wife has a final say in large household purchases; and (iii) whether the wife has a final say in purchases for daily needs. To construct these measures, I use the wife's reports of who makes the decisions in the household. Specifically, I assess a woman's decision-making ability using her answer to the question: "Who has the final say in [X] in your household?". If the woman alone has the final say, then each of the measures above equals one, and zero otherwise.

Initial bargaining power seems to play an important role in determining the employment effects on woman's decision-making power. Heath (2014) shows that women with higher bargaining power before entering the labor force are less likely to face domestic violence upon entering the labor force. In order to assess how the employment effects may differ by a wife's initial bargaining power, I also include the wife's age at marriage and education, and interactions of these with employment, as controls in the regression of employment on the decision-making power measures.

Previous evidence for Colombia suggests that women working in the cut-flower industry, via formal jobs, increased their self-esteem and gained higher decision-making power within the household Friedemann-Sánchez (2006). My results are consistent with these findings. Columns 1, 4, 7 and 10 in Table 2.9 suggest that employment is associated with higher decision-making power. The results also show that education and age at first marriage are also positively correlated with higher decision-making power (age at marriage is only statistically significant in the regression where the outcome is final say in own health care and where the outcome captures overall bargaining power.) Notably, columns 3, 6, 9 and 12 suggest that employment is associated with less decision-making power in women with more education (coefficient on interaction term is negative).[11] These results suggest that employment and education do not interact positively in raising a woman's decision-making power.

---

[11]The point estimates on the interaction between employment and education do not change drastically after controlling for husband's characteristics such as his age and education or after adding an interaction term of employment with husband's education.

2.6.1.3   Women's Empowerment as a Potential Mediator

As noted above, one potential explanation for my results is that they are driven by women's economic empowerment. To assess this, I check how much of my baseline results can be explained by decision-making power, the mediator. I do so in two ways. First, I include the mediator as a covariate in the 2SLS specification, along with IPV, the treatment of interest. This analysis is shown in Tables 2.10 to 2.13, columns 1,3 and 5. The coefficient on IPV remains significant, suggesting that its direct effect may not operate through decision-making power.

To gain a deeper understanding of the mechanism, I also use a causal mediation analysis in this section. Imai et al. (2011) explain that the goal of this analysis is to decompose the causal effect of IPV into an indirect effect, which represents the hypothesized causal mechanism, and a direct effect, which represents all the other mechanisms. My hypothesized causal mechanism is woman's decision-making power. One problem with using decision-making power in a causal mediation approach, however, is that it violates the key assumption of no *intermediate confounders*, which are consequences of IPV that also affect the mediator (decision-making power) and outcome (employment). To address this concern, I employ a method recently developed by Acharya, Blackwell, and Sen (2016) that allows the identification of causal direct effects in the face of intermediate confounders. With this method, I am able to calculate the *controlled direct effect* of IPV on employment, if I were to fix a woman's decision-making power at a particular level (that is, decision-making power has the same fixed value for all units). The *indirect effect*, in contrast, is the portion of the total effect of IPV due to the IPV effect on the mediator and the mediator's subsequent effect on employment.

To calculate the controlled direct effect, I use a two-stage estimate: the *sequential g-estimator*. For this, it is assumed that one can control for a set of covariates that satisfies the *sequential unconfoundedness* assumption that there exist no omitted variables for two

relationships: one between employment and IPV and the other between employment and bargaining power (Acharya, Blackwell, and Sen, 2016). The first group of variables, whihc relate employment and IPV, is denoted as the group of *"pre-treatment" covariates*, whereas variables in the second group, which relate employment and bargaining power, are denoted as *intermediate covariates*. Including intermediate covariates help make the sequential unconfoundedness assumption more plausible.

To estimate the direct effect of IPV using mediation analysis, I first estimate the effect of decision-making power on employment, controlling for "pre-treatment" covariates[12] and intermediate covariates,[13] as well as IPV. I then transform the outcome variable by subtracting the (predicted) effect of decision-making power to create counterfactual estimates of the outcome as if all women had the same decision-making power. Finally, I estimate the effect of IPV on this transformed variable using 2SLS with husband's childhood exposure to domestic violence as the instrument, along with the intermediate covariates. The 2SLS estimator gives the controlled direct effect of IPV on employment. One note of caution: in the context of this paper, the results from this mediation analysis are only exploratory. Although it is possible to use instrumental variables in the second stage of the sequential g-estimation, when I estimate the effect of IPV on the transformed variable, that decision-making power is a potential endogenous variable would require another estimation strategy that relies on further compliance and selection-on-observables assumptions[14] that are beyond the scope of this paper.

Estimates from this mediation analysis are reported in Tables 2.10 to 2.13. Bargaining power is proxied here with variables for final say on own health care (Table 2.10), on large purchases (Table 2.11) and on purchases for daily needs (Table 2.12). Table 2.13 presents

---

[12]This set of covariates include: Female's characteristics: age, ethnicity, age at first marriage, years of education, woman's father hit her mother when she was a child, woman's mistreated by parents when she was a child. Spouse/partner's characteristics: age, years of education,, occupation (7 cat), alcohol consumption. Household characteristics: urban/rural area.

[13]This set of covariates include: Female's characteristics: presence of children 6-18 years old, presence of children less than 6 years old.,any childbirths past year. Household characteristics: wealth quintile group

[14]For a reference, see work in progress by Blackwell (2016) on causal interaction between two treatments.

the results for a bargaining power variable that equals 1 if the woman has a final say on any of the three variables previously described. Overall, compared to the baseline estimates of Table 2.5 and the estimates in columns 1, 3 and 5 in Tables 2.10 to 2.13, these results suggest that the measures of decision-making power used have little influence on the IPV effect on work. The direct effects of IPV are similar to those in Table 2.5 and are still significant. I do not find strong evidence that the IPV positive effect on employment may operate via the decision-making power variables here considered. Nonetheless, it is still possible that these proxies for decision-making power are imperfect, but data limitations prevent me from studying other potentially better measures of decision-making power.

The last potential explanation for my results is a woman's willingness to separate/divorce from her husband in the past 12 months, the same time period for which work and abuse are reported. If, theoretically, women's decision-making power comes from their outside option, then decision-making power is still a channel in the results, which are explained by a willingness to divorce, since this also comes from increases in their outside option.

To test this mechanism, I also use the sequential g-estimation method and calculate the controlled direct effect of IPV on employment using willingness to divorce as the hypothesized mediator. Estimates from this analysis are reported in Table 2.14. The results indicate that willingness to separate seems to have an important influence on the IPV effect on employment as the IPV estimate is no longer statistically significant, despite willingness to separate itself not being statistically significant in any regression.

2.7 Concluding Remarks

This paper estimates the effect of reported experience of intimate partner violence on women's employment. An econometric estimation that ignores the potential endogeneity

problem between IPV and employment leads to biased estimates of the effects of IPV. Two sources of endogeneity are of particular concern: reverse causality and unobserved heterogeneity. In an attempt to deal with these sources of endogeneity, I employ an instrumental variables approach. I use as an instrument for IPV a dummy variable that indicates whether a husband was mistreated by his parents as a child. I find that any event of intimate partner violence is associated with a 16.1 percentage-point increase in the likelihood of employment, whereas physical and emotional violence are associated with a 28.7 and 18.5 percentage-point increase, respectively. The results suggest that the incidence of IPV does not restrain women from being active in the labor force, and instead have the opposite effect.

The evidence presented in this paper supports the hypothesis that women may behave strategically in their labor market decision-making and seek employment to improve their outside alternatives when faced with intimate partner violence, as suggested by Farmer and Tiefenthaler (2004). Women with higher initial bargaining power (proxied by age at marriage and education) are more likely to work, and employed women also seem to enjoy higher decision-making power within the household. However, upon exploring whether a woman's decision-making power mediates the positive impact of IPV on employment, using the sequential g-estimation method of Acharya, Blackwell, and Sen (2016), I find little evidence of this being the mechanism behind my results. This result, however, may reflect that the variables for decision-making power used in this paper are actually imperfect measures. To further explore this channel, I explore the mediating role of a woman's willingness to divorce, which is increasing in her outside options and perhaps in her decision-making power. Using the mediation analysis, I find suggestive evidence that willingness to divorce is mediating the positive relationship between IPV and employment.

While the wife's partner's childhood experience of violence may not be a perfect instrument because of some remaining concerns about its excludability and monotonicity,

it provides an alternative way to control for selection. If the instrument violates the exclusion restriction, its effect would be to bias the effect of IPV on women's work. That is, the effect would capture both any effect of IPV itself and also effects operating through pathways relating to the husband's direct impact on the wife due to marriage matching, for instance.

That the IPV measure used in this paper is self-reported posits another problem. Although the DHS program attempts to minimize the underreporting and measurement error of this variable by "building rapport with the respondent, ensuring privacy, providing the respondent with multiple opportunities for disclosure [. . .] not only by asking them many different times about any experience of violence, but also by asking them about many different forms of violence" (Kishor and Johnson, 2004), my estimates of the effect of IPV on employment should be interpreted cautiously. Still, despite the number of caveats to the results presented, this study sheds new light on the impact of intimate partner violence on female labor market decisions.

My findings may suggest some important policy implications. That women victims of IPV are more likely to work suggests that they may benefit from counseling and legal help inside and outside the workplace. This is particularly important since previous studies (Farmer and Tiefenthaler, 2004) suggest that IPV has negative effects on labor productivity. Provision of women's shelters and better enforcement of the law may also help women facing IPV to lessen the severity of the violence. Although the *Comisarias de Familia* program has been available to abused women in Colombia for more than 20 years, no rigourous, economic evaluation of the program has been conducted yet.

Possibilities for future research include tackling the remaining methodological issues using administrative data to study the effect of IPV on the transitions in and out of the labor market, which also requires being able to observe the full labor history of a woman.

Figure 2.1: Conley-Hansen-Rossi Bounds Test for Instrument Validity: Union of Confidence Intervals
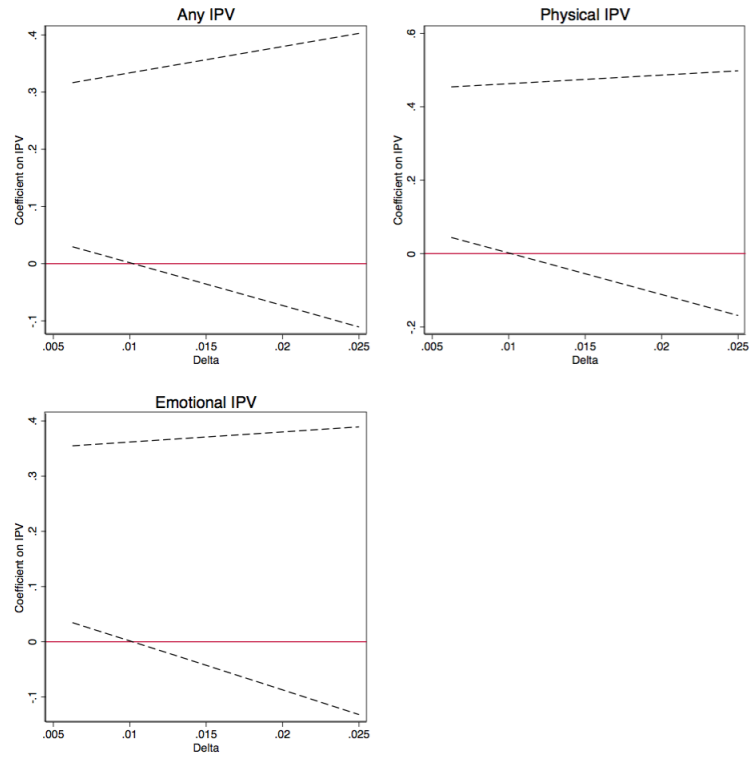
Note: All the reported bounds are for the 95% confidence intervals which have been generated from robust PSU clustered standard errors. The estimates are obtained using the STATA command `plausexog` by Clarke (2014).
This figure presents 95% confidence intervals for the estimated coefficient of IPV under the assumption that the instrumental variable has an baseline influence on employment. On the horizontal axis, I vary the baseline influence of husband's childhood experience of domestic violence on employment.

Figure 2.2: Conley-Hansen-Rossi Bounds Test for Instrument Validity: Local-to-Zero Approximation
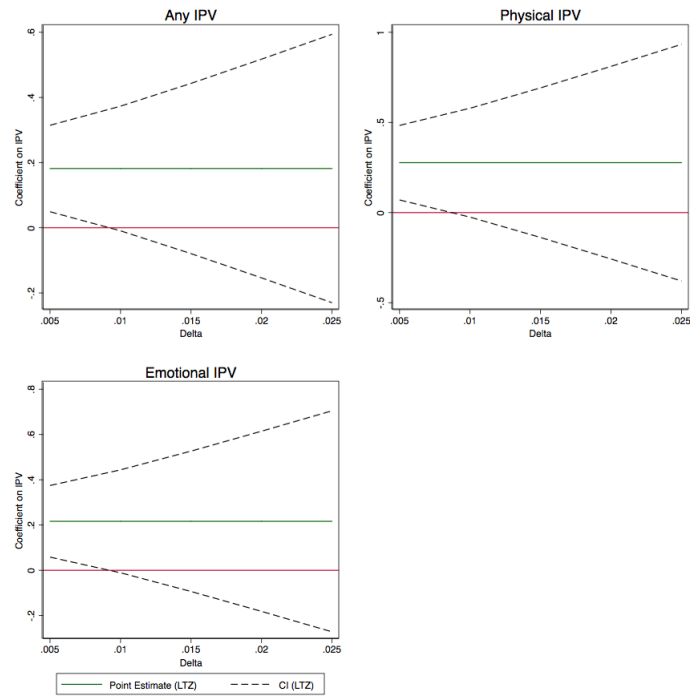
Note: All the reported bounds are for the 95% confidence intervals which have been generated from robust PSU clustered standard errors. The estimates are obtained using the STATA command `plausexog` by Clarke (2014).
This figure presents 95% confidence intervals for the estimated coefficient of IPV under the assumption that the instrumental variable has an baseline influence on employment. On the horizontal axis, I vary the baseline influence of husband's childhood experience of domestic violence on employment.

Table 2.1: Characteristics of Women in Sample (N=25,528)

Variable	All Women	IPV Victims	Non-Victims	Difference in means	P-value for t-test of diff in means
Wife worked in past 12 months	0.690	0.716	0.669	0.046	0.000
Quintile 1 of household wealth	0.200	0.184	0.213	-0.029	0.000
Quintile 2 of household wealth	0.200	0.206	0.195	0.010	0.115
Quintile 3 of household wealth	0.200	0.219	0.186	0.033	0.000
Quintile 4 of household wealth	0.202	0.214	0.193	0.020	0.003
Quintile 5 of household wealth	0.197	0.178	0.212	-0.034	0.000
Urban residence	0.766	0.792	0.746	0.046	0.000
Age of wife	33.642	32.189	34.812	-2.623	0.000
Wife's Ethnicity: No ethnicity	0.858	0.853	0.862	-0.009	0.134
Wife's Ethnicity: Indigenous	0.043	0.042	0.044	-0.002	0.483
Wife's Ethnicity: Afro-Colombian	0.099	0.104	0.094	0.010	0.044
Wife's Ethnicity: Other	0.001	0.001	0.000	0.000	0.303
Wife's education	9.090	9.047	9.124	-0.077	0.265
Wife's age at marriage	20.864	20.341	21.286	-0.945	0.000
Any children aged 6+ at home	0.590	0.567	0.610	-0.043	0.000
Any children aged 5 or less at home	0.397	0.428	0.373	0.055	0.000
Wife currently pregnant	0.044	0.045	0.043	0.003	0.383
Any childbirth in past year	0.084	0.087	0.082	0.005	0.230
Wife has final say on own health care	0.791	0.816	0.771	0.045	0.000
Wife has final say on making large household purchases	0.298	0.328	0.274	0.054	0.000
Wife has final say on making household purchases for daily needs	0.456	0.475	0.441	0.034	0.000
Husband's age	38.194	36.712	39.294	-2.582	0.000
Husband's education	10.818	10.870	10.776	0.094	0.281
Husband currently working	0.929	0.925	0.932	-0.008	0.100
Husband drinks alcohol	0.663	0.715	0.621	0.094	0.000
Wife's mother ever beaten by husband	0.359	0.417	0.312	0.105	0.000
Wife mistreated by parents in childhood	0.210	0.215	0.206	0.009	0.207
Husband mistreated by parents in childhood	0.335	0.412	0.274	0.139	0.000

Source: 2010 Colombian DHS

Table 2.2: OLS Estimates for the Likelihood of Women's Employment

Dependent Variable: Work in past 12 months	(1)	(2)	(3)
Physical and/or emotional IPV	0.044***		
	(0.008)		
Physical IPV in past 12 months		0.034***	
		(0.012)	
Emotional IPV in past 12 months			0.043***
			(0.008)
Urban Residence	0.054***	0.056***	0.054***
	(0.015)	(0.015)	(0.015)
Quintile 2 of household wealth	0.074***	0.075***	0.074***
	(0.015)	(0.015)	(0.015)
Quintile 3 of household wealth	0.079***	0.080***	0.079***
	(0.019)	(0.019)	(0.019)
Quintile 4 of household wealth	0.082***	0.084***	0.082***
	(0.020)	(0.020)	(0.020)
Quintile 5 of household wealth	0.068***	0.069***	0.068***
	(0.021)	(0.021)	(0.021)
Wife's Age Group: 26-35	0.084***	0.082***	0.083***
	(0.013)	(0.013)	(0.013)
Wife's Age Group: 36-49	0.093***	0.090***	0.092***
	(0.015)	(0.015)	(0.015)
Wife's Ethnicity: Indigenous	0.080***	0.080***	0.081***
	(0.020)	(0.020)	(0.020)
Wife's Ethnicity: Afro-Colombian	0.025*	0.026*	0.025*
	(0.014)	(0.014)	(0.014)
Wife's Ethnicity: Other	0.211*	0.220**	0.210*
	(0.109)	(0.107)	(0.109)
Wife's Education: Incomplete primary	0.015	0.013	0.014
	(0.034)	(0.034)	(0.034)
Wife's Education: Complete primary	0.034	0.033	0.033
	(0.034)	(0.034)	(0.034)
Wife's Education: Incomplete secondary	0.032	0.031	0.031
	(0.034)	(0.034)	(0.034)
Wife's Education: Complete secondary	0.104***	0.102***	0.102***
	(0.034)	(0.034)	(0.034)
Wife's Education: Higher	0.242***	0.240***	0.241***
	(0.035)	(0.035)	(0.035)
Any children aged 6+ in the household	0.005	0.005	0.005
	(0.009)	(0.009)	(0.009)
Any children aged 5 or less in the household	-0.068***	-0.068***	-0.068***
	(0.009)	(0.009)	(0.009)
Any childbirth in past year	-0.150***	-0.151***	-0.150***
	(0.014)	(0.014)	(0.014)
Husband's Age Group: 25-35	0.015	0.013	0.015
	(0.016)	(0.016)	(0.016)
Husband's Age Group: 35-49	-0.009	-0.010	-0.009
	(0.018)	(0.018)	(0.018)
Husband's Age Group: 50-65	-0.069***	-0.070***	-0.069***
	(0.022)	(0.022)	(0.022)
Husband's Age Group: 65+	-0.032	-0.034	-0.032
	(0.050)	(0.050)	(0.050)
Husband's Age Group: Unknown	0.186***	0.182***	0.187***
	(0.019)	(0.019)	(0.019)
Husband's Education: Incomplete primary	0.018	0.016	0.018
	(0.027)	(0.027)	(0.027)
Husband's Education: Complete primary	0.038	0.036	0.039
	(0.026)	(0.026)	(0.026)
Husband's Education: Incomplete secondary	0.034	0.032	0.034

Continued on next page

Table 2.2 continued

Dependent Variable: Work in past 12 months	(1)	(2)	(3)
	(0.026)	(0.026)	(0.026)
Husband's Education: Complete secondary	0.045	0.043	0.046
	(0.032)	(0.031)	(0.032)
Husband's Education: Higher	0.053*	0.049*	0.052*
	(0.028)	(0.028)	(0.028)
Husband's Education: Unknown	0.008	0.004	0.009
	(0.046)	(0.047)	(0.046)
Husband currently working	-0.033**	-0.034**	-0.033**
	(0.015)	(0.015)	(0.015)
Husband drinks alcohol	0.020**	0.023***	0.021**
	(0.008)	(0.008)	(0.008)
Constant	0.291***	0.310***	0.294***
	(0.047)	(0.047)	(0.047)
Department Fixed Effects	Yes	Yes	Yes
R-squared	0.126	0.124	0.126
Observations	21,345	21,345	21,345

Standard errors in parenthesis, clustered at the PSU level
Significant at *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Source: 2010 Colombian DHS

Table 2.3: OLS Estimation Results for the Reduced Form Relationship between Women's Work and Husband's Childhood Exposure to Domestic Violence

Dependent Variable: Work in past 12 months	(1)
Husband mistreated by parents in childhood	0.023***
	(0.008)
Constant	0.686***
	(0.006)
$R-squared$	0.001
Observations	22,668

Standard errors in parenthesis, clustered at the PSU level
Significant at *** p<0.01, ** p<0.05, * p<0.1
Source: 2010 Colombian DHS

Table 2.4: First-Stage Estimates for the Likelihood of Intimate Partner Violence

Dependent Variable: Victim of IPV in past 12 months	(1) Any IPV	(2) Physical IPV	(3) Emotional IPV
Husband mistreated by parents in childhood	0.146*** (0.010)	0.082*** (0.007)	0.127*** (0.010)
Urban Residence	0.047*** (0.015)	0.017* (0.010)	0.050*** (0.015)
Quintile 2 of household wealth	0.026* (0.015)	0.012 (0.010)	0.022 (0.016)
Quintile 3 of household wealth	0.048** (0.019)	0.019 (0.013)	0.045** (0.020)
Quintile 4 of household wealth	0.046** (0.021)	0.009 (0.014)	0.046** (0.021)
Quintile 5 of household wealth	0.016 (0.023)	-0.003 (0.014)	0.014 (0.023)
Wife's Age Group: 26-35	-0.099*** (0.015)	-0.067*** (0.011)	-0.086*** (0.015)
Wife's Age Group: 36-49	-0.166*** (0.017)	-0.101*** (0.013)	-0.149*** (0.017)
Wife's Ethnicity: Indigenous	-0.004 (0.021)	0.015 (0.015)	-0.007 (0.020)
Wife's Ethnicity: Afro-Colombian	0.038** (0.017)	0.011 (0.010)	0.040** (0.017)
Wife's Ethnicity: Other	0.116 (0.226)	-0.080*** (0.031)	0.133 (0.221)
Wife's Education: Incomplete primary	-0.060* (0.035)	-0.043* (0.024)	-0.027 (0.034)
Wife's Education: Complete primary	-0.066* (0.036)	-0.037 (0.024)	-0.025 (0.035)
Wife's Education: Incomplete secondary	-0.055 (0.036)	-0.045* (0.025)	-0.013 (0.036)
Wife's Education: Complete secondary	-0.088** (0.036)	-0.063** (0.025)	-0.045 (0.036)
Wife's Education: Higher	-0.082** (0.037)	-0.051** (0.025)	-0.043 (0.037)
Any children aged 6+ in the household	0.028*** (0.011)	0.013** (0.006)	0.023** (0.011)
Any children aged 5 or less in the household	0.021** (0.011)	0.022*** (0.007)	0.011 (0.011)
Any childbirth in past year	-0.072*** (0.016)	-0.046*** (0.011)	-0.069*** (0.016)
Husband's Age Group: 25-35	-0.047*** (0.017)	-0.023* (0.014)	-0.047*** (0.017)
Husband's Age Group: 35-49	-0.067*** (0.020)	-0.042*** (0.015)	-0.066*** (0.020)
Husband's Age Group: 50-65	-0.081*** (0.024)	-0.050*** (0.017)	-0.077*** (0.024)
Husband's Age Group: 65+	-0.099* (0.058)	-0.067*** (0.024)	-0.089 (0.057)
Husband's Age Group: Unknown	0.033 (0.023)	0.136*** (0.020)	0.014 (0.023)
Husband's Education: Incomplete primary	-0.049* (0.028)	0.004 (0.017)	-0.052* (0.028)
Husband's Education: Complete primary	-0.056** (0.029)	-0.004 (0.018)	-0.059** (0.029)
Husband's Education: Incomplete secondary	-0.037 (0.029)	0.001 (0.018)	-0.040 (0.029)
Husband's Education: Complete secondary	0.002 (0.035)	0.055** (0.024)	-0.008 (0.036)

Continued on next page

Table 2.4 continued

Dependent Variable: Victim of IPV in past 12 months	(1) Any IPV	(2) Physical IPV	(3) Emotional IPV
Husband's Education: Higher	-0.079**	-0.017	-0.074**
	(0.031)	(0.019)	(0.031)
Husband's Education: Unknown	-0.161***	-0.038	-0.159***
	(0.053)	(0.037)	(0.053)
Husband currently working	-0.011	-0.003	-0.012
	(0.018)	(0.012)	(0.018)
Husband drinks alcohol	0.088***	0.045***	0.081***
	(0.010)	(0.006)	(0.009)
Constant	0.508***	0.152***	0.451***
	(0.049)	(0.030)	(0.050)
Mean of Dependent Variable	0.447	0.146	0.415
Department Fixed Effects	Yes	Yes	Yes
First-stage F-stat	211.533	142.933	165.472
R-squared	0.019	0.015	0.015
Observations	19,085	19,085	19,085

Standard errors in parenthesis, clustered at the PSU level
Significant at *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Source: 2010 Colombian DHS

Table 2.5: 2SLS Estimates for the Likelihood of Women's Employment

Dependent Variable: Work in past 12 months	(1)	(2)	(3)
Physical and/or emotional IPV in past 12 months	0.161***		
	(0.059)		
Physical IPV in past 12 months		0.287***	
		(0.107)	
Emotional IPV in past 12 months			0.185***
			(0.068)
Urban Residence	0.046***	0.049***	0.044***
	(0.016)	(0.016)	(0.016)
Quintile 2 of household wealth	0.071***	0.071***	0.071***
	(0.017)	(0.017)	(0.016)
Quintile 3 of household wealth	0.076***	0.079***	0.076***
	(0.020)	(0.021)	(0.020)
Quintile 4 of household wealth	0.076***	0.081***	0.075***
	(0.022)	(0.022)	(0.022)
Quintile 5 of household wealth	0.068***	0.072***	0.068***
	(0.023)	(0.024)	(0.023)
Wife's Age Group: 26-35	0.089***	0.092***	0.089***
	(0.015)	(0.016)	(0.015)
Wife's Age Group: 36-49	0.104***	0.106***	0.105***
	(0.019)	(0.020)	(0.020)
Wife's Ethnicity: Indigenous	0.067***	0.062***	0.067***
	(0.022)	(0.022)	(0.022)
Wife's Ethnicity: Afro-Colombian	0.022	0.025*	0.021
	(0.015)	(0.015)	(0.015)
Wife's Ethnicity: Other	0.212	0.254**	0.206
	(0.133)	(0.119)	(0.135)
Wife's Education: Incomplete primary	0.024	0.026	0.019
	(0.038)	(0.038)	(0.039)
Wife's Education: Complete primary	0.044	0.044	0.038
	(0.038)	(0.037)	(0.038)
Wife's Education: Incomplete secondary	0.042	0.047	0.036
	(0.038)	(0.037)	(0.038)
Wife's Education: Complete secondary	0.112***	0.115***	0.106***
	(0.038)	(0.038)	(0.038)
Wife's Education: Higher	0.252***	0.253***	0.246***
	(0.039)	(0.039)	(0.039)
Any children aged 6+ in the household	-0.001	-0.000	-0.001
	(0.010)	(0.010)	(0.010)
Any children aged 5 or less in the household	-0.072***	-0.075***	-0.071***
	(0.010)	(0.010)	(0.010)
Any childbirth in past year	-0.142***	-0.141***	-0.141***
	(0.016)	(0.016)	(0.016)
Husband's Age Group: 25-35	0.022	0.021	0.023
	(0.017)	(0.018)	(0.017)
Husband's Age Group: 35-49	0.002	0.003	0.003
	(0.020)	(0.020)	(0.020)
Husband's Age Group: 50-65	-0.053**	-0.052**	-0.052**
	(0.024)	(0.024)	(0.024)
Husband's Age Group: 65+	-0.021	-0.017	-0.020
	(0.055)	(0.056)	(0.055)
Husband's Age Group: Unknown	0.187***	0.153***	0.190***
	(0.020)	(0.026)	(0.020)
Husband's Education: Incomplete primary	0.013	0.004	0.014
	(0.030)	(0.029)	(0.030)
Husband's Education: Complete primary	0.032	0.024	0.034
	(0.030)	(0.029)	(0.030)
Husband's Education: Incomplete secondary	0.031	0.025	0.032

Continued on next page

Table 2.5 continued

Dependent Variable: Work in past 12 months	(1)	(2)	(3)
	(0.030)	(0.029)	(0.030)
Husband's Education: Complete secondary	0.041	0.025	0.042
	(0.035)	(0.034)	(0.035)
Husband's Education: Higher	0.056*	0.048	0.057*
	(0.032)	(0.031)	(0.032)
Husband's Education: Unknown	-0.002	-0.017	0.001
	(0.054)	(0.053)	(0.055)
Husband currently working	-0.026*	-0.027*	-0.026
	(0.016)	(0.016)	(0.016)
Husband drinks alcohol	0.010	0.011	0.009
	(0.011)	(0.010)	(0.011)
Constant	0.240***	0.278***	0.239***
	(0.059)	(0.052)	(0.059)
Mean of Dependent Variable	0.682	0.682	0.682
Department Fixed Effects	Yes	Yes	Yes
First-stage F-stat	211.533	142.933	165.472
R-squared	0.108	0.095	0.102
Observations	19,085	19,085	19,085

Standard errors in parenthesis, clustered at the PSU level
Significant at *** p<0.01, ** p<0.05, * p<0.1
Source: 2010 Colombian DHS

Table 2.6: 2SLS Estimates for Women's Employment: Nevo-Rosen Approach with no Covariates

	OLS			2SLS: Imperfect IV			2SLS: Nevo-Rosen IV		
Dependent Variable: Work in past 12 months	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Physical and/or emotional IPV in past 12 months	0.046*** (0.007)			0.152*** (0.054)			0.015* (0.009)		
Physical IPV in past 12 months		0.064*** (0.010)			0.213*** (0.076)			0.062 (0.042)	
Emotional IPV in past 12 months			0.045*** (0.008)			0.179*** (0.064)			0.016* (0.009)
Mean of Dependent Variable	0.682	0.682	0.682	0.682	0.682	0.682	0.682	0.682	0.682
Observations	25,528	25,528	25,528	22,668	22,668	22,668	22,668	22,668	22,668

Note: No covariates included. Imperfect Instrument: Husband's Childhood Experience of Domestic Violence.
Standard errors in parenthesis. Clustered at the PSU level for OLS and IIV models. Bootstrap (200 reps.) for Rosen-Nevo instrument
Significant at *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Source: 2010 Colombian DHS

Table 2.7: 2SLS Estimates for Women's Employment: Nevo-Rosen Approach with Covariates

Dependent Variable: Work in past 12 months	OLS			2SLS: Imperfect IV			2SLS: Nevo-Rosen IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Physical and/or emotional IPV in past 12 months	0.044*** (0.008)			0.161*** (0.059)			0.019* (0.010)		
Physical IPV in past 12 months		0.034*** (0.012)			0.287*** (0.107)			0.097* (0.053)	
Emotional IPV in past 12 months			0.043*** (0.008)			0.185*** (0.068)			0.020* (0.010)
Mean of Dependent Variable	0.682	0.682	0.682	0.682	0.682	0.682	0.682	0.682	0.682
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household's Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wife's Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Husband's Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,345	21,345	21,345	19,085	19,085	19,085	19,085	19,085	19,085

Note: Imperfect Instrument: Husband's Childhood Experience of Domestic Violence.
Standard errors in parenthesis. Clustered at the PSU level for OLS and IIV models. Bootstrap (200 reps.) for Rosen-Nevo instrument.
Significant at *** p<0.01, ** p<0.05, * p<0.1
Source: 2010 Colombian DHS

Table 2.8: OLS Estimates: Relationship between Wife and Husband Characteristics, IPV and Employment

Dependent Variable: Wife worked in past 12 months	(1)	(2)	(3)	(4)	(5)	(6)
Physical and/or emotional IPV in past 12 months	0.099***	0.118**	0.134***	0.084***	0.089***	0.087**
	(0.036)	(0.046)	(0.050)	(0.033)	(0.034)	(0.039)
Wife's age at marriage	0.009***	0.010***	0.012***			
	(0.001)	(0.001)	(0.001)			
IPV × Wife's age at marriage	-0.003**	-0.004**	-0.004*			
	(0.002)	(0.002)	(0.002)			
Age Difference Husband-Wife		-0.003**	-0.002**			
		(0.001)	(0.001)			
IPV × Age Difference Husband-Wife		-0.001	-0.002			
		(0.001)	(0.002)			
Wife's education				0.024***	0.025***	0.027***
				(0.001)	(0.001)	(0.001)
IPV × Wife's education				-0.003	-0.003*	-0.003
				(0.002)	(0.002)	(0.002)
Education Difference Husband-Wife					0.003***	0.004***
					(0.001)	(0.001)
IPV × Education Difference Husband-Wife					-0.002	-0.002
					(0.002)	(0.002)
Wife's mother ever beaten by her husband			0.025**			0.028***
			(0.010)			(0.008)
Husband mistreated in childhood			0.009			0.032***
			(0.010)			(0.008)
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.040	0.045	0.023	0.074	0.075	0.061
Observations	25,528	20,595	17,994	25,528	25,090	21,711

Standard errors in parenthesis, clustered at the PSU level
Significant at *** p<0.01, ** p<0.05, * p<0.1
Source: 2010 Colombian DHS
Notes: Coefficients on Age and Age x IPV are not reported.
Columns 3 and 6 lose sample size because the wife did not report either her spouse's age or her spouse's education.

Table 2.9: OLS Estimates: Effects of Employment, Age at Marriage, and Education on Self-reported Autonomy

	Own health care			Large purchases			Daily purchases			All		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent Variable: Wife has final say												
Employment	0.104***	0.112***	0.152***	0.160***	0.040	0.006	0.098***	0.056	0.103***	0.090***	0.121***	0.174***
	(0.007)	(0.036)	(0.032)	(0.007)	(0.034)	(0.032)	(0.009)	(0.039)	(0.035)	(0.007)	(0.032)	(0.030)
Wife's age at marriage		0.004***			0.000			0.000			0.003**	
		(0.001)			(0.001)			(0.002)			(0.001)	
Employment × Wife's age at marriage		-0.002			-0.005***			-0.005**			-0.002	
		(0.002)			(0.002)			(0.002)			(0.001)	
Wife's education			0.019***			0.005***			0.015***			0.018***
			(0.002)			(0.001)			(0.002)			(0.001)
Employment × Wife's education			-0.008***			-0.004**			-0.012***			-0.009***
			(0.002)			(0.002)			(0.002)			(0.002)
Constant	0.772***	0.708***	0.567***	0.165***	0.078***	0.030	0.376***	0.143***	-0.003	0.812***	0.702***	0.564***
	(0.012)	(0.033)	(0.028)	(0.013)	(0.029)	(0.026)	(0.016)	(0.036)	(0.032)	(0.011)	(0.030)	(0.027)
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.048	0.049	0.067	0.035	0.053	0.052	0.017	0.043	0.046	0.034	0.037	0.054
Observations	25,528	25,528	25,528	25,528	25,528	25,528	25,528	25,528	25,528	25,528	25,528	25,528

Standard errors in parenthesis, clustered at the PSU level
Significant at *** p<0.01, ** p<0.05, * p<0.1
Source: 2010 Colombian DHS

Table 2.10: OLS Estimates: Effects of IPV on Work Net the Effect of Women's Decision-Making Power on Her Own Health

Dependent Variable: Work in past 12 months	(1)	(2)	(3)	(4)	(5)	(6)
Physical and/or emotional IPV in past 12 months	0.120** (0.054)					
Direct effect any IPV		0.118** (0.055)				
Physical IPV in past 12 months			0.198** (0.090)			
Direct effect physical IPV				0.194** (0.092)		
Emotional IPV in past 12 months					0.142** (0.064)	
Direct effect emotional IPV						0.139** (0.065)
Wife has final say on own health care	0.049*** (0.009)		0.049*** (0.009)		0.048*** (0.009)	
Model	2SLS	Seq. g-est	2SLS	Seq. g-est	2SLS	Seq. g-est
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.083		0.078		0.079	
Observations	17,652	17,652	17,652	17,652	17,652	17,652

Notes: For 2SLS estimation: Standard errors in parenthesis, clustered at the PSU level.
For sequential g-estimation: Bootstrap standard errors in parenthesis, clustered at the PSU level; 200 replications.
Significant at *** p<0.01, ** p<0.05, * p<0.1
Source: 2010 Colombian DHS
Pre-treatment covariates: Female's characteristics: age, ethnicity, age at first marriage, years of education
Spouse/partner's characteristics: age, years of education, work status, alcohol consumption.
Household characteristics: urban/rural area
Intermediate covariates: Female's characteristics: presence of children 6-18 y.o., presence of children less than 6 y.o., any childbirths past year. Household characteristics: wealth quintile group

Table 2.11: OLS Estimates: Effects of IPV on Work Net the Effect of Women's Decision-Making Power on Large Purchases

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Work in past 12 months						
Physical and/or emotional IPV in past 12 months	0.130** (0.054)					
Direct effect any IPV		0.130** (0.056)				
Physical IPV in past 12 months			0.215** (0.090)			
Direct effect physical IPV				0.214** (0.092)		
Emotional IPV in past 12 months					0.153** (0.064)	
Direct effect emotional IPV						0.153** (0.066)
Wife has final say on making large household purchases	0.066*** (0.009)		0.067*** (0.009)		0.065*** (0.009)	
Model	2SLS	Seq. g-est	2SLS	Seq. g-est	2SLS	Seq. g-est
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.083		0.077		0.078	
Observations	17,652	17,652	17,652	17,652	17,652	17,652

Notes: For 2SLS estimation: Standard errors in parenthesis, clustered at the PSU level.
For sequential g-estimation: Bootstrap standard errors in parenthesis, clustered at the PSU level; 200 replications.
Significant at *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Source: 2010 Colombian DHS
Pre-treatment covariates: Female's characteristics: age, ethnicity, age at first marriage, years of education
Spouse/partner's characteristics: age, years of education, work status, alcohol consumption.
Household characteristics: urban/rural area
Intermediate covariates: Female's characteristics: presence of children 6-18 y.o., presence of children less than 6 y.o., any childbirths past year. Household characteristics: wealth quintile group

Table 2.12: OLS Estimates: Effects of IPV on Work Net the Effect of Women's Decision-Making Power on Daily Purchases

Dependent Variable: Work in past 12 months	(1)	(2)	(3)	(4)	(5)	(6)
Physical and/or emotional IPV in past 12 months	0.132** (0.054)					
Direct effect any IPV		0.130** (0.056)				
Physical IPV in past 12 months			0.217** (0.090)			
Direct effect physical IPV				0.214** (0.092)		
Emotional IPV in past 12 months					0.155** (0.064)	
Direct effect emotional IPV						0.153** (0.066)
Wife has final say on making household purchases for daily needs	0.019** (0.008)		0.020*** (0.008)		0.019** (0.008)	
Model	2SLS	Seq. g-est	2SLS	Seq. g-est	2SLS	Seq. g-est
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.080		0.074		0.075	
Observations	17,652	17,652	17,652	17,652	17,652	17,652

Notes: For 2SLS estimation: Standard errors in parenthesis, clustered at the PSU level.
For sequential g-estimation: Bootstrap standard errors in parenthesis, clustered at the PSU level; 200 replications.
Significant at *** p<0.01, ** p<0.05, * p<0.1
Source: 2010 Colombian DHS
Pre-treatment covariates: Female's characteristics: age, ethnicity, age at first marriage, years of education
Spouse/partner's characteristics: age, years of education, work status, alcohol consumption.
Household characteristics: urban/rural area
Intermediate covariates: Female's characteristics: presence of children 6-18 y.o., presence of children less than 6 y.o.,
any childbirths past year. Household characteristics: wealth quintile group

Table 2.13: OLS Estimates: Effects of IPV on Work Net the Effect of Women's Decision-Making Power on Own Health, Large Household Purchases and Daily Household Purchases

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Work in past 12 months						
Physical and/or emotional IPV in past 12 months	0.118** (0.055)					
Direct effect any IPV		0.116** (0.056)				
Physical IPV in past 12 months			0.194** (0.090)			
Direct effect physical IPV				0.190** (0.092)		
Emotional IPV in past 12 months					0.139** (0.064)	
Direct effect emotional IPV						0.136** (0.066)
Wife has final say on various dimmensions	0.057*** (0.010)		0.058*** (0.010)		0.057*** (0.010)	
Model	2SLS	Seq. g-est	2SLS	Seq. g-est	2SLS	Seq. g-est
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.084		0.079		0.080	
Observations	17,652	17,652	17,652	17,652	17,652	17,652

Decision making power equals 1 if the wife has final say on at least one of own health, large household purchases and daily household purchases.
Notes: For 2SLS estimation: Standard errors in parenthesis, clustered at the PSU level.
For sequential g-estimation: Bootstrap standard errors in parenthesis, clustered at the PSU level; 200 replications.
Significant at *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Source: 2010 Colombian DHS
Pre-treatment covariates: Female's characteristics: age, ethnicity, age at first marriage, years of education
Spouse/partner's characteristics: age, years of education, work status, alcohol consumption.
Household characteristics: urban/rural area
Intermediate covariates: Female's characteristics: presence of children 6-18 y.o., presence of children less than 6 y.o., any childbirths past year. Household characteristics: wealth quintile group

Table 2.14: OLS Estimates: Effects of IPV on Work Net the Effect of Women's Willigness to Separate in Past Year

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Work in past 12 months						
Physical and/or emotional IPV in past 12 months	0.118 (0.090)					
Direct effect any IPV		0.077 (0.055)				
Physical IPV in past 12 months			0.203 (0.156)			
Direct effect physical IPV				0.111 (0.091)		
Emotional IPV in past 12 months					0.143 (0.110)	
Direct effect emotional IPV						0.089 (0.064)
Considered separating from husband in past 12 months	0.016 (0.033)		0.012 (0.036)		0.013 (0.035)	
Model	2SLS	Seq. g-est	2SLS	Seq. g-est	2SLS	Seq. g-est
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.082		0.075		0.077	
Observations	17,652	17,652	17,652	17,652	17,652	17,652

Notes: For 2SLS estimation: Standard errors in parenthesis, clustered at the PSU level.
For sequential g-estimation: Bootstrap standard errors in parenthesis, clustered at the PSU level; 200 replications.
Significant at *** p<0.01, ** p<0.05, * p<0.1
Source: 2010 Colombian DHS
Pre-treatment covariates: Female's characteristics: age, ethnicity, age at first marriage, years of education
Spouse/partner's characteristics: age, years of education, work status, alcohol consumption.
Household characteristics: urban/rural area
Intermediate covariates: Female's characteristics: presence of children 6-18 y.o., presence of children less than 6 y.o.,
any childbirths past year. Household characteristics: wealth quintile group

Chapter 3

Inequality of Opportunity in Adult Health in Colombia

Published in The Journal of Economic Inequality[1]

3.1 Introduction

The 2006 World Development Report on Equity and Development highlights that health is not only an important dimension of welfare, but that inequality in health often reinforces and reproduces over time inequality in domains such as income, education or labor (The World Bank, 2005). The traditional focus of policies that aim to reduce health inequity in both developed and developing countries is the reduction of inequality in specific health outcomes as well as in access to health care services and health insurance. Differences in opportunities driven by individual characteristics such as gender, ethnicity or place of origin have not received such consideration. However, they seem to play a key role in determining how health inequality reproduces over time and across generations. For that reason, the study of alternative policies to reduce health inequality has led to an increasing interest in the equality of opportunity literature and its empirical application to health equity (Rosa Dias and Jones, 2007; Fleurbaey and Schokkaert, 2009; Rosa Dias, 2009; Jusot, Tubeuf, and Trannoy, 2010; Donni, Peragine, and Pignataro, 2014).

Roemer (1998)'s theoretical approach to equality of opportunity is based on the idea that the sources of an individual's desirable outcome, like good health, can be separated into circumstances and efforts. Circumstances are factors that are beyond an individual's control and inequalities emerging from such circumstances should be compensated. Conversely, effort is affected by individual choice and inequalities arising from different efforts are morally and normatively acceptable. The most important implication derived from the equality of opportunity approach is that an equal-opportunity policy should aim to provide everyone with the same opportunity to achieve or enjoy an excellent outcome. A social planner, therefore, would seek to equalize opportunities rather than outcomes and would allow individuals to be fully responsible for their own choices and final results.

Inequality of opportunity, from a theoretical stance, rests on two principles: the compensation principle and the reward principle (Ramos and Van de Gaer, 2015). The com-

pensation principle indicates that inequalities due to circumstances must be compensated, whereas the reward principle indicates that individual efforts must be rewarded. The ex-ante approach to compensation suggests that equality of opportunity holds as long as all individuals face the same opportunities, regardless of each one's circumstances. Under this approach, the observation of all possible efforts is not required for empirical analysis as inequality of opportunity can then be studied focusing on the outcome distributions for different sets of circumstances.

Following an ex-ante approach, inequality of opportunity in adult health has been studied mainly in the context of developed countries. For instance, Rosa Dias (2009) finds that about 21% of health inequality in adulthood, for a cohort of British individuals born in 1956, is related to circumstances in childhood such as maternal education, spells of financial difficulties, as well as poor health and obesity in childhood. The empirical analysis developed in this chapter is also grounded on Trannoy et al. (2010) and Donni, Peragine, and Pignataro (2014). Trannoy et. al study inequality of opportunity among French adults and suggest that such inequality might be halved if the effects of individual circumstances were removed. Donni, Peragine and Pignataro, in contrast to Rosa-Dias, apply an alternative empirical approach to data from various waves of the British Household Panel Survey and estimate that about 30% of adult health inequality is due to circumstances. For developing countries, the literature is very scarce. For instance, Jusot, Mage, and Menendez (2014) study inequality of opportunity in adult health in Indonesia. The authors construct a synthetic index of global health status using information on biomarkers and self-reported health. Their most striking finding is that the existence of long-term inequalities in adult health is related mainly to variables that indicate a sense of community such as religion and language spoken.

This chapter fits in this line of research. Specifically, I address the following research question: among the set of observed circumstances, which particular earlylife circumstances have a salient long-term association with observed inequality of opportunity in

adult health Colombia as a whole, and in both rural and urban areas of the country? This study is among the first to answer this question using data from a developing country.

Colombia is undergoing rapid demographic changes. The Colombian population predominantly lives is urban areas, is aging as life expectancy at birth has increased from 65 to 75 years in the last 35 years, and a fertility rate decreasing from 4.0 in 1980 to 2.0 births per woman in 2015. Additionally, health outcomes appear to be worse in rural areas than in urban areas. Health status varies greatly between rural and urban residents: 32% of the rural population reports a poor or fair health status whereas 22% of the urban population reports a similar status. It is worth noting that access to health care services has considerably increased in the country. The The World Health Organization (2006) reports that the Colombian health system achieved 96% coverage of the population in 2013. Yet, some important differences persist between urban and rural areas. Findings from a few studies (Restrepo et al., 2007; Flórez et al., 2007) suggest that the area of residence is an important determinant of the use of health services in Colombia. Differential health care use between urban and rural residents may reflect both a major difficulty in securing the availability of health care providers in rural areas and a large concentration of private health care providers in urban areas (Vargas-Lorenzo, 2009). Besides important differences in the density of medical care access or income, exposure to different childhood circumstances may still play an important role in adult health outcomes currently observed in urban and rural areas.

I use data from the 2010 Colombian Living Standards and Social Mobility Survey, a rich dataset that provides retrospective information about individual childhood. In the empirical analysis, I use first-order stochastic dominance analysis to provide a weak test of inequality of opportunity in the conditional distributions of self-assessed health status, following Lefranc, Pistolesi, and Trannoy (2009). I also compute a dissimilarity index and a Gini-opportunity index as direct measures of inequality of opportunity (Paes De Barros, Vega, and Saavedra, 2008; Paes De Barros et al., 2009; Rosa Dias, 2009). I then use

the Shapley-value decomposition to calculate the specific contribution of childhood circumstances such as parental education and household socioeconomic status at age 10 to inequality of opportunity. The findings suggest that 8% to 10% of the circumstance-driven opportunities distinctively enjoyed by those who are healthier should be compensated for or redistributed among those who are less healthy in order to achieve equality of opportunity. Differences in household socioeconomic status during childhood and parental educational attainment appear to be the most important dimensions of inequality of opportunity in adult health. Household socioeconomic status at age 10 contributes between 15% and 22% to the dissimilarity index, whereas parental education between 10% and 13%. In contrast, Jusot, Mage, and Menendez (2014) suggest ethnicity and region of birth are more important factors for health inequity in Indonesia.

The remaining of the chapter is organized as follows. Section 2 describes the 2010 Living Standards and Social Mobility Survey and provides some descriptive statistics. Section 3 explains the empirical methods. Estimation results are presented in Section 4. Section 5 provides a discussion of the limitations of this study and concluding remarks.

3.2   Data

The main data source is the 2010 Colombian Living Standards and Social Mobility Survey (LSSM – Encuesta de Calidad de Vida y Movilidad Social) carried out by the Colombian Bureau of Statistics (Departamento Administrativo Nacional de Estadistica – DANE.) This survey provides current and retrospective measures of socioeconomic characteristics. The LSSM is representative for the entire country, urban and rural areas, and for nine different subnational regions.[2] The LSSM includes recall questions on living conditions when the respondent was 10 years old. This set of questions provides information on parental

---

[2]The regions are: Atlantic, Eastern, Central, Pacific, Orinoquia-Amazonia, Antioquia, Valle del Cauca, San Andrs and Providencia, and Bogota. Rural areas in the regions of Orinoquia-Amazonia and San Andrs and Providencia were not surveyed due to prohibitive costs and poor road access.

educational attainment and ownership of durable assets during childhood. The social mobility module in the LSSM only considers heads of household who are between 25 and 65 years old. The sample design ensures that the final sample of 2,253 individuals represents about 9.57 million heads of household in Colombia. Table 3.1 displays a summary of descriptive statistics for the full sample. Table A.1 and Table A.2 in Section A.1 in the Appendix C show the summary statistics for the urban and rural sub-samples.

The outcome of interest is health status in adulthood. It is measured by self-assessed health status, which has been demonstrated to be effective in predicting mortality (Idler and Benyamini, 1997; Van Doorslaer and Gerdtham, 2003) and health care utilization (De Salvo et al., 2005). In the survey, individuals rank their health as either poor (1), fair (2), good (3) or excellent (4) when answering the question "*In general, how do you rate your health status?.*" Around 73% of the respondents reported a good or an excellent health status whereas 2.2% reported a poor health status. By area, 78% of urban residents reported at least a good health status whereas 68% of rural residents reported a similar status.

Self-reported health status has some limitations that have been previously identified in the health literature (Jusot, Mage, and Menendez, 2014). The first limitation is that sub-groups of the population may use different thresholds and reference points when assessing their health status, although their objective health conditions are probably the same, leading to a problem known as reporting bias. The second limitation is the lack of cardinality and continuity of the self-assessed health status variable. This problem proves difficult for the use of standard inequality measures.

The set of *early-life circumstances* includes parental educational level and household socioeconomic status at age 10. Parental educational attainment is a categorical variable that indicates whether a parent completed or not a specific level (primary school, secondary school or higher education). In this sample, approximately 60% of the heads of household reported that their parents did not attend school or did not complete primary

education. In contrast, less than 9% indicated that their parents completed secondary school or a higher education level. In urban areas, 46% of fathers and 51% of mothers did not complete primary education. In rural areas, the%ages for incomplete primary education are even higher: 54% for fathers and 62% for mothers.

Household socioeconomic status at age 10 is a categorical variable that indicates the quintile group in which a household falls into, based on an asset index following the methodology by Vyas and Kumaranayake (2006).[3] For the full sample, about 25%[4] of the heads of household are assigned to the first quintile group of the socioeconomic index, according to their reports of assets ownership.[5] In urban areas, each of the five quintile groups has approximately the same number of individuals. In rural areas, in contrast, 25% of individuals belong in the first quintile group. Retrospective data are far from ideal and measurement error and recall bias could be problematic, in particular when income or earnings data are asked. It is still possible to argue that the variables for assets ownership could be remembered with some reasonable accuracy despite observing longer recall intervals for older adults, as suggested by Angulo et al. (2012).

Other variables that are likely to affect individual health status are also considered. In the set of *demographic controls*, I include ethnicity, urban or rural location of birth, and region of birth. About 9% of heads of household reported being a member of an ethnic minority. Indigenous minorities are mostly located in rural areas, in contrast with African-Colombian minorities who are uniformly distributed between urban and rural areas..[6] Regarding location of birth, about 28.4% of current residents in urban areas

---

[3]Variables in the socioeconomic status index include type of floor materials, source of water supply, type of toilet available, availability of electricity, and ownership of appliances like washing machine, vacuum cleaner, refrigerator, gas or electric stove, gas or electric oven, television set, as well as ownership of dwelling, automobile, or motorcycle.

[4]Quintiles of the wealth index do not contain equal numbers of individuals, since many respondents in rural areas have the same or very similar index scores in the lower part of the distribution.

[5]One potential concern that arises from the use of these data is the recall nature of the early-life circumstances. A threat to this analysis comes from the possibility that the information reported is less accurate for longer recall intervals, in particular, for older adults regarding assets ownership in their childhood.

[6]The choice between ethnicity and region is not of particular concern here. The correlation between these variables is low. Predicting ethnicity from region of birth, or vice versa, gives a variance inflation factor of 1, which is well below the rule of thumb of 10.

were born in rural areas, with the younger urban cohorts exhibiting a smaller proportion of rural-born adults. There are substantial socio-economic differences between regions within the country. The World Bank (The World Bank, 2015)(pp.45) documents the main regional differences in growth and inequality, which show, in particular, a Gini coefficient of inequality across regions of 0.3 and a large per capita income gap with Bogota. Throughout the analysis, additional controls include gender and age group. In the full sample, about 71% of household heads are males. The proportion of male household heads is larger in urban (79%) than rural areas (64%).

The LSSM does not provide information on individual or parental health-related behaviors. The only circumstance in the data that is partly affected by individual effort is years of education. Educational attainment is an important variable in the analysis of health inequality, as it has been shown to have a positive and large association with health status (Lleras-Muney, 2005; Arendt, 2005; Cutler, Lleras-Muney, and Vogl, 2008). The average number of years of education of the heads of household in this sample is seven years, being larger at 8.4 years for the youngest cohort (25–35 years of age).

3.3   The Measurement of Inequality of Opportunity in Adult Health

This section explains the parametric approach used to test for inequality of opportunity following Paes De Barros et al. (2009).[7] I obtain direct estimates of inequality of opportunity, controlling for age and gender, using a non-linear model for health status. The predicted probability of reporting at least a good health status is used to calculate a dissimilarity index. The index is then decomposed using the Shapley-value. The decomposition measures the contribution of each circumstance to the observed inequality of opportunity in adult health. To provide an alternative measure of inequality of oppor-

---

[7]Section A.2 in the Appendix C provides a theoretical framework and estimates of inequality of opportunity using a stochastic dominance approach (Lefranc, Pistolesi, and Trannoy, 2009). The empirical tests follow the methodology proposed by (Yalonetzky, 2013).

tunity, I also calculate a Gini-Opportunity Index.

3.3.1 Parametric Model of the Relationship between Health Status and Early-Life Circumstances

The predicted probability of achieving a good or excellent health status is obtained after the estimation of a logit model in which the dependent variable is the dichotomous health status indicator previously defined. Thereafter, I use the predicted probability to calculate the dissimilarity index. This procedure is performed for the entire sample, and for the sub-samples of urban and rural residents.

First consider a health production function such as

$$\mathbf{H} = f(C,\ D, e, u) \tag{3.1}$$

where C is a vector of individual circumstances, D a vector of demographic controls and e a vector of effort. The residual term u captures luck and other random factors that are not measured by the other variables in the health production function.

Efforts can also be affected by individual circumstances and in most cases are unobserved. In Roemer's definition of equality of opportunity, efforts are assumed orthogonal to circumstances. This assumption suggests that any other determinant of health status that is correlated with circumstances is also understood as a circumstance. One of such variables is educational attainment.

This relationship can be empirically approximated using a non-linear specification.

$$\Pr\left[H^* = 1 \mid C_i, D_i\right] = \frac{\exp\{d + C_i a + D_i b\}}{1 + \exp\{d + C_i a + D_i b\}} \tag{3.2}$$

where H^* denotes a dichotomous health outcome for individual i, C_i the vector of individual circumstances, and D_i demographic characteristics.

The following circumstances are observed in the 2010 LSSM data: ethnicity (E), father's highest educational level (FE), mother's highest educational level (ME), quintile groups of household socioeconomic status index during childhood (WS), urban or rural area of birth (LB), and region of birth (RB). The only circumstance partly affected by individual choice that is observed in the dataset is years of education (ED). Demographic controls include gender (M) and age group (AG). Therefore, $C_i \equiv \{E_i, FE_i, ME_i, WS_i, LB_i, RB_i, ED_i\}$ and $D_i \equiv \{M_i, AG_i\}$.

To estimate the global effect of observed circumstances on health status, I also clean years of education of any influence coming from the other observed circumstances. In a related study, Trannoy et al. (2010) proposed a two-step procedure to estimate the correlation of circumstances and health status in a non-linear model. The first step involves the estimation of the residuals from an auxiliary regression of each of the circumstance variables affected by individual effort on the full set of observed circumstances. In the second step, these residuals are included in the estimable health status equation along with the same vector of observed circumstances. Trannoy et al. emphasize that the residuals from step one represent effort, luck and unobserved circumstances that allow an individual to reach a higher education level, for a given vector of observed circumstances. In this chapter, I adopt Trannoy et al. (2010)'s empirical strategy.

The logistic regression model now takes the following form:

$$
\begin{aligned}
\Pr\left[H^* = 1 \mid C'_i, \hat{\varepsilon}^e_i, D_i\right] &= \frac{\exp\{d + C'_i a_1 + \hat{\varepsilon}^e_i a_2 + D_i b\}}{1 + \exp\{d + C'_i a_1 + \hat{\varepsilon}^e_i a_2 + D_i b\}} \\
\Pr\left[H^* = 0 \mid C'_i, \hat{\varepsilon}^e_i, D_i\right] &= 1 - \Pr[H^* = 1 \mid C'_i, \hat{\varepsilon}^e_i, D_i]
\end{aligned}
\tag{3.3}
$$

where $C'_i \equiv \{E_i, FE_i, ME_i, WS_i, LB_i, RB_i\}$. Vector C_i includes years of education,

whereas vector C_i' does not.

The logistic regression model now contains the term $\hat{\varepsilon}_i^e$, which corresponds to the residuals obtained from the OLS estimation of the following model:

$$ED_i = k + C_i'g + D_iw + \varepsilon_i \qquad (3.4)$$

where ε_i is a disturbance assumed to be normally distributed.

By construction, the residuals $\hat{\varepsilon}_i^e$ are orthogonal to circumstances in the equation for health status and represent the share of individual educational attainment explained by individual responsibility, luck and unobserved characteristics and circumstances, for the given vector of observed circumstances, as shown by Trannoy et al. (2010).

My interest is to gauge what circumstances are more correlated with the health status reported by residents in rural areas and respondents living in urban areas. Therefore, I estimate logistic regression models for the subsample of individuals residing in rural areas and the subsample of individuals residing in urban areas using similar specifications to those presented in equations 3.2, 3.3 and 3.4.[8] Note that I do not perform this analysis for the full sample controlling for a dichotomous variable that indicates current urban or rural residence status, because current residence is considered an effort variable in Roemer's framework that may not be controlled for in the ex-ante approach followed in this chapter.

One contribution of this study comes from the estimation of equations 3.3 and 3.4. I provide suggestive evidence regarding the possible transmission channels of health inequalities by defining whether the effect is direct or indirect. For instance, if the estimated coefficient on a particular circumstance is only statistically significant in the estimation of

---

[8]I retain both significant and insignificant coefficients in the estimation of the dissimilarity index, following Paes De Barros, Vega, and Saavedra (2008)

the education equation but not so in the estimation of the health status equation, then it can be argued that the circumstance has an indirect effect. That is, the circumstance only has an effect on self-reported health through its effect on education. Alternatively, if the coefficient on a circumstance is significant in the health status equation only, then it can be argued that the effect is direct. Note that a circumstance may also have both direct and indirect effects. In my view, this type of analysis is consistent with the transmission channels proposed by Trannoy et al. (2010). More specifically, the authors suggest that human capital investments during childhood and the transmission of parental socioeconomic status have an indirect influence on health status in adulthood, whereas a specific risk that takes place during childhood has a direct influence on adult health following a latency period.

3.3.2 The Dissimilarity Index of Inequality of Opportunity

The calculation of the dissimilarity index first requires the estimation of a logistic regression model to obtain the predicted probability of achieving a good or excellent health status ($\hat{p}_i$). In the LSSM sample, 2.2% of the respondents report a poor health status (category 1) whereas 7.1% report an excellent health status (category 4.) For the subsequent analysis, I group the two lower categories (1 and 2) and the two upper categories (3 and 4) to define a dichotomous variable which equals 0 if the respondent reports a poor or fair health status, and equals 1 if the respondent reports a good or excellent health status.

I measure inequality of opportunity using the dissimilarity index, which has been used in inequality analysis using binary outcomes (Paes De Barros et al., 2009; Paes De Barros, Vega, and Saavedra, 2008). The dissimilarity index is a measure proportional to the absolute distance between the distribution of circumstances among those with high outcomes (i.e., excellent health) and the distribution among those with low outcomes (i.e., poor health.)

Paes De Barros, Vega, and Saavedra (2008) show that a consistent estimator for the dissimilarity index for binary outcomes is given by

$$\hat{D} = \frac{1}{2\overline{p}} \sum_{i=1}^{n} w_i |\widehat{p}_i - \overline{p}| \tag{3.5}$$

where $\widehat{p}_i$ is the predicted probability of achieving a good or excellent health status for individual $i=1,\ldots,n$. The estimated conditional probability is $\overline{p} = \sum_{i=1}^{n} w_i \widehat{p}_i$, where w_i denote sampling weights.

The dissimilarity index of inequality of opportunity can be interpreted as the minimum fraction of the number of healthier persons that need to be redistributed across circumstance groups in order to achieve equal opportunity, that is, an equal proportion of less healthy persons in all circumstance groups Paes De Barros, Vega, and Saavedra (2008).[9] The index ranges from 0 to 1, with 0 indicating a situation with equality of opportunity.

Paes De Barros et al. (2009) and Yalonetzky (2012) show that the dissimilarity index for binary outcomes satisfies some important properties of inequality indexes. First, the index equals 0 if the conditional distributions of health given circumstances are identical (that is, perfect between-type equality in access to opportunities), and equals 1 when one individual always attains an excellent health status while others do not. Second, the dissimilarity index is scale-invariant, so that rescaling the outcome by some scalar does not alter the index. Third, the index exhibits anonymity as it does not vary when individuals switch between two dichotomous states of health status. Fourth, the index is invariant to population replication. Fifth, the dissimilarity index is insensitive to balanced increases in opportunities, which suggests that the index does not change when the predicted probability of achieving a better health status increases for each type in such a way that the

---

[9]An alternative interpretation: the index indicates the%age of available opportunities for enjoying a better health status that need to be reallocated from the adults who are healthier to the adults who are less healthy, in order to achieve equality of opportunity.

original distribution is preserved. That is, the index is insensitive to transfers of opportunities between circumstance groups that are above or below the average population achievement because the balanced increases do not alter the proportion of the population in each type or the proportion of the population enjoying an excellent health status.

Ersado and Aran (2014) also show that the index can only increase when new circumstances are added. Elaborating on the last property, Ferreira and Gignoux (2014) show that the measure of inequality of opportunity obtained with a set of observed circumstances is a lower bound on the true inequality of opportunity that would be captured if the full vector of circumstances was observed.

3.3.3 Gini-Opportunity Index

In order to provide a measure of inequality of opportunity that is sensitive to transfers of opportunities between circumstances (Lefranc, Pistolesi, and Trannoy, 2009), I calculate a Gini-opportunity index. This index computes the weighted sum of all the differences among areas of opportunity sets and then divides that sum by the mean outcome of the entire population.

The Gini-opportunity index has been applied to the study in health inequalities by Rosa Dias (2009). The index was first proposed by Lefranc, Pistolesi, and Trannoy (2009) to quantify the Gini index for each type G_c, so that the opportunity set for each type is denoted by $\bar{h}_c(1 - G_c)$, where $\bar{h}_c$ represents the average health outcome for type c. Rosa Dias (2009) then defines the Gini-Opportunity index in health for k types as:

$$G_{opp} = \frac{1}{\bar{h}} \sum_{i=1}^{k} \sum_{i<j} p_i p_j [\bar{h}_j (1 - G_j) - \bar{h}_i (1 - G_i)] \tag{3.6}$$

where $\bar{h}$ denotes the mean of the health distribution, p the population share, G the Gini coefficient, and i the set of circumstances.

Lefranc, Pistolesi, and Trannoy (2009) show that the index is bounded between 0 and 1, and that it satisfies almost all of the required properties of inequality indexes. The index, in particular, is not invariant to the scale in which the health outcome is measured. The most salient limitation is that the index, as currently applied, does not account for the ordinal nature of the health status measure. Moreover, the Gini opportunity index is shown to be highly sensitive to the number of types considered by the researcher (Rosa Dias, 2014).

3.3.4 Decomposition of the Dissimilarity Index through the Shapley Value

The Shapley value decomposition allows estimating what circumstances correlate the most with the observed inequality of opportunity. The Shapley value is a central solution concept in cooperative game theory and has been extended to inequality analysis by Shorrocks (2013). I follow the methodology of Hoyos Suarez and Narayan (2012) to perform the decomposition. These authors explain that the change in inequality that arises when a new circumstance is added to a set of circumstances depends on the sequence of inclusion of the different circumstance variables. The contribution of each circumstance is measured by the average change in inequality over all possible inclusion sequences. Formally, the change in the dissimilarity index when circumstance c is added to a subset M of circumstances is given by

$$\triangle D_c = \sum_{M \subset C \setminus \{c\}} \frac{|m|\,!\,(\kappa - |m| - 1)\,!}{!} \left[D(M \cup \{c\}) - D(M) \right] \tag{3.7}$$

where C denotes the entire set of κ circumstances, and M is a subset of C that includes m circumstance variables except c. $D(M)$ is the dissimilarity index for the subset M and $D(M \cup \{c\})$ is the index obtained after adding circumstance c to subset M.

Let $D(\kappa)$ be the dissimilarity index for the set of κ circumstances. Therefore, the

contribution of circumstance κ to $D(\kappa)$ is defined by

$$S_c = \frac{\triangle D_c}{D(\kappa)} \tag{3.8}$$

where $\sum_{i \in C} S_i = 1$

As a result, I have an additive decomposition of the dissimilarity index that measures the contribution (in terms of correlation, not causation) of each circumstance to observed health inequality.

3.4 Results

This section first presents a brief summary of the results obtained using non-parametric statistic tests for stochastic dominance.[10] Lefranc, Pistolesi, and Trannoy (2009) propose a criterion to assess inequality of opportunity using stochastic dominance, and show that inequality of opportunity is satisfied if and only if the distributions of health status conditional on different sets of circumstances can be ordered by first-order stochastic dominance (Please see section A.2 in the Appendix C for further details of the test.) A non-parametric test suitable for categorical variables was introduced by Yalonetzky (2013), and I provide here an extension to assess inequality of opportunity in adult health.

I then examine the estimation results of the logistic regression model for the correlates of self-assessed health status, as well as the calculation and decomposition of the dissimilarity index of inequality of opportunity. I also provide an estimation of the Gini opportunity index, a measure that is sensitive to transfers of opportunities between circumstances, in contrast to the dissimilarity index.

---

[10]Please see section A.2 in the Appendix C for further details.

3.4.1 Stochastic Dominance Tests

In the LSSM data, health status is an ordinal variable which takes on values h=1, 2, 3, 4. Responses to the health status question concentrate in categories 2 (fair) and 3 (good). Thus, for the stochastic dominance analysis, I group the lower two categories together (1 and 2) to define a new categorical variable which equals 1 if the respondent reports a poor or a fair health status, and equals 2 and 3 if the respondent reports a good and an excellent health status, respectively.

In order to compare the conditional distributions of health status, I rely on a non-parametric test proposed by Yalonetzky (2013). This test is implemented for every pair of categories within a variable of interest. In this subsection, the variables of interest are parental and maternal educational attainment and socioeconomic status at age 10.

The test results, summarized in Table 3.2, firstly show that the health distribution for the fifth quintile group of socioeconomic status at age 10 dominates the health distribution for all but the first quintile group (comparing the fifth and first quintile groups, the z_k^l statistics are all larger than -1.96, for a confidence level of 95%) and that the fourth quintile group dominates the distribution for the first and second socioeconomic status quintile groups (the z_k^l statistics are smaller than -1.96, for a confidence level of 95%). These dominance relationships are statistically significant at the 5% level. In urban areas, I find that the health distribution for the fifth quintile group dominates each of the distributions for the four remaining quintile groups. In contrast with the urban sample, the statistical tests results for rural areas suggest that the only statistically significant dominance relationship is that of the health distribution for quintile group 5 relative to the first and second quintile groups.

Concerning parental education, Table 3.2 (panel b and panel c) suggests that the higher the levels of paternal and maternal education, the better health opportunities are, especially, in urban areas. The distribution of the health status of individuals whose fathers

have some degree of education dominates the health distribution of individuals whose fathers have no education at all, which is suggestive of inequality of opportunity. These results also suggest that there is inequality of opportunity in adult health after comparing the health distribution of individuals whose mothers attained more than secondary education relative to individuals whose mothers attained no more than some primary education.

3.4.2 Estimation Results from the Logistic Regression Model for Health Status

The calculation of the dissimilarity index first requires the estimation of a logistic regression model since health status is defined as a binary outcome. In this subsection, I briefly describe the estimation results in order to suggest the potential direction of the association between reporting at least a good health status and the observed early-life circumstances.

I first examine the results obtained from the estimation of Equation 3.4, where the variable for individual years of education is cleaned from the effect of circumstances. Note that the coefficients reported in Table 3.3 on household socioeconomic status at age 10 and parental education are all statistically significant at the 5% level. In particular, the coefficient on socioeconomic status is positive, increasing with quintile group. This result suggests how relevant is the capacity of richer households to make more investments in the education of their children. A similar relationship is found for higher education levels attained by both parents. These two results hold for the urban and rural sub-samples also.

Considering the remaining individual characteristics in the estimation of the correlates of years of education, being male and born in the Central region is positively associated with higher educational attainment in the urban subsample, while the opposite is observed in rural areas. There is an important cohort effect in educational attainment in Colombia: younger cohorts in rural areas have had better access to primary and secondary schooling in the past thirty years. A similar trend was documented for Guatemala

and other developing countries by **?**(Chapter 6).

3.4.2.1   Correlates of health status in the full sample

The first two columns in Table 3.4 display the estimation results of the logistic regression model for the full sample. In column 1, the results correspond to the estimation of the model controlling for years of education as an additional circumstance (as given in Equation 3.2). In this sample, on average, males are more likely to report a good health status than females. The estimated correlation between an individual's educational attainment, measured in years of education, and reporting a good adult health status is positive and highly significant. The coefficient on the age-group variables is negative, statistically significant, and increasing with age. The effect of parental education is positive but not significant, with or without the inclusion of own years of education. Regional differences are slightly important. Being born in the Pacific or Bogota has a negative effect on perceived health status, with the Atlantic and San Andres islands being the reference region. No significant difference is observed by area of birth.

Column 2 in Table 3.4 presents the results for the binary logistic regression model controlling for years of education purged from the effect of the other observed circumstances (as given in Equation 3.3.) The variable for years of education purged from circumstances has the same point estimate and standard error as years of education, by construction. Controlling for the correlation between years of education and the circumstance variables, does not change the direction of the basic relationships described in the previous paragraph, except for socioeconomic status during childhood, which becomes highly significant and increasing with the quintile group of household wealth at age 10. Cleaning years of education from the influence of the observed circumstances allows obtaining significant and positive coefficient estimates for almost all quintile groups of the

socioeconomic status variables.

3.4.2.2   Correlates of Health Status in the Rural and Urban Subsamples

Table 3.4 also presents the estimation results for urban and rural areas. Regarding the results for the urban sub-sample (columns 3 and 4), I find that early life circumstances like household socioeconomic status and parental education have a significant effect on the likelihood of reporting at least a good health status, although the relationship is not very strong. In particular, when I purge years of education from the influence of observed circumstances, I find a positive relationship between reporting a good health status and coming from the fifth quintile group of the socioeconomic status variable.

Regarding the effect of parental education, individuals whose fathers attained no more than some years of secondary education are also more likely to report a good health status, relative to those individuals whose fathers did not complete primary education. In the case of maternal education, the only significant and positive association to better health status is that of mothers having completed secondary education or more, relative to mothers with no education or some years of primary education.

Using the sample for rural residents, I only find a positive and significant relationship between reporting a good health status and high socioeconomic status during childhood, only in the comparison of quintile groups 3, 4 and 5 against quintile group 1, which is the excluded category (columns 5 and 6.) Considering the region of birth, being born in the Eastern, the Pacific, or Antioquia has a negative effect on self-assessed health status, relative to those born in the Atlantic and San Andres islands.

I now turn to the discussion on the potential transmission channels of health inequalities in adulthood. In what follows, I refer to the results presented in Tables 3.3 and 3.4. Parental socioeconomic status and parental education attainment have both direct and indirect effects through the effect of education on self-reported health. Note that being

born in urban areas has an indirect effect, through educational attainment.

The estimated results for the sample of urban residents also support that parental socioeconomic status and parental education have both a direct and an indirect effect. In contrast, in rural areas, the effect of parental socio-economic status and parental education is realized through years an education, which is an indirect effect.

3.4.3 Dissimilarity Index of Inequality of Opportunity and the Gini-Opportunity Index

I use the predicted probabilities from the estimation of the logistic regression models, given by Equations 3.3 and 3.4, to calculate the dissimilarity index. Table 3.5 displays the index value as well as its decomposition for the full sample, and for the rural and urban subsamples.[11] The Gini-opportunity index is also tabulated in Table 3.5. In the calculation of the Gini-opportunity index, I have used two definitions of the health status variable. First, I use the four-category variable where 1 indicates that the health status is poor and 4 that the health status is excellent. Second, I use the dichotomous variable for health status to calculate the Gini-opportunity index. I present the index for the full sample and for the urban and rural subsamples.

I begin with the analysis of the results for the full sample. The dissimilarity index obtained with the LSSM data is about 8.4%. The dissimilarity index is usually interpreted as the share of total opportunities for enjoying a better health status that would need to be redistributed from individuals who feel healthier to individuals who feel less healthy for equality of opportunity to prevail.

The Shapley decomposition of the dissimilarity index shows that the early life circumstances that have the largest contributions to the dissimilarity index are: household socioeconomic status at age 10 (16%), mother's education (10%) and father's education

---

[11]For the decomposition of the dissimilarity index, I use the user-written command in Stata *hoishapley* (Hoyos Suarez, 2013).

(10.2%). Once I clean years of education from the influence of circumstances, the decomposition of the index shows a slight increase in the contributions of socioeconomic status at age 10 (22.2%), mother's education (12.4%) and father's education (13%).

The Gini-opportunity index is 0.10 when the variable for health status with four categories is taken as the outcome of interest. The index is three times larger when the outcome of interest is a dichotomous variable for self-assessed health status (which equals 0.318.) The Gini-opportunity index, likewise the Gini index, ranges between 0 and 1, so that the closer to 1 the most unequal the distribution of health status among the individuals is. Although the Gini-opportunity index could be decomposed using the Shapley-value, I do not provide estimates of the contribution that each circumstance makes to the index as this chapter focuses on the dissimilarity index.

The Gini-opportunity index obtained for the full sample is also slightly larger than that calculated for the United Kingdom by Rosa Dias (2009). In the British household panel, inequality of opportunity in adult health ranges between 0.009 and 0.018. In contrast with Rosa Dias, who only uses parental socioeconomic status as a circumstance, I use the full set of circumstances (except for the demographic variables, gender and age group) to calculate the Gini-opportunity index.

Turning to the results for the urban sample, I calculate a dissimilarity index of 7.9%, when I include years of education in the vector of circumstances. That is, 7.9% of total opportunities would need to be redistributed from individuals who are healthier to individuals who are less healthy for equality of opportunity to prevail. In rural areas, the index is relatively larger: about 10.1% of total opportunities would need to be redistributed from individuals who are healthier to individuals who are less healthy for equality of opportunity to prevail. The calculated indexes do not change considerably once I clean years of education from the influence of circumstances. For urban areas, the decomposition of the index shows a slight increase in the contributions of socioeconomic status at age 10 (from 10.5% to 13.7%), mother's education (12.9% to 16.5%) and father's

education (13% to 14.6%). For rural areas, the decomposition of the index shows a slight change in the contributions of region of birth (from 20.2% and 21.1%) and socioeconomic status at age 10 (from 35% to 40.5%), the two circumstances that are most influential in inequality of opportunity in health status in rural areas.

I present two additional sets of results in section A.3 in Appendix C. The first set of results include chronic illness and disability as control variables in the logistic regression model. These objective measures of health status have a negative and significant effect on the likelihood of reporting a good health status. This result is consistent across the full sample and the subsamples of urban and rural areas. The addition of these measures does not change the association between circumstances and adult health status previously described.

The use of self-reported and retrospective recall data could bias the results obtained here. In order to gauge if there is a systematic bias in how health status is reported, I examine how people perceive their own health status based on their economic conditions, after controlling for the set of circumstances and the presence of chronic illness and permanent disability. Self-reported health status and household income per capita (defined in both levels and logs) are strongly correlated, but once I control for circumstances and objective measures of health status this correlation attenuates at conventional significance levels. Thus, the bias created by self-reported measures should be reduced as long as more objective measures are included in the model.

In the second set of results, I analyze whether the age of an individual affects their recall of early-life circumstances in a certain direction. I estimate the logistic regression models for three age cohorts: 25–35, 36–50, and 51–65 years old. There are substantial differences by age group. For instance, maternal education seems to be more important for the 50–65 group than for the 35–50 group, for which socio-economic status at age 10 is the most prominent circumstance in inequality of opportunity. Region of birth and ethnicity are more important for the 25–35 age group than for any other group.

74

3.5   Concluding Remarks

This study measures the degree of inequality of opportunity in adult health in Colombia by employing stochastic dominance tests and a decomposition of a dissimilarity index. The empirical results suggest that household socio-economic status and parental education are the most salient early-life circumstances that affect health inequality in adulthood. These circumstances, however, do not reflect how important region of birth or ethnicity may be for different socio-economic groups. Ethnicity, for instance, is highly associated with inequality of opportunity in health in urban areas but not so in rural areas. In contrast with urban areas, region of birth is potentially one of the most important circumstances in rural areas.

Even though this study provides suggestive evidence on the various sources of adult health inequality, it has several limitations. Scholars are usually skeptical with the use of self-reported health status in developing countries. For instance, Sen (2002) argues that socially disadvantaged individuals fail to perceive and report the presence or absence of certain health conditions because they are constrained by their social environment. Moreover, their own understanding and appraisal of their health status may not agree with that of their physicians.

Self-reported health status may suffer from individual reporting heterogeneity. To the best of my knowledge, no study has provided evidence, appropriate for the Colombian context, in favor of or against the use of self-reported health in health research. Objective measures of adult health status are not observed in the LSSM dataset. Unfortunately, surveys like the Demographic and Health Survey do not provide intergenerational information for adults. The study of inequality of opportunity in adult health in Colombia faces the usual problem of data availability.

An additional problem is the use of retrospective questions about circumstances.

Household ownership of assets during childhood may not be accurately reported. This misreporting introduces bias in the estimates of the correlation between early-life circumstances and adult health. The analysis in this chapter does not allow to disentangle the effects of either genetic inheritance or parental health on investments in child's health capital, which is a weakness also identified in previous research (Trannoy et al., 2010).

The estimation of the dissimilarity index is also likely to be biased due to omitted variables if any of the unobserved circumstances is correlated with any of the observed circumstances included in the analysis. Abras et al. (2013) showed that this problem is potentially mitigated by one of the properties of the dissimilarity index: it can only increase when more circumstances are added. Of course, this property does not imply that the estimated contributions to the index also increase when more circumstances are included.

The inequality of opportunity analysis provides suggestive evidence of the lasting effects of childhood circumstances on adult health. The results presented in this study constitute a first step towards the identification of the potential channels through which health inequalities are transmitted from one generation to the next. The results in this chapter also suggest that the transmission channels of health inequality across generations operate differently in rural and urban areas. In order to achieve the goal of equality of opportunity in health, more specific policies should be designed to offset the effects of different circumstances in Colombia as a whole and in both rural and urban areas of the country.

Further research on inequality of opportunity in health in Colombia and Latin America should be based on novel longitudinal and administrative data that collect comprehensive information on the parents of tomorrow's children. Recall bias, a limitation of the data used in this study, could be minimized through a proper combination of administrative records and longitudinal information.

Table 3.1: Summary Statistics: Full Sample

Heads of Household between 25 and 65 years old. Total Number of Observations: 2,253

Variable	Observations	Mean or Proportion	Std. Dev.
Outcome			
Self-assessed Health Status	2,253	2.78	0.60
Poor	49	2.2%	0.15
Fair	556	24.7%	0.43
Good	1,487	66.0%	0.47
Excellent	161	7.1%	0.26
Early-life Circumstances			
Household Socioeconomic Status at Age 10			
Quintile 1	569	25.3%	0.43
Quintile 2	533	23.7%	0.43
Quintile 3	441	19.6%	0.40
Quintile 4	355	15.8%	0.36
Quintile 5	316	14.0%	0.35
No Information on Assets	39	1.7%	0.13
Education Level of Father			
None or Incomplete Primary	1,258	55.8%	0.50
Complete Primary and Incomplete Secondary	377	16.7%	0.37
Complete Secondary or More	194	8.6%	0.28
Unknown Father's Education	422	18.7%	0.39
No Information on Father's Education	2	0.1%	0.03
Education Level of Mother			
None or Incomplete Primary	1,345	59.7%	0.49
Complete Primary and Incomplete Secondary	447	19.8%	0.40
Complete Secondary or More	171	7.6%	0.26
Unknown Mother's Education	288	12.8%	0.33
No Information on Mother's Education	2	0.1%	0.03
Other circumstances			
Ethnicity			
Indigenous	59	2.6%	0.16
Black, mulato, raizal or palenquero	144	6.4%	0.24
No ethnic minority	2,050	91.0%	0.29
Years of Education	2,253	7.02	4.65
Born in Urban Area	1,103	49.0%	0.50
Born in Rural Area	1,144	50.8%	0.50
No Information on Area of Birth	6	0.3%	0.05
Region of Birth			
Atlantic	507	22.5%	0.42
Eastern	518	23.0%	0.42
Pacific	255	11.3%	0.32
Orinoquia-Amazonia	6	0.3%	0.05
Antioquia	251	11.1%	0.31
Valle del Cauca	160	7.1%	0.26
Bogotá	159	7.1%	0.26
San Andrés islands	2	0.1%	0.03
Central	395	17.5%	0.38
Additional Controls			
Male	1,598	70.9%	0.45
Age	2,253	44.77	11.01
Age group			
25-35	504	22.4%	0.42
35-45	594	26.4%	0.44
45-55	646	28.7%	0.45
55-65	509	22.6%	0.42

Source: 2010 Colombian LSSM Survey

Table 3.2: Stochastic Dominance Tests for Inequality of Opportunity

a. Household socioeconomic status at age 10

Quintile group	1 (lowest)	2	3	4	5 (highest)
Full sample					
1 (lowest)		~	~	~	~
2	~		~	~	~
3	~	>		~	~
4	>	>	~		~
5 (highest)	~	>	>	>	
Urban Areas					
1 (lowest)		~	~	~	~
2	~		~	~	~
3	~	~		~	~
4	>	~	~		~
5 (highest)	>	>	>	>	
Rural Areas					
1 (lowest)		~	~	~	~
2	~		~	~	~
3	~	~		~	~
4	~	~	~		~
5 (highest)	>	>	~	~	

b. Paternal Education

Level	None	Primary	Secondary and higher
Full sample			
None *		~	~
Primary **	>		~
Secondary and higher	>	~	
Urban Areas			
None *		~	~
Primary **	>		~
Secondary and higher	>	~	
Rural Areas			
None *		~	~
Primary **	~		~
Secondary and higher	~	~	

c. Maternal Education

Level	None	Primary	Secondary and higher
Full sample			
None *		~	~
Primary **	>		~
Secondary and higher	>	>	
Urban Areas			
None *		~	~
Primary **	~		~
Secondary and higher	>	>	
Rural Areas			
None *		~	~
Primary **	~		~
Secondary and higher	~	~	

Note: The symbol ">" indicates that the distribution of the type in the row first-order-stochastic dominates the distribution of the type in the column. The symbol "~" indicates that the distributions cannot be ranked using first-order stochastic dominance.
* None or incomplete primary education
** Complete primary or incomplete secondary education
Source: 2010 Colombian LSSM Survey

Table 3.3: Purging Years of Education from Circumstances: OLS Results

Dependent Variable: Years of Education	All Individuals	Urban Areas	Rural Areas
	(1)	(2)	(3)
Male	0.2172	0.6416***	−0.4885*
	(0.1885)	(0.2204)	(0.2690)
Age group (Ref. 25–35 years old):			
35–45 years old	−0.1058	0.0440	−0.7039**
	(0.2245)	(0.2749)	(0.3049)
45–55 years old	−0.2316	−0.3117	−0.8309**
	(0.2394)	(0.2849)	(0.3324)
55–65 years old	−1.1098***	−1.2353***	−1.8467***
	(0.2668)	(0.3243)	(0.3329)
Ethnicity (Ref. Not a minority):			
Indigenous	−0.0621	−0.0304	0.1704
	(0.5613)	(0.8450)	(0.6265)
Black/mulato/raizal/palenquero	0.3016	0.1005	0.2613
	(0.3615)	(0.4651)	(0.4410)
Region (Ref. Atlantic and San Andres islands):			
Eastern	0.0011	−0.3190	−0.1385
	(0.2681)	(0.3290)	(0.3445)
Pacific	0.4841	1.0698*	0.2100
	(0.3596)	(0.5568)	(0.3465)
Orinoquia- Amazonia	−0.5957	−1.0903	−0.2360
	(0.5788)	(0.7468)	(0.9172)
Antioquia	−0.0747	−0.2467	−0.0174
	(0.3158)	(0.3802)	(0.4452)
Valle	0.5982	0.5387	0.3399
	(0.4001)	(0.4505)	(0.5239)
Bogota	−0.3089	−0.5637	2.0025
	(0.3279)	(0.3598)	(1.6562)
Central	0.5395*	0.7487**	0.0573
	(0.2971)	(0.3669)	(0.3522)
Born in urban area	1.0276***	0.4466	0.3522
	(0.2204)	(0.2849)	(0.2865)
Household socioeconomic status at age 10 (Ref. Quintile group 1):			
Quintile group 2	0.7084***	1.0493***	−0.3497
	(0.2732)	(0.3525)	(0.3114)
Quintile group 3	2.0127***	2.1206***	0.4408
	(0.2874)	(0.3614)	(0.3432)
Quintile group 4	3.4114***	3.1020***	0.7434**
	(0.3255)	(0.3848)	(0.3549)
Quintile group 5	4.5999***	4.2618***	2.2478***
	(0.3554)	(0.4055)	(0.4083)
Paternal education level (Ref. None):			
Complete primary and incomplete secondary	0.9560***	0.7741**	1.2467**
	(0.3064)	(0.3550)	(0.5217)
Complete secondary or more	1.8947***	1.5467***	3.8638***
	(0.4034)	(0.4459)	(0.7869)
Unknown father's level of education	−0.7116**	−0.7402**	−0.5352*
	(0.2907)	(0.3766)	(0.2938)

Table 3.3: Purging Years of Education from Circumstances: OLS Results (continued)

Dependent Variable: Years of Education	All Individuals	Urban Areas	Rural Areas
	(1)	(2)	(3)
Maternal education level (Ref. None):			
Complete primary and incomplete secondary	1.0363***	1.1135***	0.6089
	(0.2906)	(0.3392)	(0.4195)
Complete secondary or more	2.5173***	2.5426***	2.4073**
	(0.4135)	(0.4612)	(1.0519)
Unknown mother's level of education	−0.4045	−0.1635	−0.2143
	(0.3553)	(0.4703)	(0.3390)
Constant	4.6050***	5.5638***	4.9071***
	(0.3564)	(0.4646)	(0.4833)
Observations	2,204	1,242	962
R squared	0.430	0.396	0.246

***, **, and * indicate statistical significance at the 1, 5 and 10 percent level, respectively.
Robust standard errors in parentheses
Source: 2010 Colombian LSSM Survey

Table 3.4: Log-odds Ratios for the Correlates of Health Status

Dependent variable: Self−reported health status (0 = poor or fair, 1 = good or excellent)	All Individuals		Urban Areas		Rural Areas	
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.5690***	0.5932***	0.6489***	0.7217***	0.5281**	0.4781**
	(0.1277)	(0.1280)	(0.1560)	(0.1566)	(0.2104)	(0.2089)
Age group (Ref. 25−35 years old):						
35−45 years old	−0.5462***	−0.5579***	−0.5281*	−0.5231*	−0.5544**	−0.6264**
	(0.2005)	(0.2005)	(0.2748)	(0.2748)	(0.2481)	(0.2474)
45−55 years old	−0.7550***	−0.7808***	−0.7587***	−0.7941***	−0.8692***	−0.9542***
	(0.1948)	(0.1946)	(0.2650)	(0.2647)	(0.2516)	(0.2527)
55−65 years old	−1.3172***	−1.4406***	−1.3481***	−1.4882***	−1.4127***	−1.6015***
	(0.1964)	(0.1967)	(0.2663)	(0.2669)	(0.2608)	(0.2626)
Ethnicity (Ref. Not a minority):						
Indigenous	−0.2143	−0.2213	−0.7064	−0.7099	0.5513	0.5687
	(0.4386)	(0.4386)	(0.5983)	(0.5983)	(0.4468)	(0.4469)
Black and other	−0.2408	−0.2073	−0.3739	−0.3625	−0.0548	−0.0281
	(0.2386)	(0.2385)	(0.2945)	(0.2944)	(0.3495)	(0.3493)
Region (Ref. Atlantic and San Andres islands):						
Eastern	−0.2613	−0.2612	−0.2041	−0.2403	−0.5537**	−0.5679**
	(0.1826)	(0.1826)	(0.2370)	(0.2371)	(0.2488)	(0.2494)
Pacific	−0.6624***	−0.6086***	−0.7622**	−0.6409**	−0.7878***	−0.7663***
	(0.2119)	(0.2107)	(0.3131)	(0.3099)	(0.2704)	(0.2693)
Orinoquia-Amazonia	0.3799	0.3136	0.8195	0.6959	−0.6004	−0.6246
	(0.5176)	(0.5175)	(0.7804)	(0.7804)	(0.7997)	(0.7999)
Antioquia	0.0858	0.0775	0.2955	0.2676	−0.6974**	−0.6992**
	(0.2213)	(0.2214)	(0.2864)	(0.2868)	(0.3055)	(0.3055)
Valle	0.1610	0.2275	0.2359	0.2970	−0.3386	−0.3038
	(0.3232)	(0.3235)	(0.3939)	(0.3942)	(0.4189)	(0.4185)
Bogota	−0.4860*	−0.5203*	−0.4415	−0.5054*		
	(0.2795)	(0.3047)	(0.3014)	(0.3060)		
Central	−0.2169	−0.1569	−0.1171	−0.0322	−0.4650*	−0.4591*
	(0.2017)	(0.2010)	(0.2678)	(0.2664)	(0.2543)	(0.2542)
Born in urban area	−0.0722	0.0420	−0.1611	−0.1105	0.1597	0.1957
	(0.1371)	(0.1360)	(0.1794)	(0.1793)	(0.2370)	(0.2366)
Household socioeconomic status at age 10 (Ref. Quintile group 1):						
Quintile group 2	0.1220	0.2008	0.1109	0.2299	0.1291	0.0934
	(0.1618)	(0.1604)	(0.2248)	(0.2211)	(0.2500)	(0.2498)
Quintile group 3	0.3300*	0.5538***	−0.0288	0.2117	0.7877***	0.8328***
	(0.1831)	(0.1796)	(0.2331)	(0.2282)	(0.2552)	(0.2559)
Quintile group 4	0.1149	0.4943**	−0.2175	0.1342	0.7065***	0.7825***
	(0.2148)	(0.2044)	(0.2707)	(0.2540)	(0.2576)	(0.2564)
Quintile group 5	0.4963*	1.0078***	0.3021	0.7854**	0.7044**	0.9343***
	(0.2986)	(0.2846)	(0.3614)	(0.3426)	(0.2864)	(0.2786)
Paternal education level (Ref. None):						
Complete primary and incomplete secondary	0.3043	0.4106*	0.4688*	0.5566**	−0.2181	−0.0906
	(0.2216)	(0.2217)	(0.2618)	(0.2628)	(0.3625)	(0.3596)
Complete secondary or more	−0.0745	0.1362	−0.0144	0.1610	0.4579	0.8531
	(0.3773)	(0.3788)	(0.4069)	(0.4085)	(0.7744)	(0.7731)
Unknown father's level of education	0.1135	0.0344	0.3437	0.2597	−0.3095	−0.3642
	(0.1950)	(0.1948)	(0.2674)	(0.2668)	(0.2480)	(0.2464)

Table 3.4: Log-odds Ratios for the Correlates of Health Status (continued)

Dependent variable: Self−reported health status (0 = poor or fair, 1 = good or excellent)	All Individuals		Urban Areas		Rural Areas	
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal education level (Ref. None):						
Complete primary and incomplete secondary	−0.0212	0.0940	0.0231	0.1493	−0.3439	−0.2816
	(0.2117)	(0.2109)	(0.2558)	(0.2546)	(0.3187)	(0.3173)
Complete secondary or more	0.5116	0.7915*	0.7245	1.0128**	−1.1600*	−0.9138
	(0.4441)	(0.4398)	(0.5181)	(0.5139)	(0.6946)	(0.6867)
Unknown mother's level of education	−0.0382	−0.0831	−0.0705	−0.0891	0.0485	0.0266
	(0.2310)	(0.2307)	(0.3211)	(0.3210)	(0.2663)	(0.2664)
Years of education	0.1112***		0.1134***		0.1023***	
	(0.0174)		(0.0219)		(0.0262)	
Years of education purged from circumstances		0.1112***		0.1134***		0.1023***
		(0.0174)		(0.0219)		(0.0262)
Constant	0.6589***	1.1709***	0.7384**	1.3694***	0.6988**	1.2006***
	(0.2437)	(0.2368)	(0.3416)	(0.3290)	(0.3528)	(0.3408)
Observations	2,204	2,204	1,242	1,242	956	956
Log−likelihood	−4.477e+06	−4.477e+06	−3.328e+06	−3.328e+06	−1.085e+06	−1.085e+06
Pseudo R squared	0.126	0.126	0.136	0.136	0.113	0.113

, ***, and ***** indicate statistical significance at the 1, 5 and 10 percent level, respectively.
Robust standard errors in parentheses
Source: 2010 Colombian LSSM Survey

Table 3.5: Gini-Opportunity index and Dissimilarity Index of Inequality of Opportunity, with its Decomposition

	All individuals		Residents in Urban Areas		Residents in Rural Areas	
Gini-Opportunity Index (1)	0.1019		0.1148		0.0720	
Gini-Opportunity Index (2)	0.3182		0.3550		0.2604	
Dissimilarity Index (3)	0.0838	0.0839	0.0793	0.0793	0.1016	0.1016
Decomposition of the Dissimilarity Index (in %)						
Educational attainment	46.59		45.25		30.13	
Education purged from circumstances		33.31		36.76		22.53
Circumstances	53.41	66.69	54.75	63.24	69.87	77.47
Early-Life Circumstances	35.80	47.71	36.42	44.85	44.13	49.99
Mother's education	10.04	12.93	12.90	16.50	3.54	2.20
Father's education	10.21	12.49	12.98	14.57	5.64	7.30
Household socioeconomic status at age 10	15.56	22.28	10.53	13.77	34.96	40.49
Demographics	17.61	18.98	18.33	18.39	25.73	27.49
Region of birth	11.64	11.95	13.13	13.17	20.19	21.10
Born in urban area	4.56	5.61	1.00	0.97	3.87	4.71
Ethnicity	1.42	1.42	4.20	4.25	1.67	1.69
Observations	2,204		1,242		962	

Notes:
(1) The Gini-opportunity index is calculated using a self-assessed health status variable in which 1 = poor, 2 = fair, 3 = good, and 4 = excellent.
A categorical variable for the individual's years of education has also been used in this calculation. Gender and age group are not included.
(2) The Gini-opportunity index is calculated using a self-assessed health status variable in which 0 = poor or fair, and 1 = good or excellent.
(3) The index in the first, third and fifth columns include years of education as a circumstance, whereas the second, fourth, and sixth columns include years of education purged from circumstances.

Source: 2010 Colombian LSSM Survey

Chapter 4

Foods and Fads: The Welfare Impacts of Rising Quinoa Prices in Peru

with Marc Bellemare[1] and Seth Gitter[2,3]

[1]Associate Professor, Department of Applied Economics, University of Minnesota
[2]Associate Professor, Department of Economics, Towson University
[3]Marc Bellemare, Johanna Fajardo-Gonzalez and Seth Gitter contributed equally to this work

4.1 Introduction

Riding on a wave of interest in so-called "superfoods"[4] in the United States and other rich countries, quinoa—a relatively high-protein grain that has been grown for millennia in the Andean regions of Bolivia, Colombia, Ecuador, and Peru—went in less than a decade from being a largely unknown commodity outside of South America to being an upper-class staple in those same rich countries.[5] As quinoa imports to the US increased more than tenfold, from about 5 million pounds per year in 2004 to almost 65 million pounds per year in 2013 (DePillis, 2013), the price of quinoa tripled (Blythman, 2013).

Some have questioned the consequences of this increase in the popularity of quinoa, citing concerns about the effects of rising quinoa prices on the welfare of individuals and households in places where quinoa had traditionally been produced and consumed. A January 2013 article in the *Guardian* (Manchester) made the following claim (Blythman, 2013):

> [T]here is an unpalatable truth to face for those of us with a bag of quinoa in the larder. The appetite of countries such as ours for this grain has pushed up prices to such an extent that poorer people in Peru and Bolivia, for whom it was once a nourishing staple food, can no longer afford to eat it.

Three days later, an article in the *Globe and Mail* (Toronto) made the opposite claim (Saunders, 2013):

> The people of the [Andean plateau] are indeed among the poorest in the Americas. But their economy is almost entirely agrarian. They are sellers—

[4]The Oxford English Dictionary defines superfoods as foods "considered especially nutritious or otherwise beneficial to health and well-being" (Oxford English Dictionary, 2014).

[5]With 50% of Peruvian quinoa going to the United States, the United States is the commodity's largest export market (Andina, 2016). It is followed by Canada (8%), Australia (7%), Germany (6%), the United Kingdom (6%), the Netherlands (4%), France (3%) and Israel (3%).

farmers or farm workers seeking the highest price and wage. The quinoa price

rise is the greatest thing that has happened to them.

As one might expect from media accounts, neither claim was based in serious empirical analysis. That net buyers of a commodity are made worse off and net sellers better off, at least in the short run, by an increase in the price of that commodity is well-understood by economists (Deaton, 1989a).

But what are the longer-term,[6] general equilibrium effects of that price increase for consumers? And what is the effect of an international, positive price shock on the welfare of producers-cum-consumers of that commodity? We study the welfare impacts of rising quinoa prices on those households that have traditionally produced and consumed it. To do so, we use 11 years of the Peruvian *Encuesta Nacional de Hogares* (ENAHO), a large-scale, nationally representative household survey, to look at whether: (i) there is a systematic relationship between the value of household consumption (which we use here as a proxy for household welfare; see Deaton (1989a)) and the local purchase price of quinoa for those households that report consuming quinoa; and whether (ii) there is a systematic relationship between household welfare and the price of quinoa for those households that report producing quinoa.

Our study period (i.e., 2004-2014) covers years both before and after the price of quinoa rose sharply. Because the ENAHO is a repeated cross-section and is thus not longitudinal, we use pseudo-panel techniques (Deaton, 1985; McKenzie, 2004; Christiaensen and Subbarao, 2005; Antman and McKenzie, 2007a,b; Cuesta, Ñopo, and Pizzolitto, 2011) , wherein we average over household-level measures within each geographical unit and then treat those geographical units as our primary units of observation.[7] To study the re-

---

[6]By "longer-term," we are referring to a time horizon that is longer (i.e., up to one year, given the frequency of our data) than (Deaton, 1989a)'s short-term measure of welfare, and not to the long-term as it is typically understood in economics, i.e., the length of time required for all factors of production to be variable.

[7]Peru is divided in 1,838 districts in 195 provinces in 25 departments.

lationship between the international price of quinoa and household consumption, we rely in turn on geographical unit fixed effects with: (i) year fixed effects and (ii) higher-order geographical unit-year fixed effects.[8]

Our work is most closely related to the literature on the effects of commodity price shocks. This is a sizeable literature wherein scholars look at the effects of commodity price shocks on a host of outcome variables, from child outcomes (Cogneau and Jedwab, 2012) to conflict (Dube and Vargas, 2013) and almost everything in between. Specifically, our work relates to the literature on the effects of commodity price shocks—usually, food price shocks—on welfare. In a seminal contribution, Deaton (1989b) studies the effects of higher rice prices on welfare and inequality in Thailand. He finds that higher prices redistribute income towards households in the middle of the rural income distribution, with marked regional variations. More recently, Ivanic and Martin (2008) study the effects of higher global food prices on poverty in low-income countries. Using household surveys from nine low-income countries, they find that the effects of higher food prices on poverty vary by country, but also by commodity. Wodon and Zaman (2010) review the evidence looking specifically at sub-Saharan Africa, and they find that higher food prices tend to increase the extent of poverty given that net consumers tend to outnumber net producers of food. The study that is perhaps closest in spirit to our work is a study by Zezza et al. (2008), who rely on household surveys in 11 countries to look at how different groups of households are affected differently when food prices increase in an effort to look at the distributional impacts of food price changes. One notable difference between our work and the majority of studies in the commodity price shocks literature, however, is that while that literature typically focuses on major food staples (e.g., maize, rice, wheat, etc.), we focus on a non-staple. Additionally, the production of quinoa is concentrated in a specific region of the world, and little quinoa is produced in the United States or Europe. This makes quinoa similar to other regionally produced commodities, such as

---

[8]At the district level, this means province-year fixed effects. At the provincial level, this means department-year fixed effects.

teff in Ethiopia and millet in Central Africa or India. The only other economic study of the effect of rising quinoa prices has been by Stevens (2015), who finds that cultural preference for quinoa in certain areas of Peru has not led to a worsening of nutritional outcomes.

Our results suggest that the increased international demand for quinoa and the resulting quinoa price boom have had beneficial effects for consumers as well as for producers of quinoa in Peru. First, we find a positive relationship between the price of quinoa and household welfare within the average geographical unit-year wherein quinoa was consumed, which suggests that the sharp increase in the price of quinoa has had positive general equilibrium effects on the welfare of the average household in those geographical unit-year observations.[9] Specifically, we find that for a 25% increase in the price of quinoa–a change that is commensurate to the change in the purchase price of quinoa between 2013 and 2014, when international demand spiked–total household consumption increases on average by about 1.25%.

Second, and in line with theoretical expectations (Deaton, 1989a), we find a positive relationship between household welfare and household quinoa production. More specifically, the 25% increase in the price of quinoa between 2013 and 2014 would be associated with a 3.5% to 4% increase in consumption of quinoa producing households.

The remainder of this chapter is organized as follows. In section 4.2, we present the data as well as some descriptive statistics. Section 4.3 presents the empirical framework we develop to study the impacts of rising quinoa prices on welfare, with particular emphasis on our identification strategy. In section 4.4, we present and discuss our estimation results. Section 4.5 concludes with some policy recommendations and directions for future research.

---

[9]We focus on quinoa-consuming districts, households, and departments because those are the geographical units for which quinoa prices are available.

4.2   Data and Descriptive Statistics

We use data from Peru's *Encuesta Nacional de Hogares* (ENAHO), an annual household survey conducted by the Peruvian government's *Instituto Nacional de Estadística e Informática* (National Institute of Statistics and Informatics). Because of their high quality and nationally representative character, ENAHO data have been used frequently by economists. Among others, Dell (2010) has used the ENAHO to study the long-term consequences of an extractive institution operating during colonial times in Peru, Aragón and Rud (2013) have used the ENAHO to study the effects of a gold mine on local incomes, and Galdo (2013) has used the ENAHO to study the long-run labor-market impacts of civil war.

The ENAHO sample is selected every year so as to be nationally representative. The data include household-level sampling weights, which we use throughout our analysis. We use repeated cross sections from 2004 to 2014 inclusively, which encompass 277,759 household-year observations. We discuss in Section 4.3 how the repeated cross-sectional nature of the data allows the construction of a pseudo-panel.

Our outcome of interest is the total value of household consumption,[10] Annual total consumption is computed by INEI as the sum of (i) purchases of food, clothing, housing, fuel, electricity, furniture, housewares, health, transportation, communications, and entertainment. Individuals reported information in past month or past three months depending on expenditure group; (ii) expenditures on appliances, transport and others; (iii) expenditures on food consumed outside the household; (iv) expenditures on food to be consumed inside the household, and (v) the reported value of own consumption, gifts, social programs, and payments in kind in the same expenditure groups. As we discuss further, food consumption is reported via a two-week recall in a specific module of the ENAHO. [11] In developing countries such as Peru, where many rural households produce

---

[10]We remove the value of quinoa that is produced and consumed by the household from our measure of household welfare so as to avoid biasing the relationship between quinoa prices and consumption by way of reverse causality. We explain our identification strategy further in section 4.3.

[11]ENAHO is a continuous, monthly survey. Every year, INEI visits the same primary sampling units

food for their own subsistence, it is important to include the value of all consumption, and not just purchases, in order to paint a more accurate portrait of welfare.

ENAHO also includes a battery of questions on agricultural production activities in the past 12 months. Households report the quantity produced for about 200 products, as well as the proportion of such production used for own consumption, selling, bartering, seeding and sub-products. There is also information on the selling unit price and the value of sales.

We divide our sample up into two non-mutually exclusive categories. "Quinoa producers" refers to households that report producing quinoa over the previous year, whether those households consumes quinoa or not; and "quinoa consumers" refer to households that report consuming quinoa over the last two weeks, whether those households produces quinoa or not. Although it is common in the agricultural economics literature to split households between net buyers and net sellers of a commodity (see, for example, Bellemare, Barrett, and Just (2013)), the different recall periods for production (i.e., past year) and consumption data (i.e., past two weeks) make this impossible in this chapter. However, fewer than 2% of producers reported purchasing quinoa in the last two weeks, and fewer than 1% of quinoa buyers reported producing quinoa in the past year.

A comparison of mean household consumption among households that produce quinoa and those that consumed quinoa but did not produce it is shown in Table 4.1. The most notable difference in Table 4.1 is that quinoa-producing households (third column of Table 1) consumed roughly 40 % of what quinoa-consuming households did at the beginning of the sample period.[12] Households that consumed but did not produce quinoa (fourth column of Table 4.1), however, had total household consumption about 30 % higher than that of households that neither consumed quinoa nor produce it. In other words, consumers

(*conglomerados*) during the same survey month, but selects a different sample of dwellings to conduct the survey.

[12]All monetary values are expressed in real terms in 2004 PEN. The 2004 PPP adjusted exchange rate was 1.3 Soles = $1 USD

of quinoa look like they were substantially better off than the rest of the population. This parallels how, at the international level, the demand from quinoa overwhelmingly comes from rich countries.

For all households, purchased goods represented roughly 75% of the value of total consumption (which includes household food production). For quinoa-producing households, that number was closer to 60%. In other words, 40% of the total household consumption of quinoa-producing households is from non-purchased goods, including household food production. Quinoa-producing households thus appear less integrated in markets than non-producing households.

A comparison of mean household consumption among households that produce quinoa (about 4%) and those that consumed quinoa but did not produce it (about 20%) is shown in Table 4.1. The most notable difference in Table 4.1 is that quinoa-producing households (fifth column of Table 4.1) consumed roughly 40% of what quinoa-consuming households did at the beginning of the sample period.[13] Households that consumed but did not produce quinoa (seventh column of Table 4.1), however, had total household consumption about 30% higher than that of households that neither consumed quinoa nor produce it. In other words, consumers of quinoa look like they were substantially better off than the rest of the population. This parallels how, at the international level, the demand from quinoa overwhelmingly comes from rich countries.

Figure 4.1 shows time series of the consumption levels of quinoa producers, quinoa consumers, and those that neither produced nor consumed quinoa wherein, for ease of comparison, baseline consumption is set equal to 1 for each groups.[14] Up until 2009, the welfare of quinoa consumers increased at a faster rate than that of quinoa producers. Starting in 2010, however, quinoa producers saw their welfare increase faster than quinoa consumers. In fact, and as the econometric analysis below will confirm, at the peak

---

[13] All monetary values are expressed in real terms in 2004 Peruvian Soles (PEN). The 2004 PPP adjusted exchange rate was 1.3 PEN = $1 USD

[14] Yearly departmental-level deflators are used to control for price changes

of the quinoa price boom in 2013 and 2014, the welfare of quinoa producers increased much faster than that of quinoa consumers. Comparing quinoa-producing households one the one hand with quinoa-consuming and quinoa neither consuming nor producing households on the other hand, the welfare of quinoa producers increased by over 50% over the period 2004-2014, whereas it increased by about 25% for the other two groups of households.

In Table 4.2, we take a closer look at quinoa consumers. Over the sample period, one fourth to one third of the households in our sample reported consuming quinoa in the two weeks before they were surveyed, as shown in the second column of Table 4.2.[15] Over these same two weeks, conditional on purchasing, the average household in the data purchased less than one kilogram of quinoa. Back-of-the-envelope calculations based on Table 4.2 suggest that the total effect of price rises on consumers was small: At the beginning of the sample period, households purchased roughly 22.6 kg per year (or 0.87 kg every two weeks), but the real cost of this amount of quinoa rose roughly 200 Peruvian Soles (PEN) over the sample period, which is about 0.8% of the overall consumption for those households that do not produce quinoa in 2014.

Over the sample period, quinoa purchases have fallen. Indeed, the third and fifth columns of Table 4.2 show that the amount of quinoa purchased over the two weeks before the survey fell by about 20%. Using the two-week purchase data, we estimated annual purchases by multiplying by 26 to create an annual budget share. The budget share of quinoa rose as the real price of quinoa paid by buyer more than doubled from 2004 to 2014. As noted above, quinoa represents a very small share (i.e., less than 1%) of the budget of the average household in the data, and the change in budget share between 2004 and 2013 is roughly 0.5%. Compared to the budget share of staples in low-income countries, which often average over 50% (see, for example, Barrett and Dorosh (1996)), quinoa does not seem to be a staple for households in Peru, though Stevens (2015, tables

---

[15]Quinoa production in Peru is seasonal. The sowing season usually starts in September, peaking in October-November. The harvest season usually takes place in April to June.

1 and 3) notes that there is a substantial amount of heterogeneity; in particular, averages in traditional quinoa-consuming areas such as Puno are higher than national averages, with a food budget share of quinoa of 3.6% compared to the national food budget share of quinoa of 0.5% in 2012.

Table 4.3 shows some descriptive statistics for quinoa producers and sellers. Over the period 2004-2014, roughly 3.6% of all households in the data grew any quinoa. Counter to what one might expect given the quinoa price boom of 2012-2013, the percentage of producers in the data dropped from 3.4% in 2011 to 2.8% in 2012, and then to 2.6% in 2013. In 2014, with the international quinoa price still at its peak, the proportion of producers when back to the 2011 levels.

The second column of Table 4.3 shows that in any given year, less than 0.7% of the households in our sample sold any quinoa. Most of the households who grow quinoa, consume it all. More interestingly for our purposes, the percentage of households that sold some of their quinoa production (column 2 of Table 3) almost doubled between 2010 and 2011. When looking only at the sub-sample of quinoa producers, the average household produced less than 90 kg of quinoa in the last 12 months, and over time, the volume of quinoa production has been U-shaped, with the highest output levels per household at the beginning and at the end of our sample.

In our sample, over 98% of households that produced quinoa used at least some of it for their own consumption. As shown in the fifth column of Table 3, however, the percentage of production used for a household's own consumption fell over the study period, from around 85% in 2004 to about 65% in 2014.

We mentioned earlier that the international price of quinoa had more than tripled over the period 2004-2014. Even more impressively, quinoa sellers have seen the real price of quinoa experience a more than fourfold increase during that period. The rate at which the purchase price of quinoa rose (column 4 of Table 4.2) was less than the growth in the sales

price (column 6 of Table 4.3), and the farm -to-consumer price ratio has increased from 43% to 55% between 2004 and 2014.[16] This suggests that quinoa producers have captured some of the gains from rising quinoa prices, though this is obviously not a formal test of that hypothesis, which is beyond the scope of this chapter.

Lastly, the revenue of quinoa sellers grew almost sevenfold over the period 2004-2014 (seventh column of Table 4.3), although that increase has not been steady. There are also three jumps in revenue: the first occurring between 2008 and 2009, when revenue almost doubled; the second one occurring between 2011 and 2012, when revenue increased by over 80%; and the third occurring between 2013 and 2014, when revenue almost doubled. This rise in revenue was even more pronounced when looking at all quinoa farmers (eighth column of Table 4.3), and not just to quinoa sellers.

4.3   Empirical Framework

The ENAHO is a repeated cross-sectional household survey, so the usual panel methods favored by applied microeconomists (i.e., household fixed effects) are not available in this context. A standard strategy proposed by Deaton (1985) to overcome the type of data limitations one faces with repeated cross-sections, is to rely on pseudo-panel methods. Intuitively, pseudo-panel methods treat groups of observations (rather than the observations themselves) as units of analysis. In our application, instead of treating the household as our unit of analysis, we treat geographical units as our units of analysis, and we use geographical unit-level averages as our primary data. Recall that Peru is divided in 1,838 districts in 195 provinces in 25 departments. As a check on the robustness of our results, we estimate each set of results three times, respectively treating districts, provinces, and departments as our units of observation.

---

[16]Producers reported their annual sales in an agricultural production module in the ENAHO. Consumer prices were taken from the two-week recall consumption model which included purchase prices.

The ENAHO has data on all 25 departments, on all but one of the 195 provinces, and on 1,401 of 1,838 districts. Given the random selection of communities and the nationally representative nature of the ENAHO, those missing districts should not reduce the external validity of our results.

Pseudo-panel methods like the ones we use in this chapter have been effectively used to estimate economic mobility (McKenzie, 2004; Cuesta, Ñopo, and Pizzolitto, 2011) and to study poverty in developing countries (Antman and McKenzie, 2007a,b; Christiaensen and Subbarao, 2005; Cruces et al., 2015). Again, recall that in a pseudo-panel, the outcome variable (here, household welfare) and the treatment variable (here, the price paid on average by a household for its quinoa when studying the welfare of consumers, and whether a household grows quinoa when studying the welfare of producers) are averaged across geographical unit. Because households are chosen at random within each geographic region, the average among sampled households should track the average among population households.

Our variable of interest is the total real value of household consumption (deflated using departmental deflators provided by INEI), which we use here as a proxy for household welfare. For each geographical unit g, we compute the regional sample mean of the total value of household consumption $\bar{c}_{gt}$ as the average of the total value of household consumption c_{ght} over all observed households h in the set H_{gt} of all households sampled in geographical unit g in year t, such that

$$\bar{c}_{gt} = \frac{1}{H_{gt}} \sum_{i=1}^{H_{gt}} c_{hgt} \tag{4.1}$$

Here, pseudo-panel methods have two clear benefits. First, because the ENAHO covers over 20,000 households annually, the data is rich at both the national and sub-national levels, and statistical power is not a concern. Second, as the number of households averaged over increases when computing the geographical unit-level mean, the effect of

potential error in the measurement of a particular household's consumption is reduced given that that error becomes spread out over more households. If they were available to us, individual household fixed effects would allow correcting for time-invariant measurement error; however, time-variant measurement error would still be present, and fixed effects are thought to compound measurement error problems (Wooldridge, 2010).[17] This would be an issue especially regarding food consumption, where annual data is extrapolated from two weeks' worth of food consumption. Our use of pseudo-panel methods reduces this problem.

As with many of the decisions one has to make in applied econometrics, moving from the largest (i.e., department) to the smallest (i.e., district) geographical unit involves a tradeoff. As the geographical unit gets smaller, fewer observations go into making the geographical level-unit average, which maximizes measurement error but also presents the most amount of statistical power in this context. Conversely, as the geographical unit gets larger, there are fewer units of observations available for analysis, which decreases statistical power, but which also minimizes measurement error problems. In order to examine this tradeoff, we estimate all of our specifications for each of the three levels of geographic analysis.

We use two variables as our treatment variable, depending on whether we want to study the welfare effects of rising quinoa prices on consumers or on producers of quinoa. For consumers, the treatment variable is the proportion of quinoa consumers within a geographical unit interacted with the annual international price of quinoa reported by the FAO.[18] For producers, the treatment variable is the proportion of quinoa producers within a geographical unit interacted with the annual international price of quinoa reported. Both treatment variables vary over time and across space, and it is this spatio-temporal

---

[17]More specifically, Wooldridge (2010, p. 365) writes: "It is widely believed in econometrics that ... FE transformations exacerbate measurement error bias (even though they eliminate heterogeneity bias). However, it is important to know that this conclusion rests on the classical errors-in-variables model under strict exogeneity, as well as on other assumptions."

[18]These are nominal prices received by farmers for quinoa sales as collected at the point of initial sale (prices paid at the farm-gate). They are expressed in Soles per tonne.

variation which we exploit here to identify the effects on welfare of rising quinoa prices.

For our analysis of quinoa consumers, we regress the logarithm of the total value of household consumption on the proportion of quinoa consumers within a geographical unit interacted with the annual international price of quinoa. The proportion of quinoa consumers within a geographical unit allows the extent of quinoa consumption to vary over time as households choose which products to consume each year (recall that this measure comes from an annual extrapolation of a variable reported in a two-week recall period.)

For our analysis of quinoa producers, we regress the logarithm of the total value of household consumption on the proportion of quinoa producers within a geographical unit interacted with the annual international price of quinoa. The proportion of quinoa producers within a geographical unit allows the extent of quinoa production to vary over time as households choose which crops to grow each year.

4.3.1   Estimation Strategy

4.3.1.1   Consumers

In the case of quinoa consumers, our equation of interest is such that

$$\ln c_{gt} = \alpha_0 + \alpha_1 \ln p_t \cdot S_{gt} + \delta_g + \tau_t + \epsilon_{gt} \tag{4.2}$$

where, in a slight abuse of notation, $\ln c_{gt}$ is the mean of $\ln c_{hgt}$ in geographical unit g in year t, p_t is the international price of quinoa in year t, S_{gt} is the proportion of quinoa consumers in geographical unit g in year t, δ_g is a vector of geographical-unit fixed effects, and ϵ_{gt} is an error term with mean zero. The term τ_t denotes either: (i) year fixed effects; or (ii) higher-order geographical unit-year fixed effects, whenever feasible. We cluster the

standard errors at the level of the geographical unit (i.e., district, province, or department) we use as our unit of analysis.

4.3.1.2 Producers

In the case of quinoa producers, our equation of interest is

$$\ln c_{gt} = \beta_0 + \beta_1 \ln p_t \cdot D_{gt} + \gamma_g + \theta_t + v_{gt} \tag{4.3}$$

where $\ln c_{gt}$ is the mean of $\ln c_{hgt}$ in geographical unit g, p_t is the international price of quinoa in year t, D_{gt} is the proportion of households that produce quinoa in geographical unit g, γ_g is a vector of geographical unit fixed effects, and v_{gt} is an error term with mean zero. The term θ_t denotes either: (i) year fixed effects; or (ii) higher-order geographical unit-year fixed effects, whenever feasible. We cluster the standard errors at the level of the geographical unit we use as our unit of analysis.

4.3.2 Identification Strategy

4.3.2.1 Consumers

The error term in equation (4.2) contains everything that is unobserved in that equation. Because the households surveyed in the ENAHO are randomly selected, when controlling for the passage of time, the households in a given geographical unit in a given year are similar to the households in the same geographical unit the following year, and this holds both in terms of their observable and unobservable characteristics. Thus, provided we account for the passage of time in our estimations, our use of geographical unit-level fixed effects should take care of the time-invariant heterogeneity between households.

To account for time-variant unobserved heterogeneity, we estimate several different specifications, the idea being that if we find similar effects throughout, our results are less likely to be biased. First, for our district-level analysis, on top of including district fixed effects, we estimate specifications with: (i) year fixed effects; (ii) province-year fixed effects; and (iii) department-year fixed effects. Second, for our provincial-level analysis, on top of including province fixed effects, we estimate specifications with: (i) year fixed effects; and (ii) department-year fixed effects. Finally, for our departmental-level analysis, on top of including departmental fixed effects, we estimate specifications with year fixed effects.

How do those specifications help identify the welfare effects of rising quinoa prices for consumers? To help think through this, it helps to consider the three sources of statistical endogeneity, viz. (i) reverse causality or simultaneity; (ii) unobserved heterogeneity; and (iii) measurement error. As regards reverse causality or simultaneity, quinoa appears to be a normal or a luxury good, and as consumers get better off, they are likely to start consuming more quinoa, which might cause local quinoa prices to increase. Over the period 2004-2014, however, quinoa price increases were largely due to an increased international demand for quinoa rather than to an increased domestic demand for it. Moreover, even if an increased domestic demand for quinoa had driven prices up, our use of geographical unit fixed effects would control for the average demand for quinoa in a given geographical unit, and the various means of controlling for the effect of time enumerated above would absorb much of the evolution of that demand.

As regards unobserved heterogeneity, as we discussed above, our use of pseudo-panel methods allows for matching the households in a given geographical unit from year to year along both their observable and their unobservable characteristics. Our use of fixed effects at the geographical unit level, along with the various methods we deploy to control for the effect of time, purge the error term of most of its prospective endogeneity due to unobserved heterogeneity.

One potential source of unobserved heterogeneity comes from an increase in total consumption from income effects. For example, if there is unobserved variation in income that differs between geographic regions and is not consistent with geographic time-trends this may bias the estimate through rising prices of all goods. To control for this issue we use an annual departmental level deflator for the welfare measure. Unfortunately, deflators at smaller geographic regions are not available. Our results, however, are mostly consistent across the two lower geographical levels we study.

As regards measurement error, we noted earlier that this is a concern, especially at the district level, where few observations go into making geographical unit-level averages. For this reason alone, we estimate everything at higher administrative levels (i.e., province and department). But our various layers of fixed effects and time controls control for the measurement error that is systematic at those levels. What remains is likely to be classical measurement error, which causes attenuation bias, in which case $\widehat{\alpha_1}$ is an estimate of the lower bound on the true quinoa effect of household welfare.

4.3.2.2 Producers

The error term in equation (4.3) contains everything that is unobserved in the same equation. If those unobservable factors are correlated with the variables on the RHS of equation (4.3), our estimate of the impact of quinoa production on household welfare is biased.

Again, we discuss in turn the three potential sources of statistical endogeneity, viz. (i) reverse causality or simultaneity, (ii) unobserved heterogeneity or omitted variables, and (iii) measurement error. Reverse causality or simultaneity issues might arise if the prospect of a higher welfare (as proxied by household consumption) induces some households who did not previously grow quinoa to do so, or if it induces quinoa producers to grow more quinoa within a given year, through a price effect. Our use of geographical unit fixed effects, however, would control for the average production for quinoa in a given

geographical unit, and the various means of controlling for the effect of time enumerated above would absorb the evolution of quinoa supply.

Unobserved heterogeneity issues might arise in this context if some unobservable factor is correlated with the variables on the RHS of equation (4.3). For example, it could be that households whose primary decision maker is more risk averse are more likely to grow quinoa, or that they grow more of it. In applied microeconomic studies such as this one, unobserved heterogeneity is generally the most important problem plaguing the identification of causal relationships. This problem is considerably lessened here by our use of pseudo-panel techniques. Indeed, recall that each round of our data consists of randomly selected households. Because the households selected at random in each geographical unit in each year are representative of that geographical unit, our use of geographical unit fixed effects should control for all things time-invariant within a geographical-unit, both observable and unobservable.[19] Of course, this does not control for those factors that are time-variant within a geographical unit, which are unobserved and correlated with the variables on the right-hand side of equation (4.3). Our use of year dummies should partly obviate that issue.

Finally, measurement error issues can bias our estimate of the impact of quinoa production on household welfare in two ways. With classical measurement error, our estimate of the impact of quinoa production on household welfare would be biased toward zero. With systematic measurement error, our estimate would be biased in a systematic direction, which would depend on the direction of the measurement error. Time-invariant measurement error that is systematic at the geographical unit level would be controlled for by the geographical unit fixed effects. Here, the measurement errors we should be most preoccupied with are: (i) classical measurement error; and (ii) time-variant systematic measurement error in our variable of interest, i.e., the proportion of households that produce quinoa in a geographical unit. On the former, we have discussed above

---

[19]The ENAHO includes a subsample that is resurveyed as part of a panel. These households are randomly chosen and the combination of the panel and non-panel households is nationally representative.

how the extent of measurement error is dependent upon the geographical unit we use as an observation. On the latter, there is no reason to believe that there is any systematic measurement error in this context, as there is really no incentive for respondents to systematically over- or under-report whether they produce quinoa or not.

The change in welfare over time before the rapid increase in the price of quinoa is likely similar for quinoa producers and non-producers. Figure 4.1, which plots the evolution of household welfare for quinoa producers, consumers, and non-consumers show that average household welfare followed a similar course from 2004 to 2010, after which the welfare of net sellers of quinoa has clearly evolved faster than the welfare of the other groups.

Another threat to identification when using pseudo-panel methods is the possibility that the composition of the relevant groups—here, households that produce (consume) quinoa versus households that do not produce (consume) quinoa—changes over time. In our application, it is possible that some households that did not grow (consume) quinoa decide to grow (consume) quinoa in response to higher expected levels of welfare. That said, for producers, Table 4.3 shows that the proportion of quinoa producers is relatively stable over the sample period. If anything, that proportion declines slightly toward the end of the sample period. Similarly, given that the dramatic increase in the price of quinoa in 2012-2014 was largely unpredictable and driven by an increased international demand for quinoa, we are not worried about a potential Ashenfelter dip (see Heckman and Smith (1995)).[20] Looking at Figure 4.1, it does not look as though the welfare of quinoa-producing households was significantly lower than that of other households before the quinoa price increase of 2012-2014.

---

[20]In this context, an Ashenfelter dip would involve households self-selecting into quinoa cultivation ex ante of the quinoa price spike, based on their expectation that the price of quinoa would increase significantly.

4.4 Estimation Results and Discussion

4.4.1 The Welfare Effects of Rising Quinoa Prices on Consumers

Tables 4.4 to 4.6 present estimation results for the welfare (i.e., consumption) effects of rising quinoa prices on consumers. In each table, the coefficient on the logarithm of the international price of quinoa interacted with the proportion of quinoa consumers is an estimate of the quinoa price elasticity of household welfare, on average, for those households in geographical units where quinoa was consumed for the period 2004-2014. In other words, this coefficient tells us how, for a 1% increase in the price of quinoa in those districts, provinces, and departments where quinoa was consumed for the study period, household welfare changed.

Tables 4.4 to 4.6 present estimation results at the district, provincial, and departmental levels, respectively. In almost all cases (i.e., 5 out of 6), the quinoa price elasticity of household welfare is statistically significant. In terms of economic significance, the quinoa price elasticity of household welfare ranges from 0.04 (at the district level controlling for year fixed effects, in column 2 of Table 4.4) to 0.06 (at the provincial level controlling for department-year fixed effects, in column 3 of 4.5).

For the sake of brevity, we will discuss this elasticity as being equal to about 0.05 on average, which means that a 1% increase in the price of quinoa is associated with a 0.05% increase in household welfare on average in those geographical units where quinoa is consumed in Peru for the period 2004-2014. From a macroeconomic perspective, this suggests that the increase in the price of quinoa over the period 2004-2014 has had positive general equilibrium effects extending to consumers of quinoa in addition to producers of quinoa. More specifically, between 2013 and 2014, the international price of quinoa rose by 25%, which would be associated with a 1.25% increase in household welfare. Though it is impossible to determine the precise mechanism through which this might

have happened, this likely took place via a multiplier effect.

4.4.2 The Welfare Effects of Rising Quinoa Prices on Producers

Tables 4.7 to 4.9 present estimation results for our analysis the welfare impacts of rising quinoa prices on quinoa producers. In each table, the coefficient on the logarithm of the international price of quinoa interacted with the proportion of quinoa producers is an estimate of the quinoa price elasticity of household welfare, on average, for those households in geographical units where quinoa was produced for the period 2004-2014.

The results from our core specification at the district level, shown in Table 4.7, suggest that the elasticity of household welfare with respect to the price of quinoa ranges from 0.011 (controlling for district and department-year fixed effects, in column 2 of Table 4.8) to 0.015 (controlling for district and province-year fixed effects, in column 2 of Table 4.7). More specifically, between 2013 and 2014, the price of quinoa rose by 25%, which would be associated with a 3.5% to 4% increase in household welfare of quinoa producers.

With that said, as the size of the geographical unit of observation increases, from district to province (Tables 4.7 to 4.9), and then from province to department (Tables 4.7 to 4.9), we find that the size of the point estimate decreases and even turns negative (at the provincial level controlling for year fixed effects in column 1 of Table 4.8, and at the departmental level controlling for year fixed effects in column 1 of Table 4.9.) A comparison of the average number of quinoa producers given that the region produces quinoa may help explain some of these differences in the point estimates. For departments that produce quinoa the average percentage of households that were producers was 6%, while it is 16% and 29% for province and districts that have any quinoa production. This increase in the point estimate as the geographical units get smaller is not surprising as quinoa producing departments contain provinces or districts that do not produce quinoa. Additionally, the decreasing statistical power is consistent with there being less classical

measurement error the more observations go into making the relevant averages, and so with there being less attenuation bias. The downside of considering larger geographical units, as we mentioned earlier, is that the precision of our estimates declines with the size of the geographical unit of observation, given the reduction in statistical power as the number of observations falls.

As was the case for the effect of quinoa prices for consumers, we find that the estimated coefficient on the interaction of international quinoa price and proportion of producers reverses its sign when treating the department as our unit of observation. The fact that this is similar to what we found in the case of consumers supports our hypothesis that the lack of significance at this level is due to the fact that our sample size is drastically decreased when moving from the province to the department.

4.5 Summary and Concluding Remarks

We have investigated whether the sharp rise in the international price of quinoa over the period 2004-2014 has had any impact on the welfare of quinoa consumers and producers in Peru. On the demand side, we find that an increase in the price of quinoa translates into positive effects on the welfare of consumers. Specifically, a 1% increase in the purchase price of quinoa is associated with a 0.04%-0.06% increase in the welfare of quinoa-consuming households. On the supply side, we find evidence that the rising price of quinoa has had a positive effect on the welfare of producer households. Specifically, a 1% increase in the international price of quinoa is associated with a 0.014%-0.016% increase in the welfare of quinoa-producing households.

The findings in this chapter are important for several reasons. First, Peruvian quinoa producers are particularly poor, with an average consumption that is still only about half of that of households that do not produce any quinoa. Recall that in 2013, some

people advocated that consumers in rich countries feel guilty about and reduce their consumption of quinoa because the rising international demand for quinoa was hurting those who had traditionally produced and consumed it. It is useful to know that the claim that rising quinoa prices were hurting those who had traditionally produced and consumed it—those households in our sample that produce quinoa—was patently false. Second, the positive general equilibrium effects of rising quinoa prices that we identify for those households that consume quinoa are interesting in and of themselves. Indeed, though Deaton (1989a)'s short-term, partial-equilibrium measure of the welfare impacts of an increase in the price of a commodity would suggest that quinoa consumers would be hurt by rising quinoa prices, our longer-term estimates show that for a 1% increase in the price of quinoa, household welfare increases by a modest 0.05%. These findings should assuage rich-country consumers' concerns about whether their growing demand for quinoa is having a negative influence on Andean households.

With that said, our analysis raises important questions for future research that are well beyond the scope of this chapter. For example, what about the indirect effects of rising quinoa prices? These could include nutritional and health outcomes,[21] agricultural wages, technology adoption, or educational outcomes. Second, though quinoa producers tend to be poorer, our analysis does not get into the distributional effects of rising quinoa prices, nor does it look at changes in poverty rates. For now, we leave these questions to future research.

---

[21]Stevens (2015) looks at whether the quinoa price boom has affected nutritional outcomes in the Peruvian regions where quinoa has traditionally been consumed and finds no negative effects of rising quinoa prices on nutrition.

Figure 4.1: Total Consumption and International Prices of Quinoa, 2004-2014.
Ratio of value in year *t* to value in 2004
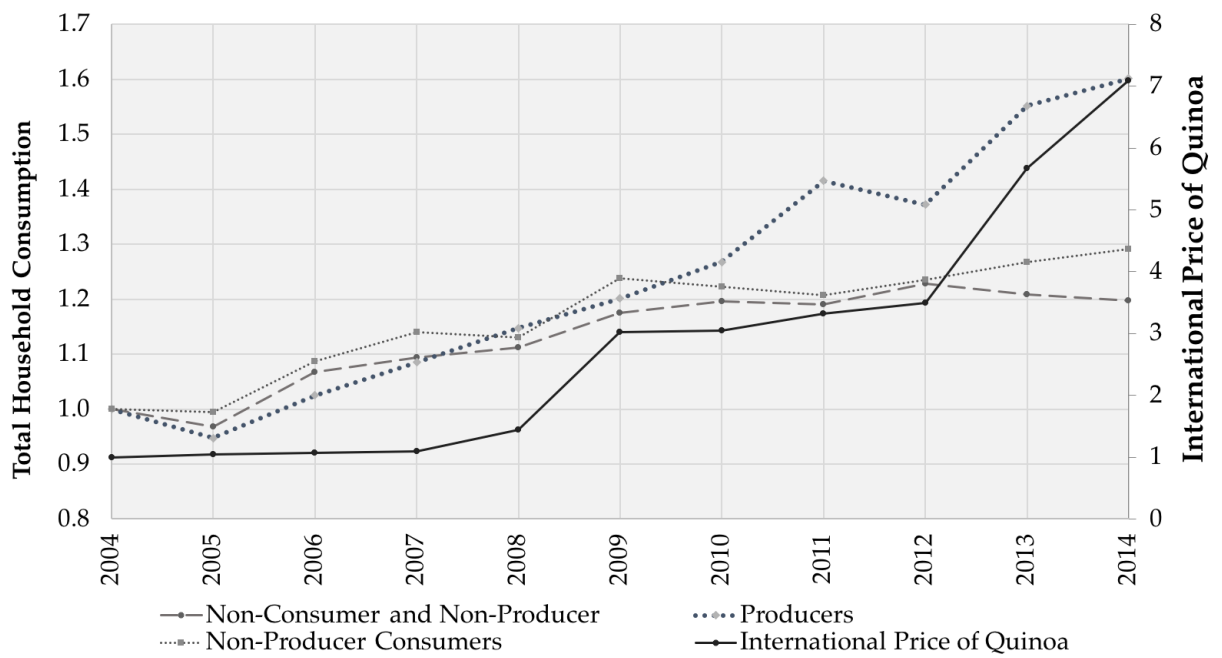
Source: ENAHO (sampling weights used) and FAO (prices)

Table 4.1: Household Welfare Trends in Constant Terms, 2004-2014

Year	Non-Consumers and Non-Producers of Quinoa		Producers of Quinoa		Non-Producer Consumers of Quinoa	
	% Households	Value	% Households	Value	% Households	Value
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2004	76.0%	14,474.84	3.69%	6,183.24	20.5%	18,952.50
2005	74.1%	14,014.12	3.92%	5,857.40	22.3%	18,852.89
2006	73.5%	15,450.09	3.90%	6,338.86	22.8%	20,584.68
2007	74.2%	15,841.39	3.69%	6,706.97	21.8%	21,605.90
2008	77.1%	16,100.50	3.06%	7,094.08	19.3%	21,410.90
2009	78.2%	17,007.19	3.38%	7,427.77	17.9%	23,461.13
2010	76.7%	17,304.01	3.56%	7,839.03	19.0%	23,168.29
2011	75.9%	17,236.37	3.38%	8,748.95	20.5%	22,890.31
2012	74.9%	17,777.43	2.81%	8,483.50	22.0%	23,397.33
2013	73.0%	17,498.58	2.63%	9,595.53	24.0%	24,013.04
2014	74.4%	17,340.45	3.34%	9,901.29	21.9%	24,480.29

Note: Figures measured in 2004 PEN and exclude consumption of cultivated and purchased quinoa.

All descriptive statistics are weighted using the sampling weights provided in the ENAHO.

In addition to being expressed in constant (i.e., 2004) terms, all prices are deflated using departmental-level deflators.

Table 4.2: Descriptive Statistics for Quinoa Consumers, 2004-2014

Year	Proportion of Quinoa-Consuming Households (%)	Kg of Whole Quinoa Purchased, Past 2 Weeks	Purchase Price of Whole Quinoa Per Kg, 2004 PEN	Budget Share of Annual Total Consumption of Quinoa, All Households (%)
(1)	(2)	(3)	(4)	(5)
2004	26.84%	0.87	3.15	0.36%
2005	30.70%	0.80	3.28	0.39%
2006	30.56%	0.84	3.17	0.37%
2007	29.60%	0.83	3.17	0.37%
2008	25.66%	0.75	4.18	0.53%
2009	24.64%	0.68	6.17	0.56%
2010	25.80%	0.73	6.29	0.54%
2011	27.90%	0.75	6.09	0.56%
2012	29.37%	0.71	6.10	0.57%
2013	30.83%	0.69	7.56	0.63%
2014	29.71%	0.64	11.27	0.73%

Note: Average purchase amount for households who purchased quinoa. In addition to being expressed in constant (i.e., 2004) terms, all prices are deflated using annual departmental-level deflators. Budget shares are imputed by multiplying the value of purchases in the previous two weeks by 26 and dividing by total household consumption.

Table 4.3: Descriptive Statistics for Quinoa Producers, 2004-2014

Year	Sample Proportion of Quinoa Producers (%)	Sample Proportion of Quinoa Sellers (%)	Quinoa Production, Past 12 Months (Kg), Quinoa Producers Only	Quinoa Production for Own Consumption (%), Quinoa Producers Only	Average Sales Price (Per Kg, 2004 PEN)	Quinoa Revenue (Quinoa Sellers Only, 2004 PEN)	Quinoa Revenue (All Quinoa Farmers, 2004 PEN)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2004	3.69%	0.30%	69.02	85.10%	1.34	173.38	14.07
2005	3.92%	0.42%	63.15	81.96%	1.57	206.41	22.28
2006	3.90%	0.37%	70.49	81.75%	1.62	202.40	19.17
2007	3.69%	0.31%	56.44	68.34%	1.42	101.37	8.38
2008	3.06%	0.20%	39.60	74.76%	1.91	191.71	12.59
2009	3.38%	0.30%	49.15	68.84%	3.35	419.34	37.40
2010	3.56%	0.29%	51.80	68.54%	2.98	299.60	24.45
2011	3.38%	0.50%	70.58	63.19%	2.99	449.60	66.53
2012	2.81%	0.46%	86.53	62.53%	3.33	787.66	128.73
2013	2.63%	0.46%	75.85	67.04%	4.44	855.47	149.02
2014	3.34%	0.66%	158.64	64.58%	6.18	2045.47	403.49

Note: All descriptive statistics are weighted using the sampling weights provided in the ENAHO.

In addition to being expressed in constant (i.e., 2004) terms, all prices are deflated using departmental-level deflators.

Table 4.4: District-Level Pseudo-Panel Regression of Total Household Consumption on the Price of Quinoa, 2004-2014

Variables	(1)	(2)	(3)
Dependent Variable: (Log) Total Value of Household Consumption			
Quinoa Consumers x (Log) International Price of Quinoa	0.042***	0.045***	0.045***
	(0.003)	(0.003)	(0.003)
Constant	8.724***	8.728***	8.724***
	(0.012)	(0.011)	(0.012)
Number of Districts	1,470	1,470	1,470
R-squared	0.204	0.455	0.283
District Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	No	No
Province-Year Fixed Effects	No	Yes	No
Department-Year Fixed Effects	No	No	Yes

Note: *, **, and *** denote statistical significance at the 10, 5, and 1 % levels, respectively. The sample only includes district-year observations where quinoa was consumed. Standard errors clustered at the district level are shown in parentheses. Each household is weighted according to the sampling weight it was given in the ENAHO. In addition to being expressed in constant (i.e., 2004) terms, all prices are deflated using departmental-level deflators.

Table 4.5: Province-Level Pseudo-Panel Regression of Total Household Consumption on the Price of Quinoa, 2004-2014

Variables	(1)	(2)
Dependent Variable: (Log) Total Value of Household Consumption		
Quinoa Consumers x (Log) International Price of Quinoa	0.051***	0.062***
	(0.010)	(0.011)
Constant	8.733***	8.721***
	(0.023)	(0.025)
Number of Provinces	194	194
R-squared	0.381	0.507
Province Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	No
Department-Year Fixed Effects	No	Yes

Note: *, **, and *** denote statistical significance at the 10, 5, and 1 % levels, respectively. The sample only includes district-year observations where quinoa was consumed. Standard errors clustered at the district level are shown in parentheses. Each household is weighted according to the sampling weight it was given in the ENAHO. In addition to being expressed in constant (i.e., 2004) terms, all prices are deflated using departmental-level deflators.

Table 4.6: Department-Level Pseudo-Panel Regression of Total Household Consumption on the Price of Quinoa, 2004-2014

Variables	(1)
Dependent Variable: (Log) Total Value of Household Consumption	
Quinoa Consumers x (Log) International Price of Quinoa	-0.010
	(0.023)
Constant	9.117***
	(0.046)
Number of Departments	25
R-squared	0.699
Department Fixed Effects	Yes
Year Fixed Effects	Yes

Note: *, **, and *** denote statistical significance at the 10, 5, and 1 % levels, respectively. The sample only includes district-year observations where quinoa was consumed. Standard errors clustered at the district level are shown in parentheses. Each household is weighted according to the sampling weight it was given in the ENAHO. In addition to being expressed in constant (i.e., 2004) terms, all prices are deflated using departmental-level deflators.

Table 4.7: District-Level Pseudo-Panel Regression of Total Household Consumption on the Price of Quinoa, 2004-2014

Variables	(1)	(2)	(3)
Dependent Variable: (Log) Total Value of Household Consumption			
Quinoa Producers x (Log) International Price of Quinoa	0.014**	0.015***	0.014***
	(0.005)	(0.005)	(0.005)
Constant	8.783***	8.793***	8.788***
	(0.011)	(0.010)	(0.011)
Number of Districts	1,470	1,470	1,470
R-squared	0.181	0.435	0.259
District Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	No	No
Province-Year Fixed Effects	No	Yes	No
Department-Year Fixed Effects	No	No	Yes

Note: *, **, and *** denote statistical significance at the 10, 5, and 1 % levels, respectively. The sample only includes district-year observations where quinoa was consumed. Standard errors clustered at the district level are shown in parentheses. Each household is weighted according to the sampling weight it was given in the ENAHO. In addition to being expressed in constant (i.e., 2004) terms, all prices are deflated using departmental-level deflators.

Table 4.8: Province-Level Pseudo-Panel Regression of Total Household Consumption on the Price of Quinoa, 2004-2014

Variables	(1)	(2)
Dependent Variable: (Log) Total Value of Household Consumption		
Quinoa Producers x (Log) International Price of Quinoa	-0.014	-0.012
	(0.013)	(0.014)
Constant	8.822***	8.827***
	(0.016)	(0.016)
Number of Provinces	189	189
R-squared	0.414	0.491
Province Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	No
Department-Year Fixed Effects	No	Yes

Note: *, **, and *** denote statistical significance at the 10, 5, and 1 % levels, respectively. The sample only includes district-year observations where quinoa was consumed. Standard errors clustered at the district level are shown in parentheses. Each household is weighted according to the sampling weight it was given in the ENAHO. In addition to being expressed in constant (i.e., 2004) terms, all prices are deflated using departmental-level deflators.

Table 4.9: Department-Level Pseudo-Panel Regression of Total Household Consumption on the Price of Quinoa, 2004-2014

Variables	(1)
Dependent Variable: (Log) Total Value of Household Consumption	
Quinoa Producers x (Log) International Price of Quinoa	-0.058
	(0.097)
Constant	9.116***
	(0.034)
Number of Departments	25
R-squared	0.761
Department Fixed Effects	Yes
Year Fixed Effects	Yes

Note: *, **, and *** denote statistical significance at the 10, 5, and 1 % levels, respectively. The sample only includes district-year observations where quinoa was consumed. Standard errors clustered at the district level are shown in parentheses. Each household is weighted according to the sampling weight it was given in the ENAHO. In addition to being expressed in constant (i.e., 2004) terms, all prices are deflated using departmental-level deflators.

Chapter 5

Conclusion

This dissertation studied three topics in development economics: (i) the effect of gender-biased violence against women on female employment in Colombia; (ii) the contribution of early life circumstances to inequality of opportunity in adult health in Colombia; and (iii) the welfare effects of rising quinoa prices in Peru.

Chapter 2 showed that the incidence of intimate partner violence increases the likelihood of female employment by about 16 percentage points for a sample of women in Colombia. Women's decision-making power likely explains this finding, as abused women may work to enhance their decision-making power and escape violent situations at home. These results suggest some important policy implications. That women victims of intimate partner violence are more likely to work suggests that they may benefit from counseling and legal help inside and outside the workplace. This is particularly important since previous studies suggest that intimate partner violence has negative effects on labor productivity (Farmer and Tiefenthaler, 2004).

Chapter 3 provided suggestive evidence of the lasting effects of childhood circumstances on adult health in Colombia. This chapter also showed how the transmission channels of health inequality across generations seem to operate differently in rural and urban areas. To equalize the opportunity to achieve a healthy adulthood, more specific policies should be designed to offset the effects of unequal circumstances, in particular those related to human capital formation early in life.

Chapter 4 showed that the effects of increasing international prices of quinoa on changes in the welfare of rural households in Peru are likely positive, although modest. In particular, for quinoa-consuming households, longer-term estimates show that for a 1% increase in the price of quinoa, household welfare increases by a modest 0.05%, which suggests that the quinoa price increase has had general equilibrium effects extending to non-producers. This analysis raises important questions for future research. For example, the effects of rising quinoa prices on nutritional and health outcomes, agricultural wages, technology adoption, or educational outcomes are yet to be studied.

Bibliography

Abras, A., A. Hoyos Suarez, A. Narayan, and S. Tiwari. 2013. "Inequality of Opportunities in the Labor Market: Evidence from Life in Transition Surveys in Europe and Central Asia." *IZA Journal Of Labor & Development* 2:2–22.

Acharya, A., M. Blackwell, and M. Sen. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110:512–529.

Ackerberg, D.A., and M. Botticini. 2002. "Endogenous Matching and the Empirical Determinants of Contract Form." *Journal of Political Economy* 110:564–591.

Agüero, J. 2013. "Causal Estimates of the Intangible Costs of Violence against Women in Latin America and the Caribbean." Inter American Development Bank Working Paper No. IDB-WP-414.

Aizer, A. 2010. "The Gender Wage Gap and Domestic Violence." *The American Economic Review* 100:18–47.

Anand, P., L. Roope, and A. Gray. 2013. "Missing Dimensions in the Measurement of Wellbeing and Happiness." Working paper, Mimeo.

Andina. 2016. "Peru: Agro-exports Exceeded US$5.20 Billion in 2015." Online; posted February 7, 2016. Available: `http://www.andina.com.pe/Ingles/noticia-peru-agroexports-exceeded-520-billion-in-2015-597855.aspx` [Last accessed: March 3,2016].

Angrist, J.D., and J.S. Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Angulo, R., J. Azevedo, A. Gaviria, and G. Páez. 2012. "Movilidad Social en Colombia." *Documentos CEDE* 43.

Antman, F., and D. McKenzie. 2007a. "Earnings Mobility and Measurement Error: A PseudoPanel Approach." *Economic Development and Cultural Change* 56:125–161.

—. 2007b. "Poverty Traps and Nonlinear Income Dynamics with Measurement Error and Individual Heterogeneity." *The Journal of Development Studies* 43:1057–1083.

Aragón, F.M., and J.P. Rud. 2013. "Natural Resources and Local Communities: Evidence from a Peruvian Gold Mine." *American Economic Journal: Economic Policy* 5:1–25.

Arendt, J.N. 2005. "Does Education Cause Better Health? A Panel Data Analysis Using School Reforms for Identification." *Economics of Education Review* 24:149–160.

Assaad, R., G. Friedemann-Sanchez, and D. Levison. 2016. "Impact of Domestic Violence on Children's Education in Colombia: Methodological Challenges." *Violence Against Women*, pp. 1–29.

Barrett, C.B., and P.A. Dorosh. 1996. "Farmers' Welfare and Changing Food Prices: Nonparametric Evidence From Rice in Madagascar." *American Journal of Agricultural Economics* 78:656–669.

Bellemare, M.F., C.B. Barrett, and D.R. Just. 2013. "The Welfare Impacts of Commodity Price Volatility: Evidence from Rural Ethiopia." *American Journal of Agricultural Economics* 95:877–899.

Bhattacharyya, M., A.S. Bedi, and A. Chhachhi. 2011. "Marital Violence and Women's Employment and Property Status: Evidence from North Indian Villages." *World Development* 39:1676–1689.

Blackwell, M. 2016. "Instrumental Variable Methods for Conditional Effects and Causal Interaction in Voter Mobilization Experiments." *Journal of the American Statistical Association*, pp. Forthcoming.

Blythman, J. 2013. "Can Vegans Stomach the Unpalatable Truth about Quinoa." *The Guardian* 16.

Christiaensen, L.J., and K. Subbarao. 2005. "Towards an Understanding of Household Vulnerability in Rural Kenya." *Journal of African Economies* 14:520–558.

Clarke, D. 2014. "plausexog: Stata Module to implement Conley et al's Plausibly Exogenous Bounds." Working paper, Boston College Department of Economics.

Cogneau, D., and R. Jedwab. 2012. "Commodity Price Shocks and Child Outcomes: The 1990 Cocoa Crisis in Cote dIvoire." *Economic Development and Cultural Change* 60:507–534.

Conley, T.G., C.B. Hansen, and P.E. Rossi. 2012. "Plausibly Exogenous." *Review of Economics and Statistics* 94:260–272.

Cruces, G., P. Lanjouw, L. Lucchetti, E. Perova, R. Vakis, and M. Viollaz. 2015. "Estimating Poverty Transitions Using Repeated Cross-Sections: A Three-Country Validation Exercise." *The Journal of Economic Inequality* 13:161–179.

Cuesta, J., H. Ñopo, and G. Pizzolitto. 2011. "Using Pseudo-Panels to Measure Income Mobility in Latin America." *Review of Income and Wealth* 57:224–246.

Cutler, D.M., A. Lleras-Muney, and T. Vogl. 2008. "Socioeconomic Status and Health: Dimensions and Mechanisms." Working paper, National Bureau of Economic Research.

De Salvo, K., V.S. Fan, M.B. Mcdonell, and S.D. Fihn. 2005. "Predicting Mortality and Healthcare Utilization with a Single Question." *Health Services Research* 40:1234–1246.

Deaton, A. 1989a. "Household Survey Data and Pricing Policies in Developing Countries." *The World Bank Economic Review* 3:183–210.

—. 1985. "Panel Data from Time Series of Cross-Sections." *Journal of Econometrics* 30:109–126.

—. 1989b. "Rice Prices and Income Distribution in Thailand: A Non-parametric Analysis." *The Economic Journal* 99:1–37.

Dell, M. 2010. "The Persistent Effects of Peru's Mining Mita." *Econometrica* 78:1863–1903.

DePillis, L. 2013. "Quinoa Should be Taking Over the World. This is Why it isnt." *Washington Post. July* 11.

Donni, P.L., V. Peragine, and G. Pignataro. 2014. "Ex-Ante and Ex-Post Measurement of Equality of Opportunity in Health: A Normative Decomposition." *Health Economics* 23:182–198.

Dube, O., and J.F. Vargas. 2013. "Commodity Price Shocks and Civil Conflict: Evidence from Colombia." *The Review of Economic Studies* 80:1384–1421.

Ersado, L., and M.A. Aran. 2014. "Inequality of Opportunity among Egyptian Children." *The World Bank Policy Research Working Paper 7026*, pp. .

Eswaran, M., and N. Malhotra. 2011. "Domestic Violence and Women's Autonomy in Developing Countries: Theory and Evidence." *Canadian Journal of Economics/Revue canadienne d'économique* 44:1222–1263.

Evans, S.E., C. Davies, and D. DiLillo. 2008. "Exposure to Domestic Violence: A Meta-Analysis of Child and Adolescent Outcomes." *Aggression and Violent Behavior* 13:131–140.

Farmer, A., and J. Tiefenthaler. 2004. "The Employment Effects of Domestic Violence." *Research in Labor Economics* 23:301–334.

Ferreira, F.H., and J.É. Gignoux. 2014. "the Measurement of Educational Inequality: Achievement and Opportunity." *The World Bank Economic Review* 28:210–246.

Filmer, D., and L.H. Pritchett. 2001. "Estimating Wealth Wffects Without Expenditure Dataor Tears: An Application to Educational Enrollments in States of India." *Demography* 38:115–132.

Flake, D.F., and R. Forste. 2006. "Fighting Families: Family Characteristics Associated with Domestic Violence in Five Latin American Countries." *Journal of Family Violence* 21:19–29.

Fleurbaey, M., and E. Schokkaert. 2009. "Unfair Inequalities in Health and Health Care." *Journal of Health Economics* 28:73–90.

Flórez, C., V. Soto, O. Acosta, C. Karl, J. Misas, N. Forero, and C. Lopera. 2007. "Avances y Desafos de la equidad en el Sistema de Salud Colombiano." *Bogot á: Fundacinón Corona, Departamento Nacional De Planeacinón, Universidad Del Rosario, Universidad De Los Andes*, pp. .

Friedemann-Sánchez, G. 2006. *Assembling Flowers and Cultivating Homes: Labor and Gender in Colombia*. Lexington Books.

Friedemann-Sánchez, G., and R. Lovatón. 2012. "Intimate Partner Violence in Colombia: Who Is at Risk?" *Social Forces* 91:663–688.

Galdo, J. 2013. "The Long-Run Labor-Market Consequences of Civil War: Evidence from The Shining Path in Peru." *Economic Development and Cultural Change* 61:789–823.

Graham, C. 2008. "Happiness and Health: Lessonsand Questionsfor Public Policy." *Health affairs* 27:72–87.

Graham-Bermann, S.A., L.E. Castor, L.E. Miller, and K.H. Howell. 2012. "The Impact of Intimate Partner Violence and Additional Traumatic Events on Trauma Symptoms and PTSD in Preschool-Aged Children." *Journal of Traumatic Stress* 25:393400.

Heath, R. 2014. "Women's Access to Labor Market Opportunities, Control of Household Resources, and Domestic Violence: Evidence from Bangladesh." *World Development* 57:32–46.

Heckman, J.J., and J.A. Smith. 1995. "Ashenfelter's Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies.", pp. .

Howes, S. 1996. "A new test for inferring dominance from sample data." *World Bank, Washington, DC*, pp. .

Hoyos Suarez, A. 2013. "Hoishapley: Stata Module to Perform Shapley Decomposition of the Human Opportunity Index." Working paper.

Hoyos Suarez, A., and A. Narayan. 2012. "Inequality of Opportunities among Children: How Much Does Gender Matter?" *Background Paper For the 2013 World Development Report*, pp. .

Idler, E.L., and Y. Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior*, pp. 21–37.

Imai, K., L. Keele, D. Tingley, and T. Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105:765–789.

Ivanic, M., and W. Martin. 2008. "Implications of Higher Global Food Prices for Poverty in Low-Income Countries1." *Agricultural economics* 39:405–416.

Jepsen, L.K., and C.A. Jepsen. 2002. "An Empirical Analysis of the Matching Patterns of Same-Sex and Opposite-Sex Couples." *Demography* 39:435–453.

Jusot, F., S. Mage, and M. Menendez. 2014. "Inequality of Opportunity in Health in Indonesia." *Dial (DÉVeloppement, Institutions Et Mondialisation) Working Papers Dt/2014–6*, pp. .

Jusot, F., S. Tubeuf, and A. Trannoy. 2010. "Inequality of Opportunities in Health in Europe: Why So Much Difference Across Countries?" *Health, Econometrics and Data Group (Hedg) Working Papers* 10:26.

Kalmijn, M. 1994. "Assortative Mating by Cultural and Economic Occupational Status." *American journal of Sociology*, pp. 422–452.

Kishor, S., and K. Johnson. 2004. "Profiling Domestic Violence: A Multi-Country Study." Measure DHS+ and ORC Macro Publication OD31, Available as `http://www.dhsprogram.com/publications/publication-OD31-Other-Documents.cfm`. Accessed: May 7, 2014.

Lam, D. 1988. "Marriage Markets and Assortative Mating with Household Public Goods: Theoretical Results and Empirical Implications." *Journal of Human Resources*, pp. 462–487.

Lefranc, A., N. Pistolesi, and A. Trannoy. 2009. "Equality of Opportunity and Luck: Definitions and Testable Conditions, with an Application to Income in France." *Journal of Public Economics* 93:1189–1207.

Lleras-Muney, A. 2005. "The Relationship Between Education and Adult Mortality in the United States." *The Review of Economic Studies* 72:189–221.

Lloyd, S. 1997. "The Effects of Domestic Violence on Women's Employment." *Law and Policy* 19:139–167.

Lloyd, S., and N. Taluc. 1999. "The Effects of Male Violence on Female Employment." *Violence Against Women* 5:370–392.

Martin, S.L., K.E. Moracco, J. Garro, A.O. Tsui, L.L. Kupper, J.L. Chase, and J.C. Campbell. 2002. "Domestic Violence Across Generations: Findings from Northern India." *International Journal of Epidemiology* 31:560 – 572.

McKenzie, D.J. 2004. "Asymptotic Theory for Heterogeneous Dynamic Pseudo-Panels." *Journal of Econometrics* 120:235–262.

Meisel, J., D. Chandler, and B.M. Rienzi. 2003. "Domestic Violence Prevalence and Effects on Employment In Two California TANF Populations." *Violence Against Women* 9:1191–1212.

Montiel Olea, J.L., and C.E. Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business and Economic Statistics* 31:358–369.

Morrison, A.R., and M.B. Orlando. 1999. "Social and Economic Costs of Domestic Violence: Chile and Nicaragua." In A. Morrison and L. Biehl, eds. *Too Close to Home: Domestic Violence in the Americas*. Inter American Development Bank, pp. 51–80.

Nevo, A., and A.M. Rosen. 2012. "Identification with Imperfect Instruments." *Review of Economics and Statistics* 94:659–671.

Oxford English Dictionary. 2014. "Superfood." Online. Available: `http://www.oed.com/view/Entry/194186?redirectedFrom=superfood#eid69476470` [Last accessed: October 8, 2015].

Paes De Barros, R., F. Ferreira, J. Molinas, and J. Saavedra. 2009. *Measuring Inequality of Opportunities in Latin America and the Caribbean*. The World Bank.

Paes De Barros, R., J.R.M. Vega, and J. Saavedra. 2008. "Measuring Inequality of Opportunities for Children." Working paper, Unpublished, The World Bank.

Profamilia. 2011. *Encuesta Nacional de Demografía y Salud 2010*. Bogota, Colombia: Gerencia de Evaluacion e Investigacion Profamilia.

Ramos, X., and D. Van de Gaer. 2015. "Approaches to Inequality of Opportunity: Principles, Measures and Evidence." *Journal of Economic Surveys*, pp. .

Restrepo, J., A. Zambrano, M. Velez, and M. Ramirez-Gomez. 2007. "Health Insurance as a Strategy for Access: Streamlined Facts of the Colombian Health Care Reform." Working paper.

Rios-Avila, F., and G. Canavire-Bacarreza. 2017. "The Effect of Intimate Partner Violence on Labor Market Decisions: Evidence from a Multi-Ethnic Country." *International Journal of Social Economics* 44:75–92.

Roemer, J.E. 1998. *Equality of Opportunity*. Harvard University Press.

Rosa Dias, P. 2014. "Equality of opportunity in health." In A. J. Culyer, ed. *Encyclopedia of Health Economics*. Elsevier, pp. 282–287.

Rosa Dias, P. 2009. "Inequality of Opportunity in Health: Evidence From a UK Cohort Study." *Health Economics* 18:1057–1074.

Rosa Dias, P., and A.M. Jones. 2007. "Giving Equality of Opportunity a Fair Innings." *Health Economics* 16:109–112.

Saunders, D. 2013. "Killer Quinoa? Time to Debunk These Urban Food Myths." Online; In: *The Globe and Mail*, January 19, 2013. Available: `http://www.theglobeandmail.com/globe-debate/chow-down-on-quinoa-and-three-modern-food-fallacies/article7536845/` [Last accessed: October 19, 2014].

Sen, A. 2002. "Health: Perception Versus Observation: Self-Reported Morbidity Has Severe Limitations and Can Be Extremely Misleading." *British Medical Journal* 324:860–861.

Shorrocks, A.F. 2013. "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value." *Journal of Economic Inequality*, pp. 1–28.

Staggs, S.L., and S. Riger. 2005. "Effects of Intimate Partner Violence on Low-Income Women's Health and Employment." *American Journal of Community Psychology* 36:133–145.

Stevens, A. 2015. "Quinoa Quandary: Cultural Tastes and Nutrition in Peru.", pp. . Unpublished Draft.

Straus, M.A. 1979. "Measuring Intrafamily Conflict and Violence: The Conflict Tactics (CT) Scales." *Journal of Marriage and the Family*, pp. 75–88.

Straus, M.A., S.L. Hamby, S. Boney-McCoy, and D.B. Sugarman. 1996. "The Revised Conflict Tactics Scales (CTS2): Development and Preliminary Psychometric Data." *Journal of Family Issues* 17:283–316.

Swanberg, J.E., and T. Logan. 2005. "Domestic violence and Employment: A Qualitative Study." *Journal of Occupational Health Psychology* 10:3.

The World Bank. 2015. *Colombia: Systematic Country Diagnostic*. The World Bank.

—. 2005. *World Development Report 2006: Equity and Development*. Oxford University Press, Incorporated.

The World Health Organization. 2006. *Country Cooperation Brief Strategy*. The World Health Organization.

Tolman, R.M., and H.C. Wang. 2005. "Domestic Violence and Women's Employment: Fixed Effects Models of Three Waves of Women's Employment Study Data." *American Journal of Community Psychology* 36:147–158.

Trannoy, A., S. Tubeuf, F. Jusot, and M. Devaux. 2010. "Inequality of opportunities in health in France: A First Pass." *Health Economics* 19:921–938.

Van Doorslaer, E., and U.G. Gerdtham. 2003. "Does Inequality in Self-Assessed Health Predict Inequality in Survival by Income? Evidence from Swedish Data." *Social Science & Medicine* 57:1621–1629.

Vargas-Lorenzo, I. 2009. "Barreras en el Acceso a la Atención en Salud en Modelos de Competencia Gestionada: Un Estudio de Caso en Colombia."

Vyas, S., and L. Kumaranayake. 2006. "Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis." *Health Policy and Planning* 21:459–468.

Whitfield, C.L., R.F. Anda, S.R. Dube, and V.J. Felitti. 2003. "Violent Childhood Experiences and The Risk of Intimate Partner Violence in Adults." *Journal of Interpersonal Violence* 18:166–185.

Wodon, Q., and H. Zaman. 2010. "Higher Food Prices in Sub-Saharan Africa: Poverty Impact and Policy Responses." *The World Bank Research Observer* 25:157–176.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. The MIT press.

World Health Organization. 2013. *Global and Regional Estimates of Domestic Violence Against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-Partner Sexual Violence*. Geneva: World Health Organization Department of Reproductive Health and Research.

Yalonetzky, G. 2012. "A Dissimilarity Index of Multidimensional Inequality of Opportunity." *Journal of Economic Inequality* 10:343–373.

—. 2013. "Stochastic dominance with ordinal variables: conditions and a test." *Econometric Reviews* 32:126–163.

Zezza, A., B. Davis, C. Azzarri, K. Covarrubias, L. Tasciotti, G. Anriquez, et al. 2008. "The Impact of Rising Food Prices on The Poor." *FAO–ESA Working Paper*, pp. .

Appendix A.
Inequality of Opportunity in Adult Health

A.1 Additional Tables

Table A.1: Summary Statistics: Urban Subsample

Variable	Observations	Mean or Proportion	Std. Dev.
Outcome			
Self-assessed Health Status	1,263	2.85	0.60
Poor	25	2.0%	0.14
Fair	258	20.4%	0.40
Good	856	67.8%	0.47
Excellent	124	9.8%	0.30
Early-life Circumstances			
Household Socioeconomic Status at Age 10			
Quintile Group			
1 (Lowest)	265	21.0%	0.41
2	252	20.0%	0.40
3	253	20.0%	0.40
4	243	19.2%	0.39
5 (highest)	237	18.8%	0.39
No information on assets available	13	1.0%	0.10
Education Level of Father			
None or Incomplete Primary	585	46.3%	0.50
Complete Primary and Incomplete Secondary	289	22.9%	0.42
Complete Secondary or More	177	14.0%	0.35
Unknown Father's Education	210	16.6%	0.37
No information on father's education	2	0.2%	0.04
Education Level of Mother			
None or Incomplete Primary	647	51.2%	0.50
Complete Primary and Incomplete Secondary	333	26.4%	0.44
Complete Secondary or More	151	12.0%	0.32
Unknown Mother's Education	130	10.3%	0.30
No information on mother's education	2	0.2%	0.04
Other circumstances			
Ethnicity			
Indigenous	22	1.7%	0.13
Black, *mulato, raizal* or *palenquero*	80	6.3%	0.24
No ethnic minority	1,161	91.9%	0.27
Years of Education	1,263	8.83	4.54
Born in Urban Area	899	71.2%	0.45
Born in Rural Area	359	28.4%	0.45
No information on area of birth	5	0.4%	0.06
Region of Birth			
Atlantic	259	20.5%	0.40
Eastern	325	25.7%	0.44
Pacific	74	5.9%	0.23
Orinoquia-Amazonia	5	0.4%	0.06
Antioquia	146	11.6%	0.32
Valle del Cauca	102	8.1%	0.27
Bogotá	153	12.1%	0.33
San Andrés islands	2	0.2%	0.04
Central	197	15.6%	0.36
Additional Controls			
Male	811	64.2%	0.48
Age	1,263	45.13	10.96
Age group			
25–35	275	21.8%	0.41
35–45	315	24.9%	0.43
45–55	385	30.5%	0.46
55–65	288	22.8%	0.42

Note: Heads of Household between 25 and 65 years old. Total Number of Observations: 1,263
Source: 2010 Colombian LSSM Survey

Table A.2: Summary Statistics: Rural Subsample

Variable	Observations	Mean or Proportion	Std. Dev.
Outcome			
Self-assessed Health Status	990	2.69	0.58
Poor	24	2.4%	0.15
Fair	298	30.1%	0.46
Good	631	63.7%	0.48
Excellent	37	3.7%	0.19
Early-life Circumstances			
Household Socioeconomic Status at Age 10			
Quintile Group			
1 (lowest)	246	24.8%	0.43
2	158	16.0%	0.37
3	181	18.3%	0.39
4	194	19.6%	0.40
5 (highest)	185	18.7%	0.39
No information on assets available	26	2.6%	0.16
Education Level of Father			
None or Incomplete Primary	673	68.0%	0.47
Complete Primary and Incomplete Secondary	88	8.9%	0.28
Complete Secondary or More	17	1.7%	0.13
Unknown Father's Education	212	21.4%	0.41
Education Level of Mother			
None or Incomplete Primary	698	70.5%	0.46
Complete Primary and Incomplete Secondary	114	11.5%	0.32
Complete Secondary or More	20	2.0%	0.14
Unknown Mother's Education	158	16.0%	0.37
Other circumstances			
Ethnicity			
Indigenous	37	3.7%	0.19
Black, *mulato, raizal* or *palenquero*	64	6.5%	0.25
No ethnic minority	889	89.8%	0.30
Years of Education	990	4.71	3.66
Born in Urban Area	204	20.6%	0.41
Born in Rural Area	785	79.3%	0.40
No information on area of birth	1	0.1%	0.03
Region of Birth			
Atlantic	248	25.1%	0.43
Eastern	193	19.5%	0.40
Pacific	181	18.3%	0.39
Orinoquia-Amazonia	1	0.1%	0.03
Antioquia	105	10.6%	0.31
Valle del Cauca	58	5.9%	0.23
Bogotá	6	0.6%	0.08
Central	198	20.0%	0.40
Additional Controls			
Male	787	79.5%	0.40
Age	990	44.31	11.06
Age group			
25–35	229	23.1%	0.42
35–45	279	28.2%	0.45
45–55	261	26.4%	0.44
55–65	221	22.3%	0.42

Note: Heads of Household between 25 and 65 years old. Total Number of Observations: 990
Source: 2010 Colombian LSSM Survey

A.2 Stochastic Dominance Test for Ordinal Variables and Its Application to Inequality of Opportunity in Adult Health in Colombia

A.2.1 Stochastic Dominance and Inequality of Opportunity

Roemer (1998) defines equality of opportunity as a situation where individuals with similar efforts reach similar outcomes, regardless of their circumstances. More formally, under equality of opportunity, the probability distribution of health status H given effort e does not depend on circumstances C or C'. That is,

$$\forall C \neq C, \forall e, F(H \mid C, e) = F(H \mid C', e) \tag{A.1}$$

where $F(H \mid C, e)$ denotes the cumulative probability function.

Lefranc, Pistolesi, and Trannoy (2009) suggest that different health-related outcomes can be seen as alternative lotteries resulting from the effect of luck and other random factors that are equally distributed across individuals sharing the same efforts and circumstances.[1] These authors then show that a consistent definition of inequality of opportunity formulates that different conditional distributions of health can be ordered according to expected utility theory. In their paper, Lefranc, Trannoy and Pistolesi propose a criterion to assess inequality of opportunity using stochastic dominance relationships. The authors assume that health status is increasing in effort and that the relative effort can be inferred from the observation of health status and circumstances. Thus, inequality of opportunity is satisfied if and only if the distributions of health status conditional on different sets of circumstances can be ordered by first-order stochastic dominance, such that

---

[1] The authors also note that luck could lead to differences in individual health outcomes as long as it remains neutral with respect to circumstances.

$$\forall C \neq C', F(H \mid C) \succeq_{FSD} F(H \mid C') \tag{A.2}$$

A.2.2 A Stochastic Dominance Test for Ordinal Variables

Self-assessed health status is a categorical variable. In this case, the stochastic dominance test is performed using a non-parametric test proposed by Yalonetzky (2013), as the more familiar statistical tests for stochastic dominance such as the Kolmogorov-Smirnov or the Davidson-Duclos cannot be directly applied to outcomes that are ordinal and lack any cardinal meaning.

Anand, Roope, and Gray (2013) provide the univariate extension of the stochastic dominance test proposed by Yalonetzky (2013). In this appendix, I follow closely Anand, Roope and Gray's notation.

Let A be the subgroup of individuals who share exposure to circumstance category a (e.g., individuals whose mothers have incomplete primary education), and B the subgroup who share exposure to circumstance category b (e.g., individuals whose mothers have incomplete secondary education). The sample size of each group is denoted by n_A and n_B, respectively. Each individual in each group $g \in \{A, B\}$ reports a health status which lies in one of $S \in \mathbb{N}$ ordinal categories. Suppose there are N_g individuals in group $g \in \{A, B\}$. Each individual indicates a health status which lies in one of $S \in \mathbb{N}$ ordinal categories, in our case $S = 3$

Let $\mathbf{h_g} \in \mathbb{N}_{\uparrow}^{N_g}$ be a vector of health status scores, where the $\uparrow$ subscript indicates that the ordinal categories are ordered in terms of their desirability from the least to the most desired one. The i-th element of $\mathbf{h_g}$ is given by $h_{ig} \in \{1, \ldots, S\}$

For $k \in \{1, \ldots, S\}$, let $F_g(k) \equiv Pr(h_{ig} \leq k)$ denote the cumulative probability function. Furthermore, the difference in cumulative probability functions is defined as $\triangle F(\cdot) \equiv$

$$F_A(\cdot) - F_B(\cdot)$$

Now, let p_{kg} be the probability that a randomly selected individual from $G = \{1, \ldots, N_g\}$ has a health status in category $k \in \{1, \ldots, S\}$, and $\mathbf{p_g} \in [0,1]^S$ be the corresponding vector of probabilities. The empirical estimate of p_{kg} from a random sample $n_g \leq N_g$ is given by

$$\widehat{p_{kg}} = \frac{1}{n_g} \sum_{i=1}^{n_g} I(k_i) \tag{A.3}$$

where $I(k_i)$ is an indicator function that equals 1 when $k_i = k$.

The empirical estimates for the probability that a randomly selected individual from group g has a health status in category $j \in \{1,2\}$ are denoted by $\widehat{p_{jA}}$ and $\widehat{p_{jB}}$, respectively. Let $\widehat{\mathbf{p_g}}$ be the vector of empirical estimates of $\mathbf{p_g}$. Formby, Smith and Zheng (2004) show that the corresponding asymptotic result is given by

$$\sqrt{n_g}(\widehat{p_g} - p_g) \rightarrow_d N(0, \Omega_g) \tag{A.4}$$

where Ω_g is a S-dimensional covariance matrix whose (k,l)-th element is equal to $p_{kg}(1 - p_{kg})$ whenever $k = l$, and $-p_{kg}p_{lg}$ whenever $k \neq l$

Thus, under the null hypothesis that groups A and B are identically distributed, $\Omega_g = \Omega$ for any $g \in \{A, B\}$, so that

$$(\widehat{p_A} - \widehat{p_B}) \rightarrow_d N(0, \frac{n_A + n_B}{n_A n_B}\Omega) \tag{A.5}$$

The empirical estimate of Ω_g has corresponding elements $\widehat{p_{kg}(1 - p_{kg})}$ whenever $k = l$, and $-\widehat{p_{kg}p_{lg}}$ whenever $k \neq l$.

Let $\widehat{\triangle F}$ be the S-vector with $k-th$ element given by $\widehat{\triangle F} = \sum_{j=1}^{k}(\widehat{p_{jA}} - \widehat{p_{jB}})$ and $\boldsymbol{L}$ be a S-dimensional lower triangular matrix of ones. Under the assumption that A and B are independent, the estimated covariance matrix of the empirical difference in cumulative probability functions is given by

$$var(\widehat{\triangle F}) = L(\frac{1}{n_A}\Omega_A + \frac{1}{n_B}\Omega_B)L' \tag{A.6}$$

Thus, for each $k \in \{A, \ldots, B\}$, the corresponding z-statistic z_k^l is obtained by dividing $\widehat{\triangle F}$ by its respective standard error, which is given by the squared root of the $k-th$ diagonal element of $var(\widehat{\triangle F})$. More formally, a test for the hypothesis that A does not first-order-stochastic dominate B against the alternative that A first-order-stochastic dominates B is given by

$$H_0 = \triangle F(k) \geq 0 \quad for\ some \quad k \in \{1,2\}$$
$$H_1 = \triangle F(k) < 0 \quad for\ all \quad k \in \{1,2\} \tag{A.7}$$

The corresponding z-statistic, z_k^l, is given by

$$z_k^l = \frac{\sum_{j=1}^{k}(p_{jA} - p_{jB})}{\sqrt{\sum_{j=1}^{k}(\frac{p_{jA}(1-p_{jA})}{n_A} + \frac{p_{jB}(1-p_{jB})}{n_B} - \frac{p_{jA}}{n_A}\sum_{l=1,l\neq j}^{k}\widehat{p_{lA}} - \frac{p_{jB}}{n_B}\sum_{l=1,l\neq j}^{k}\widehat{p_{lB}})}} \tag{A.8}$$

The rejection rule proposed by Howes (1996) suggests that H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all $k \in \{1, \ldots, S-1\}$, where $-z^*$ is the left-tail critical value for a desired level of statistical significance.

A.2.3 Results

I perform $\dfrac{m!}{[(m-2)!]\,2}$ pairwise tests for each circumstance variable c that has m response categories. To assess the differences in inequality of opportunity between urban and rural residents, I perform separate statistical tests for the sample of all individuals, the subsample of individuals residing in rural areas, and the subsample of individuals residing in urban areas.

In this section, I empirically assess inequality of opportunity using the stochastic dominance approach. I analyze one circumstance at a time. In what follows, I refer to the group of individuals who share exposure to a particular circumstance category as "subgroup" (in Roemer (1998), a subgroup is referred to as "type").

In the LSSM data, health status is an ordinal variable which takes on values $h = 1, 2, 3, 4$. Most responses concentrate in categories 2 (fair) and 3 (good). Thus, for the stochastic dominance analysis, I group the lower two categories together (1 and 2) to define a new categorical variable which equals 1 if the respondent reports a poor or a fair health status, and equals 2 and 3 if the respondent reports a good and an excellent health status, respectively.

In the following subsections, I particularly focus on the following *childhood circumstances*: parental education and household socioeconomic status at age 10.

A.2.3.1 Parental Educational Attainment

To illustrate the application of the first-order stochastic dominance test in the context of the LSSM data, I define three subgroups based on maternal educational attainment: 1. Individuals whose mothers have incomplete primary school, 2. Mothers with complete primary school or incomplete secondary school, and 3. Mothers with complete secondary school or higher. Recall that higher values of the self-assessed health status denote a

better health status reported. I also define three subgroups based on paternal educational attainment, following the same definitions given for maternal educational attainment.

I examine the ranking of the conditional distributions of self-assessed health status using the non-parametric test proposed by Yalonetzky (2013). Appendix Table A.3 displays the test results for the comparison of health status across different maternal education levels for all individuals in the sample. Comparing the distributions for the first two subgroups shown in panel a of Appendix Table A1, at the 5 percent significance level and with a value of $-z^*$ of -1.645, the test suggests that the distribution for complete primary or incomplete secondary first-order-stochastically dominates the distribution for incomplete primary or no education in the LSSM sample. Regarding the first and the third subgroups (see Appendix Table A.3, panel b), the distribution for complete secondary or more dominates the distribution for primary education or less given the unanimously negative values and the significance of the z-statistic. A similar conclusion is suggested regarding the relationship between complete secondary or more and complete primary or incomplete secondary given the results presented in panel c of Appendix Table A.3. These results suggest that there is inequality of opportunity in adult health when a mother attains more education relative to a mother who obtains no more than some primary education.

Regarding urban areas, I find that the health distribution for mothers having completed secondary school dominate the health distribution for mothers who did not complete primary education. No dominance relationship can be established between the distribution for complete primary and incomplete primary as the z-statistic is not statistically significant for the first row, when I analyze the health category poor or fair. In rural areas, I find that no dominance relationship, at the first order, can be derived for the distributions of health status by each subgroup of maternal educational attainment (see Appendix Table A.4)

The statistical test results for stochastic dominance using the subgroups defined by

father's education level (see Appendix Table A.5) suggest that each of the distributions for complete primary and complete secondary dominates the distribution for incomplete primary at the first order. From these results, the dominance relationship between the distributions for complete primary and complete secondary is not clear. A similar result is obtained for the sample of urban residents, whereas no dominance relationship can be determined for rural residents (see Appendix Table A.6).

A.2.3.2   Household Socioeconomic Status in Childhood

I define five subgroups using the quintile groups of the socioeconomic status index calculated using information on ownership of assets by the individual's household at age 10. The non-parametric test results shown in Appendix Table A.7 suggest that the health distribution for the fifth quintile group dominates the distribution for all but the first quintile group, and that the fourth quintile group dominates the distribution for the first and second socioeconomic status quintile groups.

Turning to the urban subsample (see Appendix Table A.8), I find that the health distribution for the fifth quintile dominates each of the distributions for the four remaining quintile groups. These dominance relationships are statistically significant at the 5 percent level. In contrast with the urban sample, the statistical tests results for rural areas suggest that the only statistically significant dominance relationship is that of the health distribution for quintile 5 relative to the first and second quintile groups (see Appendix Table A.9).

The stochastic dominance analysis is limited in the sense that we cannot observe how different circumstances are related to each other. I can only focus on one circumstance at a time, and any potential conclusions derived from this analysis alone can be misleading. The regression approach is potentially more useful and allows to control for how different circumstances interact with each other.

Table A.3: Distribution of Health Status by Mothers Education Level: Full Sample

All individuals

a. Complete Primary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete primary school or incomplete secondary school			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	379	28.17	28.17	76	16.99	16.99	-0.112	-5.179 ***
2 = Good	884	65.72	93.89	330	73.84	90.83	-0.031	-2.022 ***
3 = Excellent	82	6.11	100	41	9.17	100		
Total	1,345	100		447	100			

b. Complete Secondary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	379	28.17	28.17	14	8.02	8.02	-0.202	-8.354 ***
2 = Good	884	65.72	93.89	113	65.79	73.81	-0.201	-5.863 ***
3 = Excellent	82	6.11	100	45	26.2	100.01		
Total	1,345	100		171	100			

c. Complete Secondary vs. Complete Primary

Health Status	Complete primary school or incomplete secondary school			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	76	16.99	16.99	14	8.02	8.02	-0.090	-3.282 ***
2 = Good	330	73.84	90.83	113	65.79	73.81	-0.170	-4.690 ***
3 = Excellent	41	9.17	100	45	26.2	100.01		
Total	447	100		171	100			

Note: *** denote that the statistic is significant at the 5 percent significant level. Source: 2010 Colombian LSSM Survey

Note for all tables in this section: The null hypothesis is given by $H_0 = \Delta F(k) \geq 0$ for some $k \in \{1,2\}$ and the alternative is given by $H_1 = \Delta F(k) < 0$ for all $k \in \{1,2\}$. $\widehat{\Delta F}(k)$ indicates the estimated difference between the cumulative probability functions, $\widehat{F_B(k)} - \widehat{F_A(k)}$, where $\widehat{F_B(k)}$ indicates the cumulative probability function for the subgroup in the most-right panel and $\widehat{F_A(k)}$ for the most-left panel, for row k. H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all k $\in \{1,2\}$, where $-z^* = -1.645$ is the left-tail critical value at the 5% significance level. No ordering can be established if the two values for z_k^l do not have the same direction.

Table A.4: Distribution of Health Status by Mothers Education Level: Urban and Rural Subsamples

	Residents in Urban Areas							
	a. Complete Primary vs. Incomplete Primary							
Health Status	Incomplete primary school or none			Complete primary school or incomplete secondary school			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	174	26.91	26.91	53	15.96	15.96	-0.110	-4.119 ***
2 = Good	426	65.88	92.79	248	74.54	90.5	-0.023	-1.204
3 = Excellent	47	7.22	100.01	32	9.5	100		
Total	647	100.01		333	100			

	b. Complete Secondary vs. Incomplete Primary							
Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	174	26.91	26.91	11	7.05	7.05	-0.199	-7.311 ***
2 = Good	426	65.88	92.79	99	65.65	72.7	-0.201	-5.336 ***
3 = Excellent	47	7.22	100.01	41	27.3	100		
Total	647	100.01		151	100			

	c. Complete Secondary vs. Complete Primary							
Health Status	Complete primary school or incomplete secondary school			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	53	15.96	15.96	11	7.05	7.05	-0.089	-3.080 ***
2 = Good	248	74.54	90.5	99	65.65	72.7	-0.178	-4.489 ***
3 = Excellent	32	9.5	100	41	27.3	100		
Total	333	100		151	100			

	Residents in Rural Areas							
	a. Complete Primary vs. Incomplete Primary							
Health Status	Incomplete primary school or none			Complete primary school or incomplete secondary school			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	220	31.57	31.57	30	25.88	25.88	-0.057	-1.275
2 = Good	456	65.31	96.88	77	67.85	93.73	-0.032	-1.333
3 = Excellent	22	3.13	100.01	7	6.27	100		
Total	698	100.01		114	100			

	b. Complete Secondary vs. Incomplete Primary							
Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	220	31.57	31.57	6	28.61	28.61	-0.030	-0.289
2 = Good	456	65.31	96.88	14	68.63	97.24	0.004	0.097
3 = Excellent	22	3.13	100.01	1	2.76	100		
Total	698	100.01		20	100			

	c. Complete Secondary vs. Complete Primary							
Health Status	Complete primary school or incomplete secondary school			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	30	25.88	25.88	6	28.61	28.61	0.027	0.250
2 = Good	77	67.85	93.73	14	68.63	97.24	0.035	0.814
3 = Excellent	7	6.27	100	1	2.76	100		
Total	114	100		20	100			

Note: *** denote that the statistic is significant at the 5% significant level. Source: 2010 Colombian LSSM Survey

Table A.5: Distribution of Health Status by Fathers Education Level: Full Sample

All individuals									
a. Complete Primary vs. Incomplete Primary									
Health Status	Incomplete primary school or none			Complete primary school or incomplete secondary school			$\widehat{\Delta F}(k)$	z_k^l	
	Freq.	%	Cumul. %	Freq.	%	Cumul. %			
1 = Poor/Fair	359	28.54	28.54	58	15.32	15.32	-0.132	-5.876	***
2 = Good	829	65.87	94.41	281	74.52	89.84	-0.046	-2.711	***
3 = Excellent	70	5.59	100	38	10.16	100			
Total	1,258	100		377	100				
b. Complete Secondary vs. Incomplete Primary									
Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l	
	Freq.	%	Cumul. %	Freq.	%	Cumul. %			
1 = Poor/Fair	359	28.54	28.54	22	11.23	11.23	-0.173	-6.658	***
2 = Good	829	65.87	94.41	123	63.23	74.46	-0.200	-6.240	***
3 = Excellent	70	5.59	100	50	25.54	100			
Total	1,258	100		194	100				
c. Complete Secondary vs. Complete Primary									
Health Status	Complete primary school or incomplete secondary school			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l	
	Freq.	%	Cumul. %	Freq.	%	Cumul. %			
1 = Poor/Fair	58	15.32	15.32	22	11.23	11.23	-0.041	-1.396	
2 = Good	281	74.52	89.84	123	63.23	74.46	-0.154	-4.399	***
3 = Excellent	38	10.16	100	50	25.54	100			
Total	377	100		194	100				

Note: *** denote that the statistic is significant at the 5% significant level. Source: 2010 Colombian LSSM Survey

Table A.6: Distribution of Health Status by Fathers Education Level: Urban and Rural Subsamples

Residents in Urban Areas

a. Complete Primary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete primary school or incomplete secondary school			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	160	27.39	27.39	42	14.45	14.45	-0.129	-4.670 ***
2 = Good	387	66.11	93.5	216	74.88	89.33	-0.042	-2.002 ***
3 = Excellent	38	6.5	100	31	10.67	100		
Total	585	100		289	100			

b. Complete Secondary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	160	27.39	27.39	20	11.14	11.14	-0.163	-5.419 ***
2 = Good	387	66.11	93.5	111	62.65	73.79	-0.197	-5.698 ***
3 = Excellent	38	6.5	100	46	26.21	100		
Total	585	100		177	100			

c. Complete Secondary vs. Complete Primary

Health Status	Complete primary school or incomplete secondary school			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	42	14.45	14.45	20	11.14	11.14	-0.033	-1.054
2 = Good	216	74.88	89.33	111	62.65	73.79	-0.155	-4.120 ***
3 = Excellent	31	10.67	100	46	26.21	100		
Total	289	100		177	100			

Residents in Rural Areas

a. Complete Primary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete primary school or incomplete secondary school			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	212	31.44	31.44	21	23.78	23.78	-0.077	-1.570
2 = Good	439	65.26	96.7	62	71.02	94.8	-0.019	-0.771
3 = Excellent	22	3.3	100	5	5.21	100.01		
Total	673	100		88	100.01			

b. Complete Secondary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	212	31.44	31.44	2	13.57	13.57	-0.179	-2.103 ***
2 = Good	439	65.26	96.7	14	79.46	93.03	-0.037	-0.591
3 = Excellent	22	3.3	100	1	6.97	100		
Total	673	100		17	100			

c. Complete Secondary vs. Complete Primary

Health Status	Complete primary school or incomplete secondary school			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	21	23.78	23.78	2	13.57	13.57	-0.102	-1.079
2 = Good	62	71.02	94.8	14	79.46	93.03	-0.018	-0.268
3 = Excellent	5	5.21	100.01	1	6.97	100		
Total	88	100.01		17	100			

Note: *** denote that the statistic is significant at the 5% significant level. Source: 2010 Colombian LSSM Survey.

Table A.7: Distribution of Health Status by Household Socioeconomic Status in Childhood

	a. Quintile 5 vs. Quintile 1								b. Quintile 5 vs. Quintile 2							
Health Status	Quintile 1			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l	Quintile 2			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %			Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	207	36.5	36.5	37	11.8	11.8	-0.246	-0.246	162	30.3	30.3	37	11.8	11.8	-0.185	-6.873 ***
2 = Good	342	60.1	96.5	208	66.0	77.8	-0.188	-0.188	347	65.2	95.5	208	66.0	77.8	-0.178	-7.085 ***
3 = Excellent	20	3.5	100.0	70	22.3	100.0			24	4.5	100.0	70	22.3	100.0		
Total	569	100		316	100				533	100		316	100			

	c. Quintile 5 vs. Quintile 3								d. Quintile 5 vs. Quintile 4							
Health Status	Quintile 3			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l	Quintile 4			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %			Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	94	21.4	21.4	37	11.8	11.8	-0.096	-3.600 ***	61	17.2	17.2	37	11.8	11.8	-0.054	-1.997 ***
2 = Good	313	71.0	92.4	208	66.0	77.8	-0.146	-5.508 ***	261	73.5	90.7	208	66.0	77.8	-0.129	-4.602 ***
3 = Excellent	34	7.6	100.0	70	22.3	100.0			33	9.3	100.0	70	22.3	100.0		
Total	441	100		316	100				355	100		316	100			

	e. Quintile 4 vs. Quintile 1								f. Quintile 4 vs. Quintile 2							
Health Status	Quintile 1			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l	Quintile 2			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %			Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	207	36.5	36.5	61	17.21	17.21	-0.192	-6.767 ***	162	30.3	30.3	61	17.2	17.2	-0.131	-4.645 ***
2 = Good	342	60.1	96.5	261	73.45	90.66	-0.059	-3.411 ***	347	65.2	95.5	261	73.5	90.7	-0.049	-2.716 ***
3 = Excellent	20	3.5	100.0	33	9.34	100			24	4.5	100.0	33	9.3	100.0		
Total	569	100		355	100				533	100		355	100			

	g. Quintile 4 vs. Quintile 3								h. Quintile 3 vs. Quintile 1							
Health Status	Quintile 3			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l	Quintile 1			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %			Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	94	21.4	21.4	76	17.2	17.2	-0.042	-1.501	207	36.5	36.5	94	21.4	21.4	-0.150	-0.238
2 = Good	313	71.0	92.4	252	73.5	90.7	-0.017	-0.872	342	60.1	96.5	313	71.0	92.4	-0.041	-0.129
3 = Excellent	34	7.6	100.0	27	9.3	100.0			20	3.5	100.0	34	7.6	100.0		
Total	441	100		355	100				569	100		441	100			

	i. Quintile 3 vs. Quintile 2								j. Quintile 2 vs. Quintile 1							
Health Status	Quintile 2			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l	Quintile 1			Quntile 2			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %			Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	162	30.3	30.3	94	21.4	21.4	-0.089	-3.198 ***	207	36.5	36.5	162	30.3	30.3	-0.061	-2.159 ***
2 = Good	347	65.2	95.5	313	71.0	92.4	-0.031	-2.009 ***	342	60.1	96.5	347	65.2	95.5	-0.010	-0.873
3 = Excellent	24	4.5	100.0	34	7.6	100.0			20	3.5	100.0	24	4.5	100.0		
Total	533	100		441	100				569	100		533	100			

Note: *** denote that the statistic is significant at the 5% significant level. Source: 2010 Colombian LSSM Survey.

Table A.8: Distribution of Health Status by Household Socioeconomic Status in Childhood: Residents in Urban Areas

a. Quintile 5 vs. Quintile 1

Health Status	Quintile 1			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	86	32.5	32.5	20	8.5	8.5	-0.239	-7.041 ***
2 = Good	164	62.0	94.5	157	66.2	74.8	-0.197	-6.256 ***
3 = Excellent	15	5.5	100.0	60	25.3	100.0		
Total	265	100.0		237	100.0			

b. Quintile 5 vs. Quintile 2

Health Status	Quintile 2			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	58	23.2	23.2	20	8.5	8.5	-0.147	-4.558 ***
2 = Good	178	70.4	93.6	157	66.2	74.8	-0.189	-5.874 ***
3 = Excellent	16	6.4	100.0	60	25.3	100.0		
Total	252	100.0		237	100.0			

c. Quintile 5 vs. Quintile 3

Health Status	Quintile 3			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	55	21.7	21.7	20	8.5	8.5	-0.132	-4.175 ***
2 = Good	178	70.3	92.0	157	66.2	74.8	-0.173	-5.243 ***
3 = Excellent	20	8.0	100.0	60	25.3	100.0		
Total	253	100.0		237	100.0			

d. Quintile 5 vs. Quintile 4

Health Status	Quintile 4			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	49	20.1	20.1	20	8.5	8.5	-0.116	-3.692 ***
2 = Good	171	70.5	90.6	157	66.2	74.8	-0.159	-4.688 ***
3 = Excellent	23	9.4	100.0	60	25.3	100.0		
Total	243	100.0		237	100.0			

e. Quintile 4 vs. Quintile 1

Health Status	Quintile 1			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	86	32.5	32.5	49	20.14	20.14	-0.123	-3.193 ***
2 = Good	164	62.0	94.5	171	70.48	90.62	-0.038	-1.646 ***
3 = Excellent	15	5.5	100.0	23	9.38	100		
Total	265	100.0		243	100			

f. Quintile 4 vs. Quintile 2

Health Status	Quintile 2			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	58	23.2	23.2	49	20.1	20.1	-0.031	-0.824
2 = Good	178	70.4	93.6	171	70.5	90.6	-0.030	-1.243
3 = Excellent	16	6.4	100.0	23	9.4	100.0		
Total	252	100.0		243	100.0			

g. Quintile 4 vs. Quintile 3

Health Status	Quintile 3			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	55	21.7	21.7	53	20.1	20.1	-0.016	-0.435
2 = Good	178	70.3	92.0	171	70.5	90.6	-0.014	-0.557
3 = Excellent	20	8.0	100.0	19	9.4	100.0		
Total	253	100.0		243	100.0			

h. Quintile 3 vs. Quintile 1

Health Status	Quintile 1			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	86	32.5	32.5	55	21.7	21.7	-0.107	-0.172
2 = Good	164	62.0	94.5	178	70.3	92.0	-0.024	-0.069
3 = Excellent	15	5.5	100.0	20	8.0	100.0		
Total	265	100.0		253	100.0			

i. Quintile 3 vs. Quintile 2

Health Status	Quintile 2			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	58	23.2	23.2	55	21.7	21.7	-0.015	-0.393
2 = Good	178	70.4	93.6	178	70.3	92.0	-0.016	-0.697
3 = Excellent	16	6.4	100.0	20	8.0	100.0		
Total	252	100.0		253	100.0			

j. Quintile 2 vs. Quintile 1

Health Status	Quintile 1			Quntile 2			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	86	32.5	32.5	58	23.2	23.2	-0.093	-2.367 ***
2 = Good	164	62.0	94.5	178	70.4	93.6	-0.008	-0.403
3 = Excellent	15	5.5	100.0	16	6.4	100.0		
Total	265	100.0		252	100.0			

Note: *** denote that the statistic is significant at the 5% significant level. Source: 2010 Colombian LSSM Survey.

Table A.9: Distribution of Health Status by Household Socioeconomic Status in Childhood: Residents in Rural Areas

a. Quintile 5 vs. Quintile 1

Health Status	Quintile 1			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	111	45.1	45.1	36	19.5	19.5	-0.256	-5.947 ***
2 = Good	128	51.9	97.0	136	73.5	92.9	-0.040	-1.851 ***
3 = Excellent	7	3.0	100.0	13	7.1	100.0		
Total	246	100.0		185	100.0			

b. Quintile 5 vs. Quintile 2

Health Status	Quintile 2			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	59	37.4	37.4	36	19.5	19.5	-0.180	-3.726 ***
2 = Good	95	60.3	97.7	136	73.5	92.9	-0.048	-2.169 ***
3 = Excellent	4	2.3	100.0	13	7.1	100.0		
Total	158	100.0		185	100.0			

c. Quintile 5 vs. Quintile 3

Health Status	Quintile 3			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	56	30.7	30.7	36	19.5	19.5	-0.113	-2.510 ***
2 = Good	118	65.2	95.9	136	73.5	92.9	-0.030	-1.253
3 = Excellent	7	4.1	100.0	13	7.1	100.0		
Total	181	100.0		185	100.0			

d. Quintile 5 vs. Quintile 4

Health Status	Quintile 4			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	47	24.1	24.1	36	19.5	19.5	-0.047	-1.104
2 = Good	141	72.7	96.8	136	73.5	92.9	-0.039	-1.729 ***
3 = Excellent	6	3.2	100.0	13	7.1	100.0		
Total	194	100.0		185	100.0			

e. Quintile 4 vs. Quintile 1

Health Status	Quintile 1			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	111	45.1	45.1	47	24.12	24.12	-0.209	-4.740 ***
2 = Good	128	51.9	97.0	141	72.71	96.83	-0.001	-0.072
3 = Excellent	7	3.0	100.0	6	3.16	99.99		
Total	246	100.0		194	99.99			

f. Quintile 4 vs. Quintile 2

Health Status	Quintile 2			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	59	37.4	37.4	47	24.1	24.1	-0.133	-2.702 ***
2 = Good	95	60.3	97.7	141	72.7	96.8	-0.009	-0.527
3 = Excellent	4	2.3	100.0	6	3.2	100.0		
Total	158	100.0		194	100.0			

g. Quintile 4 vs. Quintile 3

Health Status	Quintile 3			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	56	30.7	30.7	60	24.1	24.1	-0.066	-1.438
2 = Good	118	65.2	95.9	126	72.7	96.8	0.009	0.475
3 = Excellent	7	4.1	100.0	8	3.2	100.0		
Total	181	100.0		194	100.0			

h. Quintile 3 vs. Quintile 1

Health Status	Quintile 1			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	111	45.1	45.1	56	30.7	30.7	-0.143	-0.211
2 = Good	128	51.9	97.0	118	65.2	95.9	-0.010	-0.040
3 = Excellent	7	3.0	100.0	7	4.1	100.0		
Total	246	100.0		181	100.0			

i. Quintile 3 vs. Quintile 2

Health Status	Quintile 2			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	59	37.4	37.4	56	30.7	30.7	-0.067	-1.297
2 = Good	95	60.3	97.7	118	65.2	95.9	-0.018	-0.969
3 = Excellent	4	2.3	100.0	7	4.1	100.0		
Total	158	100.0		181	100.0			

j. Quintile 2 vs. Quintile 1

Health Status	Quintile 1			Quntile 2			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	111	45.1	45.1	59	37.4	37.4	-0.076	-1.527
2 = Good	128	51.9	97.0	95	60.3	97.7	0.008	0.490
3 = Excellent	7	3.0	100.0	4	2.3	100.0		
Total	246	100.0		158	100.0			

Note: *** denote that the statistic is significant at the 5% significant level. Source: 2010 Colombian LSSM Survey.

A.3   Accounting for Health Conditions and Retrospective Recall

As a first additional estimation, I include variables for self-reported chronic illness and self-reported disability as control variables (results are presented in Tables A.10 and A.11). Self-reported chronic illness is a dichotomous variable that indicates whether the individual suffers from a chronic or long-standing illness like diabetes, heart disease or cancer. Self-reported disability is a dichotomous variable that indicates the presence of a permanent disability.

These objective measures of health status have a negative and significant effect on the likelihood of reporting a good health status. This result is consistent across the full sample and the subsamples of urban and rural areas. Following the results in table A.10, the associations between circumstances and adult health status previously described do not change after including these health variables in the estimations. The equation for years of education (results available upon request) does not include the objective health measures. Thus, by construction, the coefficients and standard errors for chronic illness and permanent disability are the same in both the estimation of the non-linear model for health status including years of education and the estimation including years of education purged from the effect of circumstances. These objective measures of health status, however, highly depend on the respondents access to health care services. The distribution of health services in the country is not necessarily random. For instance, the differential health care use between urban and rural areas may reflect both a major difficulty in securing the availability of health care providers in rural areas and a large concentration of private health care providers in urban areas (Vargas-Lorenzo, 2009). Chronic illness and permanent disability are not perfect indicators of health status on their own either. Individuals may experience psychological adjustment and adaptation to permanent health problems that, in turn, affect how they perceive and report their health status (Graham, 2008).

Table A.11 shows the estimation of the inequality of opportunity indexes. The Gini-opportunity index is below the index presented in the main document. The index now ranges between 0.042 and 0.077, with rural areas exhibiting the lowest estimate, as in the main results. Note here that the outcome of interest is the health status variable with four categories. The dissimilarity indexes, on the other hand, are now larger than the indexes reported in the main document.

Regarding the decomposition of the dissimilarity index, it can be observed that all circumstances but own education, have a contribution of between 36% and 50%, with socioeconomic status at age 10 and region of birth being the most important early life circumstances. In urban areas, besides the aforementioned variables, paternal education is perhaps the most important factor in inequality of opportunity, whereas in rural areas, socioeconomic status at age 10 stands out as the most influential variable. Overall, it can be argued that the results are robust to the inclusion of objective measures of health status.

The use of self-reported and retrospective recall data could bias the results here obtained. In order to gauge if there is a systematic bias in how health status is reported, I examine how people perceive their health status based on their economic conditions, after controlling for the set of circumstances and the presence of chronic illness and permanent disability. Self-reported health status and household income per capita (defined in both levels and logs) are strongly correlated, but once I control for circumstances and objective measures of health status this correlation attenuates at conventional significance levels. Thus, the bias created by self-reported measures should be reduced as long as more objective measures are included in the model.

To check for one conceivable source of bias induced by retrospective recall, I analyze whether the age of an individual affects their recall of birth circumstances in a certain direction. In particular, I estimate the logistic regression models for three age cohorts: 2535, 3650, and 5165 years old. The results suggest that self-reported health suffers from

reporting bias in view of the substantial differences by age group. Reporting bias constitutes a threat to the analysis in this study as it compromises the comparisons between individuals with different socioeconomic characteristics.

The estimation results from the logit models for each age group are shown in Table A.12. Being a male is positively associated with reporting a good health for all age-groups. Note for the 2535 age-group that having a mother who completed primary but not secondary education has a negative association with good health status. In contrast, the opposite is true for the 5165 age-group. Higher quintile groups of household socioeconomic status at age 10 are only statistically significant and positively associated with a good self-assessment of health for individuals between 36 and 50 years of age.

Table A.13 shows the estimation of the inequality of opportunity indexes. The Gini-opportunity index ranges between 0.03 and 0.10, with the 5065 age-group exhibiting the highest coefficient estimate. Note here that the outcome of interest is also the health status variable with four categories. The dissimilarity indexes range between 0.04 and 0.10, with the highest value in the 5065 group.

Regarding the decomposition of the dissimilarity index, all circumstances but own education, have a contribution of between 59% and 78%. The contribution of each circumstance varies by age cohort. For instance, maternal education seems to be more important for the 5065 group than for the 3550 group, for which socioeconomic status at age 10 is the most prominent circumstance in inequality of opportunity. Region of birth and ethnicity are more important for the 2535 age group than for any other

Table A.10: Log-odds Ratios, controlling for presence of chronic illness or permanent disabilities

Dependent variable: self-reported health status (0=poor or fair, 1= good or excellent)	All Individuals		Urban Areas		Rural Areas	
	(1)	(2)	(3)	(4)	(5)	(6)
Any chronic illness (1=Yes)	−1.9755***	−1.9755***	−2.0409***	−2.0409***	−1.7436***	−1.7436***
	(0.1761)	(0.1761)	(0.2068)	(0.2068)	(0.2678)	(0.2678)
Any permanent disability (1=Yes)	−1.4031***	−1.4031***	−1.5184***	−1.5184***	−1.2382**	−1.2382**
	(0.3701)	(0.3701)	(0.5360)	(0.5360)	(0.5053)	(0.5053)
Male	0.4603***	0.4863***	0.5621***	0.6401***	0.4354**	0.3888*
	(0.1373)	(0.1375)	(0.1685)	(0.1686)	(0.2188)	(0.2171)
Age group (Ref. 25–35 years old):						
35–45 years old	−0.5017**	−0.5144**	−0.4825*	−0.4772	−0.5309**	−0.6043**
	(0.2094)	(0.2093)	(0.2914)	(0.2914)	(0.2540)	(0.2535)
45–55 years old	−0.4342**	−0.4619**	−0.3978	−0.4357	−0.6642**	−0.7509***
	(0.2071)	(0.2072)	(0.2857)	(0.2859)	(0.2588)	(0.2600)
55–65 years old	−0.8310***	−0.9638***	−0.8056***	−0.9556***	−1.0912***	−1.2735***
	(0.2108)	(0.2108)	(0.2921)	(0.2924)	(0.2767)	(0.2772)
Ethnicity (Ref. Not a minority):						
Indigenous	−0.1588	−0.1663	−0.4919	−0.4956	0.4388	0.4555
	(0.3975)	(0.3975)	(0.5402)	(0.5402)	(0.4583)	(0.4583)
Black/mulato/raizal/palenquero	−0.1288	−0.0927	−0.2521	−0.2399	−0.0422	−0.0178
	(0.2604)	(0.2606)	(0.3303)	(0.3304)	(0.3843)	(0.3837)
Region (Ref. Atlantic and San Andres islands):						
Eastern	−0.1640	−0.1639	−0.1203	−0.1590	−0.4932*	−0.5061**
	(0.1922)	(0.1922)	(0.2514)	(0.2512)	(0.2546)	(0.2552)
Pacific	−0.5767**	−0.5188**	−0.6139*	−0.4840	−0.7038**	−0.6838**
	(0.2287)	(0.2277)	(0.3561)	(0.3534)	(0.2801)	(0.2790)
Orinoquia and Amazonia	0.2593	0.1880	0.7666	0.6341	−0.7848	−0.8086
	(0.4692)	(0.4690)	(0.6800)	(0.6798)	(0.7630)	(0.7633)
Antioquia	0.1878	0.1788	0.3785	0.3486	−0.5712*	−0.5727*
	(0.2334)	(0.2334)	(0.3046)	(0.3047)	(0.3144)	(0.3145)
Valle	0.3126	0.3842	0.3487	0.4141	−0.1455	−0.1129
	(0.3235)	(0.3236)	(0.3891)	(0.3891)	(0.4842)	(0.4839)
Bogota	−0.5127*	−0.5496*	−0.4760	−0.5445*		
	(0.2826)	(0.2831)	(0.3103)	(0.3110)		
Central	−0.0846	−0.0201	0.0448	0.1358	−0.4093	−0.4044
	(0.2104)	(0.2102)	(0.2829)	(0.2821)	(0.2592)	(0.2591)
Born in urban area	−0.1281	−0.0052	−0.2701	−0.2159	0.1596	0.1944
	(0.1451)	(0.1434)	(0.1928)	(0.1924)	(0.2469)	(0.2465)
Household socioeconomic status at age 10:						
Quintile Group 2	0.0974	0.1821	0.0538	0.1812	0.1404	0.1063
	(0.1696)	(0.1682)	(0.2357)	(0.2323)	(0.2598)	(0.2598)
Quintile Group 3	0.4048**	0.6455***	0.0342	0.2918	0.8708***	0.9125***
	(0.1983)	(0.1955)	(0.2609)	(0.2567)	(0.2621)	(0.2625)
Quintile Group 4	0.2750	0.6830***	−0.0029	0.3738	0.7075***	0.7801***
	(0.2261)	(0.2180)	(0.2850)	(0.2712)	(0.2688)	(0.2674)
Quintile Group 5	0.8770***	1.4271***	0.7342*	1.2518***	0.9375***	1.1488***
	(0.3115)	(0.2959)	(0.3773)	(0.3572)	(0.3138)	(0.3063)
Paternal education level (Ref. None):						
Complete primary and incomplete secondary	0.3285	0.4428**	0.5145*	0.6086**	−0.2968	−0.1746
	(0.2219)	(0.2219)	(0.2646)	(0.2657)	(0.3712)	(0.3693)
Complete secondary or more	−0.1788	0.0478	−0.0889	0.0990	0.0872	0.4411
	(0.3903)	(0.3908)	(0.4313)	(0.4318)	(0.7347)	(0.7313)
Unknown father's level of education	0.0902	0.0051	0.3461	0.2562	−0.3840	−0.4351*
	(0.2038)	(0.2033)	(0.2810)	(0.2802)	(0.2546)	(0.2528)

Table A.10: Log-odds Ratios, controlling for presence of chronic illness or permanent disabilities (continued)

Dependent variable: self-reported health status (0=poor or fair, 1= good or excellent)	All Individuals		Urban Areas		Rural Areas	
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal education level (Ref. None):						
Complete primary and incomplete secondary	−0.1319	−0.0079	−0.1069	0.0283	−0.3793	−0.3155
	(0.2109)	(0.2096)	(0.2582)	(0.2559)	(0.3129)	(0.3113)
Complete secondary or more	0.4682	0.7693*	0.6236	0.9324*	−0.7971	−0.5783
	(0.4583)	(0.4541)	(0.5500)	(0.5452)	(0.6692)	(0.6610)
Unknown mother's level of education	−0.1725	−0.2209	−0.2442	−0.2641	0.0170	−0.0030
	(0.2360)	(0.2360)	(0.3259)	(0.3259)	(0.2771)	(0.2772)
Years of education	0.1196***		0.1215***		0.0961***	
	(0.0182)		(0.0231)		(0.0278)	
Years of education purged from circumstances		0.1196***		0.1215***		0.0961***
		(0.0182)		(0.0231)		(0.0278)
Constant	0.7647***	1.3154***	0.8864**	1.5622***	0.8300**	1.3060***
	(0.2635)	(0.2576)	(0.3727)	(0.3609)	(0.3667)	(0.3511)
Observations	2,204	2,204	1,242	1,242	956	956
Region of Birth Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Log−likelihood	−4.044e+06	−4.044e+06	−2.964e+06	−2.964e+06	−1.018e+06	−1.018e+06
Pseudo R squared	0.211	0.211	0.230	0.230	0.168	0.168

Note: ***, **, and * indicate statistical significance at the 1, 5 and 10 percent level, respectively.
Robust standard errors in parentheses
Own calculations. Source: 2010 Colombian LSSM Survey.

Table A.11: Gini-Opportunity index and Dissimilarity Index of Inequality of Opportunity, with its Decomposition, controlling for presence of chronic illness or permanent disability

	All individuals		Residents in Urban Areas		Residents in Rural Areas	
Gini-Opportunity Index (1)	0.0777		0.0735		0.0429	
Dissimilarity Index (2)	0.1033	0.1034	0.0990	0.0999	0.1227	0.1226
	Decomposition of the Dissimilarity Index (in %)					
Educational Attainment	50.87		42.19		26.09	
Education purged from circumstances		36.12		41.41		19.83
Circumstances	49.13	63.88	57.81	58.59	73.91	80.17
Early Life Circumstances	45.00	31.27	38.80	38.76	53.70	58.94
Mother's Education	8.99	6.15	10.43	11.96	4.16	3.05
Father's Education	10.14	7.74	12.71	13.56	8.57	9.35
Household Socioeconomic Status at age 10	25.86	17.38	15.65	13.24	40.97	46.54
Demographics	18.89	17.85	19.01	19.82	20.21	21.22
Region of Birth	13.46	13.14	16.35	17.07	17.33	17.86
Born in Urban Area	4.32	3.64	0.56	0.90	1.82	2.26
Ethnicity	1.11	1.07	2.11	1.85	1.06	1.11
Observations	2,204		1,242		962	

Bootstrapped standard errors in parentheses. 100 replications.
Own calculations. Source: 2010 Colombian LSSM

Notes:
(1) The Gini-opportunity index is calculated using a self-assessed health status variable in which 1=poor, 2=fair, 3=good, and 4=excellent. A categorical variable for the individual's years of education has also been used in this calculation. Gender and age group are not included.
(2) The index in the first, third and fifth columns include years of education as a circumstance, whereas the second, fourth, and sixth columns include years of education purged from circumstances.

Table A.12: Log-odds Ratios for the Correlates of Self-Assessed Health Status by Age Group

Dependent variable: self-reported health status (0=poor or fair, 1= good or excellent)	Age group: 25−35		36−50		51−65	
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.7924**	0.7271**	0.5218***	0.5611***	0.5171***	0.5651***
	(0.3150)	(0.3162)	(0.1979)	(0.1990)	(0.2003)	(0.2018)
Ethnicity (Ref. Not a minority):						
Indigenous	−1.2907	−1.2854	0.1894	0.2864	0.0161	−0.1839
	(0.7920)	(0.7921)	(0.5639)	(0.5636)	(0.7473)	(0.7466)
Black/*mulato/raizal/palenquero*	−0.4458	−0.4976	−0.4391	−0.3624	0.1345	0.1635
	(0.4735)	(0.4720)	(0.3821)	(0.3827)	(0.4120)	(0.4118)
Region (Ref. Atlantic and San Andres islands):						
Eastern	−0.3581	−0.3333	−0.2892	−0.3258	−0.1536	−0.0746
	(0.5251)	(0.5248)	(0.2749)	(0.2757)	(0.2780)	(0.2771)
Pacific	−0.9042*	−0.8942*	−0.6281*	−0.5816*	−0.7038**	−0.6137*
	(0.4620)	(0.4616)	(0.3425)	(0.3406)	(0.3515)	(0.3490)
Orinoquia and Amazonia	0.0000	0.0000	0.2286	0.0964	−0.0296	−0.2552
	(0.0000)	(0.0000)	(0.6933)	(0.6922)	(0.9000)	(0.8989)
Antioquia	0.6988	0.7516	−0.0448	−0.0612	−0.0082	0.0004
	(0.6109)	(0.6142)	(0.3545)	(0.3551)	(0.3351)	(0.3349)
Valle	−0.5004	−0.4549	0.6391	0.6139	−0.1454	0.0859
	(0.7554)	(0.7549)	(0.5005)	(0.5008)	(0.4494)	(0.4488)
Bogota	−0.4951	−0.6110	−0.4700	−0.4831	−0.4874	−0.4970
	(0.6106)	(0.6154)	(0.4525)	(0.4526)	(0.4520)	(0.4521)
Central	0.0089	0.0556	−0.1130	−0.0816	−0.4387	−0.3035
	(0.5189)	(0.5184)	(0.3295)	(0.3291)	(0.3021)	(0.2990)
Born in urban area	0.1192	0.2132	−0.2122	−0.0309	0.0884	0.1610
	(0.4015)	(0.3989)	(0.2100)	(0.2057)	(0.2100)	(0.2093)
Household socioeconomic status at age 10:						
Quintile Group 2	0.9255*	0.9853*	0.2990	0.4309*	−0.2145	−0.1829
	(0.5268)	(0.5236)	(0.2433)	(0.2407)	(0.2479)	(0.2469)
Quintile Group 3	0.1625	0.5102	0.8799***	1.1013***	−0.0481	0.1371
	(0.4791)	(0.4726)	(0.2919)	(0.2902)	(0.2784)	(0.2701)
Quintile Group 4	−0.1975	0.2942	0.5725*	0.9566***	−0.0799	0.2757
	(0.5514)	(0.5258)	(0.3080)	(0.2996)	(0.3666)	(0.3514)
Quintile Group 5	0.4275	1.0926*	0.9503**	1.4916***	0.0380	0.5081
	(0.6903)	(0.6312)	(0.4699)	(0.4481)	(0.4653)	(0.4386)
Paternal education level (Ref. None):						
Complete primary and incomplete secondary	0.3920	0.5783	0.4352	0.4840	0.0960	0.2276
	(0.4682)	(0.4693)	(0.3598)	(0.3595)	(0.3887)	(0.3890)
Complete secondary or more	0.4664	0.8151	0.3590	0.4415	−0.6995	−0.3575
	(0.6931)	(0.6845)	(0.5559)	(0.5554)	(0.6188)	(0.6162)
Unknown father's level of education	−0.3563	−0.3242	0.3718	0.2627	0.2458	0.1583
	(0.4181)	(0.4205)	(0.3394)	(0.3375)	(0.2855)	(0.2850)

Table A.12: Log-odds Ratios for the Correlates of Self-Assessed Health Status by Age Group (continued)

Dependent variable: self-reported health status (0=poor or fair, 1= good or excellent)	Age group: 25−35		36−50		51−65	
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal education level (Ref. None):						
Complete primary and incomplete secondary	−0.9342**	−0.8795*	−0.1845	−0.0424	0.7081*	0.8547**
	(0.4580)	(0.4582)	(0.3050)	(0.3055)	(0.3920)	(0.3913)
Complete secondary or more	1.2847	1.5335	−0.2113	0.1268	1.0861	1.4160*
	(1.0338)	(1.0314)	(0.6177)	(0.6086)	(0.7485)	(0.7471)
Unknown mother's level of education	0.4241	0.4291	−0.5432	−0.7347*	0.1256	0.1816
	(0.5115)	(0.5113)	(0.3760)	(0.3779)	(0.3222)	(0.3224)
Years of education	0.1433***		0.1158***		0.1042***	
	(0.0461)		(0.0259)		(0.0264)	
Years of education purged from circumstances		0.1433***		0.1158***		0.1042***
		(0.0461)		(0.0259)		(0.0264)
Constant	0.4941	1.2478**	−0.0921	0.4363	−0.4141	−0.1043
	(0.5416)	(0.5181)	(0.2871)	(0.2748)	(0.2963)	(0.2885)
Observations	541	541	918	918	735	735
Region of Birth Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Log−likelihood	−716710	−716710	−1.823e+06	−1.823e+06	−1.816e+06	−1.816e+06
Pseudo R squared	0.151	0.151	0.113	0.113	0.0817	0.0817

Note: ***, **, and * indicate statistical significance at the 1, 5 and 10 percent level, respectively.
Robust standard errors in parentheses
Own calculations. Source: 2010 Colombian LSSM Survey.

Table A.13: Gini-Opportunity index and Dissimilarity Index of Inequality of Opportunity, with its Decomposition, by Age Group

	Age group: 25-35		35-50		50-65	
Gini-Opportunity Index (1)	0.0331		0.0920		0.1029	
Dissimilarity Index (2)	0.0473	0.0473	0.0720	0.0720	0.1018	0.1018
	Decomposition of the Dissimilarity Index (in %)					
Educational Attainment	21.97		22.88		28.38	
Education purged from circumstances		28.60		38.14		41.30
Circumstances	78.03	71.40	77.12	61.86	71.62	58.70
Early Life Circumstances	50.32	45.51	55.52	42.12	58.42	46.80
Mother's Education	20.47	19.43	9.53	6.86	26.24	21.08
Father's Education	8.78	6.85	9.62	7.61	13.73	12.50
Household Socioeconomic Status at age 10	21.07	19.23	36.37	27.64	18.44	13.23
Demographics	27.71	25.89	21.60	19.74	13.20	11.89
Region of Birth	19.32	18.53	14.88	13.99	6.59	6.71
Born in Urban Area	0.80	0.31	5.54	4.42	6.04	4.76
Ethnicity	7.59	7.04	1.18	1.34	0.58	0.42
Observations	541		918		735	

Bootstrapped standard errors in parentheses. 100 replications.
Own calculations. Source: 2010 Colombian LSSM

Notes:

(1) The Gini-opportunity index is calculated using a self-assessed health status variable in which 1=poor, 2=fair, 3=good, and 4=excellent. A categorical variable for the individual's years of education has also been used in this calculation. Gender and age group are not included.

(2) The index in the first, third and fifth columns include years of education as a circumstance, whereas the second, fourth, and sixth columns include years of education purged from circumstances.