



Education and the Informal Sector: Evidence from Venezuela and Brazil

Susana Carpio

► To cite this version:

Susana Carpio. Education and the Informal Sector: Evidence from Venezuela and Brazil. Economics and Finance. Université d'Auvergne - Clermont-Ferrand I, 2014. English. NNT : 2014CLF10443 . tel-01168572

HAL Id: tel-01168572

<https://theses.hal.science/tel-01168572>

Submitted on 26 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université d'Auvergne, Clermont-Ferrand 1
Faculté des Sciences Économiques et de Gestion
École Doctorale des Sciences Économiques, Juridiques et de Gestion
Centre d'Études et de Recherches sur le Développement International (CERDI)

Education and the Informal Sector

Evidence from Venezuela and Brazil

Pour l'obtention du titre de Docteur en Sciences Économiques

Par

Susana Carpio

Sous la direction de

M. le Professeur Jean-Louis ARCAND

Membres du jury

Martina VIARENGO (Professeur à l'Institut de Hautes Études Internationales et du Développement, Genève)	Rapporteur
Pablo ACOSTA (Économiste Senior à la Banque Mondiale, Washinton DC)	Rapporteur
Martine AUDIBERT (Directeur de Recherche, CNRS-CERDI)	Suffragant
Jean-Louis ARCAND (Professeur à l'Institut de Hautes Études Internationales et du Développement, Genève)	Directeur de thèse

Acknowledgements

I would like to express my very great appreciation to my jury, Martina Viarengo, Martine Audibert and Pablo Acosta, for their willingness to give their time and for their valuable and constructive suggestions. My special thanks are extended to my advisor, first for giving me the liberty of choosing the subjects that I was interested in, and second, for his patient guidance and his valuable and always detailed suggestions. Thanks to him I was submerged into the deep econometric world, in which I must say that I enjoyed running Monte Carlo logarithms for days. I would also like to extend my thanks to all the staff at my research department (CERDI) for the enriching academical experience in the middle of the beautiful chain of volcanoes of Puy de Dôme. I am particularly grateful to my colleague Catherine Korachais for giving me support with the administrative requirements, and moreover for her generous hospitality. I wish to acknowledge the help provided by the staff at the Venezuelan Ministry of Planning, by means of which I could have access to the data for my research. I would also like to thank my ex-colleagues from the Labor Market Unit at the Inter-American Development Bank for their useful critiques and encouragement.

I am particularly grateful for my dearest friends, Gaby, Jenny and Eli, who although never finished reading my thesis they gave me moral support and encouraged me to achieve my goals. To my family, my grandmothers who taught me how to cook and survive with delicious and soulful food during bad times. My cousin Ailed who accompanied me throughout all my dissertation experience in spite of physical distance. My brothers and sister for believing in me and make me try harder every time to be an example for them. To my kind mother, Isaura, for always fulfilling me with lots of love and for believing in my capacity to accomplish my dreams. To my grumpy yet loving father, Oswaldo, for tirelessly pressuring me into finishing my thesis. I would not have pursued this goal without his encouragement. To my beloved husband, Alejandro, most of all for his incredible patience all these years. I wish to thank him for his iron fist treatment towards me, he has given me both the strength and the love needed to finish my thesis. And, to the most important person in my life and my inspiration, my son Simón, thanks for being such an adorable boy and letting me have some space to work in my dissertation.

I dedicate my thesis to my father and husband.

Résumé

Le principal objectif de ce travail de recherche est d'analyser de façon empirique, à l'aide d'une approche nouvelle et robuste, les déterminants de la fréquentation scolaire, le rôle du secteur informel dans l'abandon scolaire, et les liens entre le travail des jeunes et la productivité dans les petites entreprises. Premièrement, nous sommes intéressés à déterminer la magnitude des rendements privés de l'éducation au Venezuela. Pour cela, nous utilisons des données en coupe transversale répétée provenant de l'enquête des ménages. Le désavantage d'utiliser des données de pseudo-panel est la présence élevée d'erreurs de mesure. Nous abordons ce problème à travers l'utilisation de la méthode de variables instrumentales basée sur les moments empiriques d'ordre supérieur à deux. Par la suite, nous cherchons à comprendre les causes de l'abandon scolaire dans l'enseignement secondaire au Venezuela. Notre contribution à la littérature de l'abandon scolaire est double. Tout d'abord, l'inclusion de l'informalité (représentée par les parents des jeunes qui travaillent au noir) comme l'un des principaux facteurs qui influent sur la probabilité que les enfants culminent leurs études secondaires. Deuxièmement, l'utilisation de données de panel dans l'évaluation du problème des jeunes décrocheurs du secondaire, donne un aperçu nouveau et plus robuste sur les goulets d'étranglement de l'éducation du Venezuela. L'approche économétrique est basée sur l'utilisation des corrections de Mundlak afin d'éliminer le biais d'endogénéité. Pour terminer, nous analysons l'effet de jeunes travailleurs sur la productivité des petites entreprises non enregistrées au Brésil. Nous utilisons la méthode de la variable instrumentale afin de veiller à ce que nos résultats ne soient pas biaisés.

Mots clés: rendements de l'éducation, l'abandon scolaire, travail des jeunes, économie informelle, Amérique Latine, pseudo-panel, variables instrumentales

Summary

The purpose of this thesis is to empirically assess, through the use of novel and econometrically robust approaches, the drivers of school attendance, the role of informality in school dropout, and the linkages between youth work and productivity in small firms. First, we are interested in determining how important the private returns to education are in Venezuela. To this matter, we construct a pseudo-panel data by means of the repeated cross-sections from the household survey. The drawback of using pseudo-panel data is the high presence of measurement errors. We address this issue through the use of consistent instrumental variables estimators based on sample moments of order higher than two. Second, we seek to understand the causes of school dropout in secondary education in Venezuela. The econometric approach consists in eliminating the endogeneity bias by using Mundlak corrections, since there is neither a good instrument nor can fixed-effects estimators be used in this estimation. Our contribution for school dropout's literature is twofold. First, the inclusion of informality (informal worker parents) as one of the main factors affecting the likelihood of children completing high school. Second, the use of recent panel data in assessing the problem of early high school dropouts, provides new and more robust insight into Venezuela's educational bottlenecks. Finally, we analyze the effect of the share of young workers on small unregistered firms' productivity in Brazil. We use the method of Instrumental Variable in order to ensure that our results are not biased.

Keywords: returns to education, school dropouts, youth work, informal sector, Latin America, pseudo-panel, instrumental variables.

Contents

Introduction	x
1 The Venezuelan Context: Economy and Education System	1
1.1 Education System	2
1.1.1 Access to School and Government Expenditure	4
1.1.2 Quality of Education	6
1.1.3 Reforms and Legal Framework	7
1.2 Tables and Figures	12
2 Errors in Variables in Pseudo-Panel Data: A Simple Correction	18
2.1 Introduction	18
2.2 Estimation strategy	19
2.2.1 Basic notation	19
2.2.2 An instrumental variables approach	20
2.3 Montecarlo evidence	21
2.4 An application to a large Venezuelan pseudo-panel	23
2.4.1 Data	23
2.4.2 Specification and existing estimates	24
2.4.3 Results	25
2.5 Concluding remarks	26
2.6 Tables and Figures	30
3 High School Dropouts and the Informal Sector	37
3.1 Introduction	37
3.2 Theoretical Model Review	39
3.3 Empirical Approach	43
3.4 Data	44
3.4.1 Informality: What are we measuring?	45
3.5 Results	46
3.5.1 Children's activities	49
3.6 Concluding remarks	49
3.7 Tables and Figures	55

4	Youths in the Informal Sector: Evidence from Brazil	60
4.1	Introduction	60
4.2	Informality in Brazil	62
4.2.1	Youth in the informal sector	64
4.3	Data	65
4.3.1	Descriptive statistics	66
4.4	Econometric strategy	67
4.5	Results	69
4.6	Concluding remarks	71
4.7	Tables	76
	Conclusion	81

List of Figures

1.1	Public spending on education and the price of oil. Source: Sistema Integrado de Indicadores Sociales de Venezuela (SISOV)	13
1.2	Schooling in Venezuela, 1994 and 2004 (%). Source: EHM household survey . . .	13
1.3	Gross enrollment rate. Lower and upper secondary school (%). Source: UNESCO	14
1.4	High school dropouts by type of activity and by semester (as % of total dropouts). Source: Author's calculations based on EHM	16
1.5	High school dropout rate across age and by semester. Source: Author's calculations based on EHM	16
1.6	Education expenditure in primary, secondary, and tertiary level as % of total education expenditure (%) in 2006. Source: UNESCO	17
2.1	Geographic division of Venezuelan territory by regions. Source: made by the author	30
3.1	Share of households with at least one informal working parent by three definitions of informality. Source: Author's calculations based on EHM	56

List of Tables

1.1	Comparison of GDP and income inequality measures among oil producing countries. Source: United Nations.	12
1.2	Comparison of GDP and income inequality measures among South American countries. Source: United Nations.	12
1.3	Net enrollment ratio. Source: UNESCO	14
1.4	Status of high school-age children (as % of total high school-age children). Source: Author's calculations based on the national household survey (<i>EHM</i>)	15
1.5	Share of early high school leavers. Source: Author's calculations based on <i>EHM</i> .	15
1.6	Reasons for not attending school (as % of total dropouts, 2005-2007 average). Source: Author's calculations based on Encuesta de Hogares por Muestreo (<i>EHM</i>)	17
2.1	Montecarlo experiment results: a comparison of the bias and mean-squared error (MSE) of five different estimators as applied to pseudo-panel data. $T = 5, C = 8, \rho = 0.7$	31
2.2	Montecarlo experiment results: a comparison of the bias and mean-squared error (MSE) of five different estimators as applied to pseudo-panel data. $T = 5, C = 30, \rho = 0.7$	32
2.3	Montecarlo experiment results: a comparison of the bias and mean-squared error (MSE) of five different estimators as applied to pseudo-panel data. $T = 5, C = 80, \rho = 0.7$	32
2.4	Montecarlo experiment results: a comparison of the bias and mean-squared error (MSE) of five different estimators as applied to pseudo-panel data. $n_c = 100$ and $\rho = 0.7$. N, T and C vary.	33
2.5	Montecarlo experiment results: a comparison of the bias and mean-squared error (MSE) of five different estimators as applied to pseudo-panel data. $T=10, C=30, N=1500, n_c=50$. ρ varies.	34
2.6	States by region. Source: Plan Nacional de Desarrollo Regional 2001, Ministerio de Desarrollo y Planificacin.	34
2.7	Construction of pseudo-individuals. Sample restricted to employees who are between 15 and 65 years of age. Source: authors' calculations based on the 1994 to 2004 <i>EHM</i> household surveys.	35
2.8	Descriptive statistics on grouped pseudo-panel data: 528 observations. Source: 1994 to 2004 <i>EHM</i> household surveys.	35

2.9	The returns to education in Venezuela, 1994-2004. Pseudo-panel regressions using various estimators. The number of observations = 528. Standard errors in parentheses.	36
3.1	Descriptive statistics 2005-2007. Per capita income is measured in <i>Bolivares</i> at constant prices, base 2007. Source: Author's calculations based on EHM	55
3.2	Random effects logit model on informal employment. Dummies for state, activity, and time are included. Agriculture is the omitted activity. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1	56
3.3	Dependent variable: Child Dropouts - pooled logit, random effects logit and panel GEE - 2005-2007. Dummies for state, activity, and time are included. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1	57
3.4	Dependent variable: Child Dropouts - random effects logit with Mundlak correction - 2005-2007. Dummies for state, activity, and time are included. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1	58
3.5	Multinomial logit model (MNL) for three alternatives - Work Only, Study & Work, Idle - 2006-2007. Dummies for state, activity, and time are included. Coefficients as odd ratios. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1	59
4.1	Descriptive Statistics 2003. Remuneration is measured in <i>Reais</i> at constant prices. The last two columns present the mean of each variable by age group. N=5594. The asterisks show the statistical significance of differences: *** p<0.01, ** p<0.05, * p<0.1. Source: Author's calculations based on ECINF 2003.	76
4.2	Differences between young (<24 years) and not young (>24 years) workers (measured as % of each age group unless mentioned otherwise). Source: Author calculation based on ECINF 2003.	77
4.3	Share of Brazilian young workers by industry (2003). Source: Author calculation based on ECINF	77
4.4	Dependent variable: Log of value added per worker - 2003. Dummies for industry and region are included. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1	78
4.5	Dependent variable: Log of value added per hours worked - 2003. Dummies for industry and region are included. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1	79
4.6	First-stage. Dependent variable: Share of young workers - 2003. Dummies for industry and region are included. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1	80

Introduction

The positive effect of education on economic growth is undeniable. Pscharopoulos (1984)¹ presents a good summary of the first research works on the subject. He argues that most studies underestimate the positive contribution of education to economic growth and on social welfare, because they do not take into account the indirect effects through other explanatory variables (e.g. the efficient use of new inputs or being healthy).

Thanks to the increasing release of micro data, education indicators have been also used to explain other issues such as productivity, quality of institutions, fertility and health. It is not only relevant to analyze what the contribution of the education is to the social welfare, but also to try to understand how the lack of education will affect society. One of the serious consequences of the school dropout is higher crime rates (see Lochner and Moretti 2003,² Rud et al 2013³) which, in turn, can also increase the unemployment rate of the community and have a negative impact on economic growth in the region.

Few studies, as in Pritchett (2008),⁴ show that the effect of education on the GDP growth is small, and for some countries this effect is non-existent. However, he argues that government should continue to invest in education due to its contribution in reducing other important barriers to development such as child mortality.

We analyze some of the challenges in the education sector in Venezuela. The country is an interesting case of study, given that it possesses the largest proven oil reserves on earth, has benefited from rising oil prices in the last decade (250% average increase since year 2000 compared to 1990s), a growth in GDP per capita of around 21% since 1990 (US\$ 6,415 in 2012), and at the same time, it has managed to consistently reduce poverty from 40% in 1990 to 24% in 2012, and income inequality, measured by the Gini index, from 0.47 in 1990 to 0.41 in 2012. Nevertheless, educational outcomes are mixed. In a sense, Venezuela is an upper-middle income country with challenges similar to those of low and low-middle income countries. The country is perhaps the best recent illustration of the natural resource curse (Auty, 1993)⁵, where despite

¹Pscharopoulos (1984) "The Contribution of Education to Economic Growth: International Comparisons" In *International Productivity Comparisons and the Causes of the Slowdown*, ed. J. Kendrick. Ballinger for the American Enterprise Institute, 1984. 335-60.

²Lochner, Lance and Moretti, Enrico (2003) "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Report" NBER Working Papers 8605, National Bureau of Economic Research, Inc.

³Rud, I & Van Klaveren, C. & Groot, W. and Maassen van den Brink, H. (2013) "Education and Youth Crime: a Review of the Empirical Literature" Working Papers 48, Top Institute for Evidence Based Education Research.

⁴Pritchett, Lant (2006) "Does Learning to Add up Add up? The Returns to Schooling in Aggregate Data" *Handbook of the Economics of Education*, Elsevier.

⁵Auty, Richard (1993) "Sustaining Development in Mineral Economies: The Resource Curse Thesis" London: Routledge.

enormous windfalls from rocketing commodity prices, poor management of public finances and weak institutions have hampered economic performance and human development, particularly in education.

The purpose of this thesis is to empirically assess, through the use of novel and econometrically robust approaches, the drivers of school attendance, the role of informality in school dropout, and the linkages between youth work and productivity in small firms.

The document is divided into four chapters. In the **first chapter**, we begin with a summary of the economic context of Venezuela, followed by a survey of the education system in Venezuela from 19th century until present, focusing on issues related to access to education, quality of education and the main legal instruments that affect education and child labor in the country. The aim of this first chapter is to provide a background of the situation in the education sector, in order to get a better understanding of the context in which most of our empirical results are derived from. The first chapter reveals that the most serious problems in the Venezuelan education system in the 21st century are dropouts in secondary school and the quality of education. Since reliable data on quality is not available, we focus on the phenomenon of school dropouts.

In the **second chapter**, we take a close look at the dynamics between income and education, a potential driver of school attendance or, broadly speaking, investments in human capital, through the classical Mincer approach. We address two main question in this chapter. First, from an economic perspective, we are interested in determining how important the returns to education are in Venezuela, put differently, what is the market premium of an additional year of schooling? If returns are high, then structural drivers, beyond expected wages and income, may explain the (low) school attendance. Second, from an econometric perspective, we address the issue of using repeated cross-sections in the absence of panel data. Our approach consists in a) aggregating groups of individuals (cohorts) sharing common time-invariant characteristics, and b) providing an alternative to the standard measurement error model approach to pseudo-panels, through the use of consistent instrumental variables estimators based on sample moments of order higher than two.

Building on our previous findings we know that the number of dropouts increases in secondary school, more specifically the number of school dropouts rises at the end of lower secondary school and at the beginning of upper secondary school (around 9 years of schooling). Consequently, the aim of the **third chapter** is to understand the causes of school dropout in secondary education. For instance, household survey data shows that the main reasons in Venezuela for not attending to school are: “does not want to study” (37.45%), “work” (17.46%), “handicap” (5.49%), “no vacancies” (5.42%), and “cannot afford it” (4.53%). As we can see, the lack of enthusiasm for attending school is possibly a major factor associated with dropouts. This illustrates that factors other than income may also be important when deciding whether to keep children out of school. At the same time, a child’s interests may be affected by other factors, not always measured in the survey, e.g., ability, school quality, and access to school. The econometric approach consists in eliminating the endogeneity bias by using Mundlak corrections, since there is neither a good instrument nor can fixed-effects estimators be used in this study because many relevant variables do not vary over time. The endogeneity bias may come from the unobserved heterogeneity, which may have an effect on the decision of dropping out

of school and also on some of the right hand side variables. Our contribution in this chapter is twofold. First, the inclusion of informality (informal worker parents) as one of the main factors affecting the likelihood of children completing high school. This issue has not been fully explored in the available literature on the subject. Second, the use of recent panel data in assessing the problem of early high school dropouts, provides new and more robust insight into Venezuela's educational bottlenecks.

Chapter 4 takes a closer look at the linkages between young workers and informality from a small firm perspective. For instance, there are several risks that young workers may face if their first job is informal salaried. Ultimately, this may affect the firm's productivity where they work. First, applicable to all young workers, work keeps them from allocating time to their studies. Therefore, the process of human capital accumulation is undermined and consequently, they fail to get a better job. Second, in the event of a crisis, unregistered workers are more likely to become unemployed. The third risk related to informal jobs is that young workers who drop out of school before completing secondary education are less likely to migrate to formal jobs in the future. Fourth and the focus of this chapter, hiring young workers whose characteristics are low skills and lack of experience can negatively affect productivity of the business. One mechanism can be the fact that young workers have less professional experience and lack the "soft" skills than older workers may have improved throughout their professional life. Another explanation may be that young workers are more likely to study and have a part time job. Hence, it would be difficult for them to be fully committed and highly productive. In this chapter, we analyze the effect of the share of young workers on small unregistered firms' productivity in Brazil,⁶ following an instrumental variable approach. In this type of specification, two sources of endogeneity problems emerge. First, unobserved heterogeneity can affect both changes in the share of young workers and changes in productivity. The second source of endogeneity is the simultaneity or reverse causality between productivity and age composition. We use the method of Instrumental Variable in order to ensure that our results are not biased. We try to answer the following questions: do parents owning an unregistered firm incorporate their daughters and sons into the business? If that is the case, how does the presence of youths in the firm affect the productivity of the business?

⁶Since surveys on small firms are inexistant in Venezuela, we use data from Brazilian informal firms.

Chapter 1

The Venezuelan Context: Economy and Education System

Venezuela is the seventh oil-producing and oil-exporting country in the world. At the national level, oil-related activities are an important determinant of socioeconomic dynamics. In 2006, oil-related activities represented 15% of GDP, and accounted for almost 88% of total exports. Oil is extremely important on the public finance front, contributing 48.6% of total current fiscal income. This high dependence on the oil sector makes the Venezuelan economy very vulnerable to fluctuations in the international price of crude. Positive oil shocks generally induce an excessive increase in public expenditures and very significant currency appreciations. On the other hand, negative oil shocks translate into severe cuts in public sector expenses, exchange rate depreciations and increases in the unemployment rate.

Table 1.1 compares Venezuela to other oil-exporting countries in terms of five indicators: the Human Development Index, GDP per capita, the Gini coefficient, the rich/poor ratio and an education enrollment indicator.¹ According to the World Bank classification, Venezuela is an upper-middle income country (GDP per capita of US\$ 4,214 in 2004). Venezuela displays a lower per capita GDP than Saudi Arabia or Mexico, whereas its GDP per capita is higher than that of Ecuador or of Russia. While Venezuela may be relatively well-off in terms of GDP per capita, this is not the case when it comes to measures of inequality. As shown in Table 1.1, Venezuela's income distribution is highly unequal, especially when compared to other oil-exporting countries. The richest 10 percent of the Venezuelan population earns approximately 20.4 times more than the bottom 10 percent. This is in contrast to an oil-exporting country such as Norway (one of the most egalitarian societies in the world), where the corresponding figure is 6.1.

As shown in Table 1.2, Venezuela's GDP per capita is among the highest in Latin America, while its performance in terms of inequality is relatively good within a region characterized by highly unequal income distributions. On other hand, Venezuela's performance in the education sector is poor when compared to other countries in the region. The combined gross enrollment ratio for primary, secondary and tertiary schools for Venezuela is 74.2% (as a percentage of the population of official school age for the three levels), situating the country amongst the bottom

¹The method used for calculating the two inequality measures is not homogeneous among countries, and figures are therefore not strictly comparable. Inequality data are not available for some Arab countries.

three in Latin America, alongside Colombia and Paraguay.

During the past five years, the Venezuelan labor market has been relatively slack, with the unemployment rate averaging 15%. The informal sector accounts for half of all workers.² This sector has been a growing concern for Venezuelan policymakers: in 1975, the ratio of informal workers to total workers in the economy was about 38%, increasing to 50% in 2002 [Ortega, 2003]. According to Patrinos and Sakellariou [2005b], this expansion of the informal sector has generated a fall of average earnings, given that wages are lower than those in the formal sector. The authors also argue that the difference in wages between the formal and informal sectors is a key determinant of income inequality.

Labor market imbalances and limited access to jobs have an important effect on the returns to education. For instance, the fact that skilled wages are greater than unskilled wages can be explained not only by the higher level of human capital of skilled workers, but also by an excess demand for their services.

As mentioned above, oil prices play a crucial role in explaining socioeconomic changes in Venezuela. It is thus not surprising, as suggested by Patrinos and Sakellariou [2006], that economic fluctuations have had an impact on the demand for and the supply of education, as well as on skilled/non-skilled labor demand, which in turn, also affects the returns to education. A typical example, in the context of an oil price downturn, is the case of children or youths who drop out of schools and start working because of the financial difficulties encountered by her family in the short term. This phenomenon might increase the low skilled labor offer. On the other hand, a supply side example is given by the increases in public expenditures on education when oil prices are high, as illustrated in Figure 1.1.

1.1 Education System

Since the early 19th century, several Venezuelan governments have put education at the top of their priorities. This political will is illustrated by the increasing investment and the changes in the legislation on the national education system. Education policy has been centered on eradicating illiteracy and increasing the enrollment rate in primary school.³ However, results have not always been as successful as expected. Two main factors had a negative impact on social investment in human capital before the 20th century. First, after Venezuela gained its independence from Spain at the beginning of the 19th century, the country was ravaged by almost a century of civil conflicts. Second, the discovery of oil in the late 1800s set the country on a process of deep structural changes. From an agriculture-oriented country, Venezuela became an oil exporter. The high dependency on the oil industry, a sector characterized by significant price volatility, contributed to increase social vulnerability. Hence, these two factors generated an environment of political instability that made sustainable education policies unlikely.

For the last 15 years, education has been a very controversial issue in Venezuela due to the implementation of numerous social programs and important reforms at different levels of

²This definition of informal sector includes: non professional workers with their own businesses, companies with less than five workers and those who work in the domestic service sector.

³See Pérez and Carrasquel [1996] for more information about legislation on the national education system.

education.⁴ The literacy rate of 95.2% in Venezuela is relatively high among Latin American countries [UNESCO, 2007].

Figure 1.2 shows that, in 1994, 44.36% of the Venezuelan population above 5 years of age had completed primary schooling, while 16.39% reached the secondary level and just 3.23% obtained a university degree. Ten years later, in 2004, the situation slightly improved, with the corresponding numbers being 44.09%, 20.82%, and 3.82%.

One of the main problems still to be addressed is school attendance, particularly in upper secondary school. Although education indicators have been improving over recent years, Venezuela's education system is lagging behind those of other Latin American countries. Regarding primary level, Table 1.3 shows a high enrollment ratio irrespective of age group. It means that either primary school-age children have no problem attending school or that a group of children, under/over primary school age, attend primary school. Although on average, 106% of children are enrolled in primary school, only 92% are primary school-age children.

In contrast to primary level, at the secondary level, conditions are alarming in terms of the gross enrollment ratio. Regardless of age group, in 2006, 89% and 64% of children were enrolled in lower and upper secondary school, respectively. In order to demonstrate how low the access to secondary school is in Venezuela, comparative statistics are presented for Latin American countries, Spain, and the United States. Figure 1.3 shows that even if the number of children attending lower secondary school increased from 66% to 89% during this period, Venezuela still has one of the worst performances in the region, along with Ecuador and Mexico. Results are not better if the gross enrollment ratio at upper secondary level is considered. Clearly, a significant number of children leave school after having finished the primary grades, but even more desert after the lower secondary level. In fact, school life expectancy from primary to secondary level, was 10.3 years in 2009 [UNESCO, 2007], which is equivalent to reaching the first year of upper secondary school.

The micro data from the national household survey also corroborates dropout rates in secondary school. Table 1.4 shows four activities in which high school-age children spend their day. In 2005-2007, the high school dropout rate remained over 9% (see Table 1.5). Among early school leavers, on average, 65% have remunerated or non-remunerated work and the rest, (35%), falls under the "idle" category (see Figure 1.4). For those who report being involved in an economic activity, almost 50% do not receive any remuneration; the two more relevant types of activities declared by them are household chores (most girls fall this category) and non remunerated family work (author's calculation using the national household survey).⁵ Even though there are more working than idle children, the share of the latter is still considerable (more than 30%). Baland and Robinson [2000] argue that children who report doing "nothing" not only lose educational opportunities but also suffer the lack of social and cognitive skills. Thus, to understand the causes of dropouts, it is important to distinguish between working and idle children.

The high dropout rate might be related to the factors of demand and/or supply. Demand-

⁴Since 1999, the Venezuelan government has implemented the *Proyecto Educativo Nacional* (PEN); see Bonilla-Molina [2004].

⁵Blanco and Valdivia [2006] also find a sort of "gender specialization", in which girls are in charge of household chores, and boys do activities more related to the labor market.

side factors, such as socioeconomic conditions within households, are included in this analysis, given their current data availability. On the other hand, supply-side factors could be divided into two types, child labor laws or infrastructure/quality status. However, these are difficult to measure. Regarding the former, a commonly used complementary policy to compulsory schooling is a ban on child labor. According to the Law for Child and Youth Protection (*Ley Orgánica para la Protección del Niño y Adolescente*) and the Labor Law (*Ley Orgánica del Trabajo*) in Venezuela, 14 is the minimum age for admission to a nonhazardous employment or work. Although we do not intend to evaluate the impact of this law in this work, it is interesting to see whether the distribution of the incidence of high school dropouts shows a breaking point at the age of 14. Figure 1.5 shows no sudden changes in the slope; instead there is a steady increase of high school dropouts across ages. Blanco [2009] finds that 70% of children, between 10 and 17 years old, are involved in child labor. This group includes children below the minimum age and teenagers between 14 and 17 years old working in hazardous places.⁶ As we mentioned above, we are not able to include the educational infrastructure or quality situation in the analysis due to data limitations. Yet the next two subsections provide the context concerning the access to school and the quality of education in Venezuela.

1.1.1 Access to School and Government Expenditure

As shown in Table 1.6, about 10% of secondary school-age children leave school because either their parents cannot afford it or there are no vacancies. These subjective responses can shed light on how the direct and indirect costs of education affect educational attainment. This section looks into the cost of schooling in terms of access and infrastructure. The lack of schools and the cost of education are the main problems affecting access to school. Some studies dealing with access to school use “distance to school” as a proxy for transportation costs. Unfortunately, the survey that we use does not include this valuable piece of information. However, we do analyze school conditions in terms of lack of educational infrastructure and government expenditure on education.

The amount of monetary resources devoted to education since the 20th century has been very unstable in Venezuela; moreover, it has been very difficult to enforce laws on education. The introduction of universal education in the Constitution of 1961 was translated into an expansion of the education system in terms of students registered, notably at the tertiary level. However, public expenditure on education has suffered from ups and downs for the last 50 years [Bonilla-Molina, 2004]. This instability has been caused mostly by the debt crisis in the 80s and significant oil-price fluctuations (70s, early 80s, 1994, 1998, and 2002).

When an expansionary fiscal policy is implemented, two questions arise: how much is the share of public expenditures on education, but principally, how are those resources used? Public expenditure on education as a percentage of total public expenditure attained its maximum in 2000 (23%), exactly when many social programs were put in place.⁷ The share of education

⁶Moehling [1999] finds that minimum age limits contributed little to the long run decline in child labor. The author uses data from the 1880, 1900 and 1910 United States’ censuses. Whereas Angrist and Krueger [1991] find, also using data from United States, that compulsory school legislation was effective for all men born between 1920 and 1929.

⁷There was a fall in 2002 due to the oil industry and private sector strikes [UNESCO, 2007].

in the total budget was 21% in 2004; compared to those of other countries in the same year such as Chile (18.5%), Mexico (25.6%), or France (11.3%),⁸ it is not an insignificant amount. Nevertheless, even if it represents a significant amount of resources, its relative importance, compared to those of other public expenditures, has remained unchanged for the last 15 years. The UNESCO-IBE report (2006) also mentions that the share of public expenditure on education in total social expenditure (43%) in early 2000 is similar to the share registered in the 90s, even lower than in 1994. In terms of GDP, public expenditure on education has increased and surpassed the 4% barrier.⁹

If we take a look at the distribution within expenditures on education, Figure 1.6 shows an interesting view of the situation. The tertiary level takes the biggest share, compared to those of the primary and secondary levels. This does not necessarily mean that tertiary education is well supplied; rather, the costs of salaries, inputs, and infrastructure are higher at this level of education. More alarming is the very low amount of resources assigned to the secondary level, particularly upper secondary. The low level of expenditure on this key factor might generate important disincentives in finishing secondary school, either due to insufficient secondary schools or serious deficiencies in the quality of education.

Bonilla-Molina [2004] argues that the capacity of both private and public primary schools to accommodate school-age children decreased during the 1970-1990 period. Indeed, the ratio of primary school-age population to primary institution was among the highest in Latin America. This means that the lack of primary schools was a major cause of low primary enrollment rates at that time, as mentioned above. Yet this ratio began to recover from 2000 when the education reform called PEN was implemented, and education programs called *misiones* were put in place. These initiatives involved the construction of schools,¹⁰ as well as legal reforms to implement a more “humanistic and participative” education. To our knowledge, no formal analysis has been conducted about the specific effects of these reforms on enrollment rates or student performance. Unfortunately, due to unavailable data, we are not able to analyze this issue for the secondary level, which currently shows a low enrollment rate. However, we are able to generate some information about the supply side of education for the secondary level from Bonilla-Molina [2004]. According to the author, 18,184 institutions are devoted to primary education in Venezuela, whereas only 5,142 institutions are in charge of secondary education. Since secondary and primary schools have, on average, 2.3 and 0.9 classrooms per grade, respectively, we can roughly infer that almost half of outgoing primary students cannot enroll in secondary-level institutions. There is a serious supply problem of secondary schools, which is corroborated by low enrollment rates at this level (60%-80%).

The last report of Guadalupe, Castro, and Taccari [2007] points out that Venezuela is among the 5 worst performers in terms of culmination rate (i.e., less than 75%) at the lower secondary level in Latin America and the Caribbean. Moreover, the authors show that Venezuela, with 16 other countries, failed to ensure the completion of upper secondary education for 50% of the

⁸See UNESCO [2007].

⁹See UNESCO and International Bureau of Education [IBE] [2006].

¹⁰According to Duflo [2000], a huge program aimed at increasing the supply of schools in Indonesia had positive effects on children’s enrollment rates in primary school. It is worth noting that even though no efforts were made to improve quality, this simple and straightforward program based on building schools increased the school life expectancy.

population between 20 and 39 years old. These facts highlight the seriousness of the problem at the secondary level.

1.1.2 Quality of Education

The regional office of Education for Latin America and the Caribbean [UNESCO-OREALC] carried out a challenging project in 1990, which sought to establish an evaluation system of the quality of education. Although at that time, illiteracy was showing a steady decline and education coverage was improving slowly, student performance was harshly criticized [Bonilla-Molina, 2004]. It was pointed out that there were no standard measures to evaluate the quality of education systems, neither by level of education nor by topic. One of UNESCO's initiatives was to determine the knowledge level for a sample of developing and poor countries. Thirty-one countries participated in the project, and a representative group of 13-year-old children took an external test of reading comprehension. According to Bonilla-Molina [2004], the results showed that Venezuela ranked 28th out of 31 countries, just above Nigeria, Zimbabwe, and Botswana. The study also found a lack of capacity and quality among teachers and a strong centralization of the Venezuelan education system. Thanks to this project in 1991, the Ministry of Education and UNESCO-OREALC defined 13 indicators to evaluate the performance in literature and mathematics of second and fourth graders in primary school (as cited in Bonilla-Molina [2004]).

The government established a set of policies in 1989, aiming at reforming the state in terms of modernization and efficiency at all levels. In the case of education, some of the measures were to facilitate professional training and other benefits for teachers, as well as to harmonize the curricula with the changing political, social, economic, and cultural aspects of society. This reform also took into account the need to keep children enrolled in school and prevent dropouts, so the government launched a series of social programs. Basically, those programs were based on subsidizing the cost of school supplies and assigning food vouchers. In spite of all these efforts, the education system could not avoid the damage from the structural crisis shaking the country at that time.

According to Guadalupe, Castro, and Taccari [2007], the quality of education can be measured in 5 dimensions: relevance, appropriateness, equity, efficacy, and efficiency. Each dimension has a set of indicators; some of them are qualitative, coming from laws or regulations, while others are quantitative statistics like the enrollment rate. One of the most relevant drawbacks about the quality of education in Venezuela is the poor promotion of information technology and communication as a tool for knowledge [Guadalupe et al., 2007]. In terms of appropriateness, the indicators show that the curricula are quite diversified and flexible, thus encouraging participation. However, efforts must address the curricula's adaptability to the social and cultural diversity of pupils. Regarding the efficacy, the report underlines that although literacy and primary education are well developed, the coverage of pre-school and secondary school is very poor. In terms of efficiency, the repetition rate, either in the primary or the secondary level, is less than 10%. Thus, the relationship between the children falling behind and those dropping out of school might not be very strong. Finally, with respect to equity, there is a slight difference of genders, whereas the differences between the richer and poorer quantiles are very noticeable. Girls (or richer quantile) are more likely to enroll in and complete secondary

school than boys (or poorer quantile). Thus, economic factors seem to be related to dropouts at the secondary level. Regarding education quality issues, it is also important to look into the performance of teachers and their environment. Since the teacher can pay more attention to each pupil in a small class, it is more profitable teaching a smaller class than a crowded one.

Pre-primary, primary, and secondary classes in Venezuela do not exceed 20 pupils per teacher, on average in 2005. Thus, the education system does not seem to suffer from crowded classes. By comparison, Spain and Argentina have a pupil-teacher ratio of 14 and 17, respectively, in their primary schools, whereas in Venezuela the ratio was 19 in 2005. However, this ratio is lower for the pre-primary and the secondary levels (15 and 11, respectively).

The percentage of trained teachers ranged from 82% to 86% on average, among all levels of education. Thus, roughly 14% to 18% of classes had low-skilled teachers, probably having a negative impact on children's performance. This drawback in the education system is related to the neglect of the teaching profession in the mid-1900s [Bonilla-Molina, 2004]. The segment of the Education Law referring to the professional training of teachers came into effect in 1941, but only in 1980 did teacher training become a profession at the university level. Then regulations in 1970 introduced the competition system as a means to select the most qualified teachers. However, the report identified some pitfalls in the capabilities and quality related to the teacher training in 1990.

1.1.3 Reforms and Legal Framework

Since more than 27% of dropout children are working, it is pertinent to have some insight about laws related to education and child labor. In this subsection we look through the laws and regulations linked to the education system in Venezuela.

The Organic Law of Education, promulgated in 1980, established 10 years of compulsory education instead of 7 years, which means that children are obliged to finish basic school. That covers kindergarten (1 year), primary (6 years), and lower secondary education (3 years). The purpose of this reform was to increase the cultural level of the people, to improve the quality of teaching, and also to satisfy the expectations of the Venezuelan society [Pérez and Carrasquel, 1996]. This new law was part of a regional agreement among Ministers of Education in a meeting celebrated in Mexico in 1979. They agreed to improve the access to school for children aged between 6 and 14 years-old. The main objective of all these measures was to cope with the low enrollment ratio at the primary level; a crucial issue at the time.

From an empirical perspective, a valuable attempt to analyze the effect of the compulsory education reform in Venezuela is carried out by Patrinos and Sakellariou [2005a]. They argue that compulsory schooling has a significant effect on the level of education attainment. They compare the impact of this reform on two groups of individuals, the treated and the control group. The treated group is composed of those affected by the reform, that is the children who were less than 12 years-old in 1980. The control group consists of children that were aged above 12 in 1980.

Given that there is no contractual group at the same period of time, treated and non-treated individuals made the decision of continuing school at different times. We think that it is more difficult to measure the real impact since the reform affects all citizens at the same

time. Moreover, it is hard to monitor how the government enforces the law.

A commonly used complementary policy to compulsory schooling, is a ban on child labor. According to the law for Child and Youth Protection and the Labor Law in Venezuela, the minimum age for admission to a non-hazardous employment or work is 14 years-old. In some cases, children under the minimum age are allowed to work as long as it does not jeopardize education activities. Since a 14 years-old child is, on average, about to finish lower secondary, this regulation can be linked to the fact that children abandon school after having achieved lower secondary level (10 years of education). Laws offer children the same rights than adult workers in terms of benefits like vacations and social security. Unlike eight hours working day for adults, children are not allowed to work more than six hours a day. Assuming the same conditions at work, the salary must be the same as for an adult, either in rural or urban work.

It is interesting to look at the effect of this supply-side policies on the number of working children. According to the theoretical approach developed by Basu and Van [1998], the child labor model has multiple equilibria, one in which adults and children work and the other one in which only adults work. They demonstrate that a policy intervention banning child labor will have a positive effect, reducing the number of working children. As children leave their jobs, the salary will increase and therefore families would not need to send children to work. They argue that once the economy reach the equilibrium where only adults work, the law is no longer needed.¹¹ The authors also support this type of ban if it is accompanied by monitoring school attendance records. However, they alert about some situations in which the effect of a ban can be ambiguous; for instance, when there is unemployment in the economy or if the ban does not reach all economic activities in the country. In the case of economies showing extremely low and rigid wages, this policy can worsen household subsistence level.

Using early data from the United States (1880-1910), Moehling [1999] demonstrates that more than minimum age limit regulations, the driving forces of the decline in child labor are economic dynamics like technological change, immigration and household optimizing behaviors. As the legal restriction in schooling was not imposed in all states in the U.S., the author was able to construct a control group, indispensable when evaluating the effect of this kind of reform. She compares occupation rates for 13 and 14 years-old, *before and after* the introduction of minimal age law and between states that enacted the minimum age limit and states that do not comply with this law. By means of *difference-in-difference-in difference* analysis (*DDD*), she finds that the estimated effect of legal restrictions on the probability of child labor is small and statistically insignificant. Since compulsory schooling is the other main restriction concerning child labor, she also makes the difference between three kind of scenarios: children who fulfill both minimal age limits to work and compulsory schooling, those with just one of them, and finally those who do not satisfy with neither of these laws. The results do not change.¹²

As we mentioned above given the massive dropout rates at the end of lower secondary level, it is worth studying the secondary level in depth. A curricula reform was made in Venezuela at

¹¹Baland and Robinson [2000] suggest that a marginal ban on child labor is Pareto improving when the positive effect on parent's current wages is greater than the fall of future wages.

¹²Similarly, Angrist and Krueger [1991] applied a *natural experiment* statistical technique, to evaluate the impact of compulsory schooling laws in the U.S. The authors perform a *difference-in-difference* analysis based on within-state and within-age-group comparisons. They find a statistically significant effect of this legal restriction in the enrollment rate in 1960 and 1970 in US. They also pointed out that this effect vanished over time.

the beginning of 70's. This reform consisted in a new composition of secondary school [Pérez and Carrasquel, 1996], by dividing it into two levels, lower and upper. The principal reason was to differentiate between general subjects for everyone (lower secondary), and specialized subjects, mainly Major in Humanities and Major in Science (upper secondary).¹³ This kind of reform might have an ambiguous effect on the attendance rate. According to World Bank [2005], there is a gap between the knowledge acquired at the schools and the skills required by the labor market. Therefore, the secondary school plays an essential role in preparing children either for further education or for a professional field. An important question here is how general or specialized the curricula must be to retain children at schools. The authors argue that it depends on the dynamic of the labor market and the need of a changing and globalized society. For instance, some countries are moving away from vocational curricula toward general programs, due to the rates of returns in the latter are higher.

¹³Students can choose what option or major they want to enroll in.

REFERENCES

- J. Angrist and A. Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.
- J. Baland and J. Robinson. Is child labor inefficient? *The Journal of Political Economy*, 108(4):663–679, 2000.
- K. Basu and P. H. Van. The economics of child labor. *The American Economic Review*, 88(3):412–427, 1998.
- F. Blanco. *Trabajo infantil en Venezuela: 1998-2007. Understanding Children’s Work Project*, 2009.
- F. Blanco and C. Valdivia. Child labour in venezuela: children’s vulnerability to macroeconomic shocks. *Understanding Children’s Work Project*, 2006.
- L. Bonilla-Molina. *Historia Breve de la Educación en Venezuela*. Gato Negro, Caracas, Venezuela, 2004.
- E. Duflo. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *NBER*, (7860), 2000.
- C. Guadalupe, I. Castro, and D. Taccari. The state of education in latin america and the caribbean: Guaranteeing quality education for all. Technical report, UNESCO-PRELAC, 2007.
- C. Moehling. State child labor laws and the decline of child labor. *Explorations in Economic History*, 36:72–106, 1999.
- D. Ortega. *Descripción y Perfiles de Desigualdad de Ingresos en Venezuela, 1975-2002. Documento para la Discusión, Programa de las Naciones Unidas para el Desarrollo*, 2003.
- H. Patrinos and C. Sakellariou. Schooling and labor market impacts of a natural policy experiment. *Labour*, 19(4):705–719, 2005a.
- H. Patrinos and C. Sakellariou. Schooling and labor market impacts of a natural policy experiment. *LABOUR*, 19(4):705–719, December 2005b.
- H. Patrinos and C. Sakellariou. Economic volatility and returns to education in venezuela: 1992-2002. *Applied Economics*, 38(17):1991–2005, 2006.
- E. Pérez and P. Carrasquel. *Sistema Educativo Nacional de Venezuela*. <http://www.oei.es/quipu/venezuela/#sis>, 1996. *Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura, and Ministerio de Educación de Venezuela*.
- UNESCO. Institute for statistics. data centre. http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=136&IF_Language=eng&BR_Topic=0, 2007. Accessed: 2009-09-10.

UNESCO and International Bureau of Education [IBE]. World data on education. <http://www.ibe.unesco.org/Countries/WDE/2006/index.html>, 2006.

World Bank. *Expanding Opportunities and Building Competencies for Young People. A New Agenda for Secondary Education*. Washington, DC, 2005.

1.2 Tables and Figures

Table 1.1: Comparison of GDP and income inequality measures among oil producing countries. Source: United Nations.

Country	HDI Ranking (2006)	GDP per capita \$US 2004	richest 10% ÷ poorest 10% (2006)	Gini (2006)	Combined enrollment ratio (% , 2004)
Norway	1	54,465	6.1	25.8	100
UAE	49	24,121	60
Mexico	53	6,518	24.6	49.5	75
Russia	65	4,042	12.7	39.9	88
Venezuela	72	4,214	20.4	44.1	74
Saudi Arabia	76	10,462	59
China	81	1,490	18.4	44.7	70
Ecuador	83	2,322	44.9	43.7	..
Iran	96	2,439	17.2	43.0	72
Algeria	102	2,616	9,6	35.3	73
Indonesia	108	1,184	7.8	34.3	68
Nigeria	159	560	17.8	43.7	55

Table 1.2: Comparison of GDP and income inequality measures among South American countries. Source: United Nations.

Country	HDI Ranking (2006)	GDP per capita \$US 2004	richest 10% ÷ poorest 10% (2006)	Gini (2006)	Combined enrollment ratio (% , 2004)
Argentina	36	3,988	34.5	52.8	89.3
Chile	38	5,836	40.6	57.1	81.3
Uruguay	43	3,842	17.9	44.9	89.4
Brazil	69	3,284	57.8	58.0	85.7
Colombia	70	2,176	63.8	58.6	72.9
Venezuela	72	4,214	20.4	44.1	74.2
Peru	82	2,490	40.5	54.6	86.4
Ecuador	83	2,322	44.9	43.7	..
Paraguay	91	1,220	73.4	57.8	69.7
Bolivia	115	974	168.1	60.1	86.5

Figure 1.1: Public spending on education and the price of oil. Source: Sistema Integrado de Indicadores Sociales de Venezuela (SISOV)

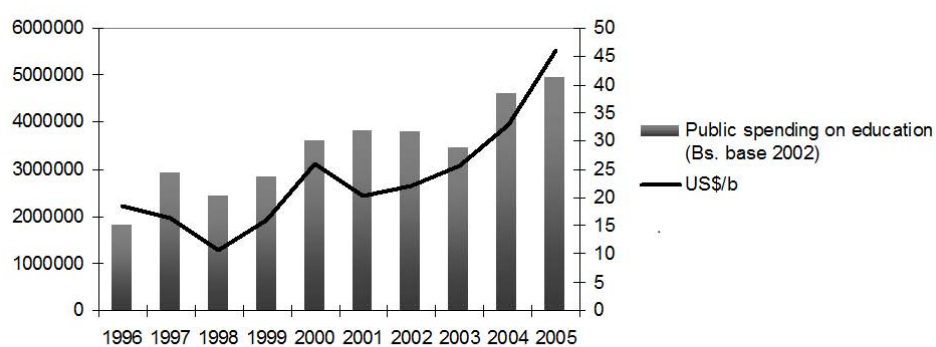


Figure 1.2: Schooling in Venezuela, 1994 and 2004 (%). Source: EHM household survey

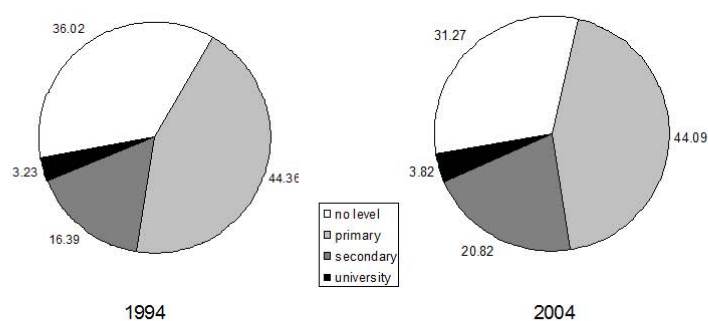


Table 1.3: Net enrollment ratio. Source: UNESCO

Level of Education	1999	2001	2003	2004	2005	2006	2007
Gross enrollment rate							
Primary	99.92	104.13	104.64	104.68	104.45	104.32	106.04
Lower secondary	66.36	75.66	82.43	84.49	85.98	87.48	89.21
Upper secondary	40.30	44.54	48.99	53.07	57.05	61.27	64.70
Net enrollment rate							
Primary	85.57	89.94	92.44	91.72	90.89	91.07	92.20
Secondary	47.46	54.09	58.72	61.22	63.01	66.33	68.41

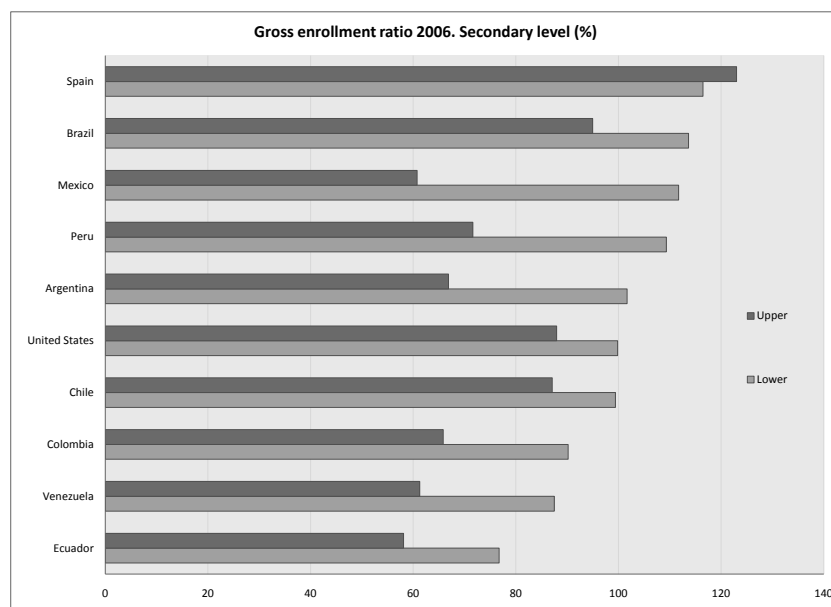
Figure 1.3: Gross enrollment rate. Lower and upper secondary school (%). Source: UNESCO

Table 1.4: Status of high school-age children (as % of total high school-age children). Source: Author's calculations based on the national household survey (*EHM*)

What do HS-age children do?	2005-1	2005-2	2006-1	2006-2	2007-1	2007-2
Study	88.50	88.49.05	88.02	86.85	89.11	87.43
Work	5.43	5.55	5.84	6.07	5.37	5.95
Nothing	3.94	3.57	3.81	4.00	3.14	3.65
Study&Work	2.13	2.39	2.34	3.08	2.39	2.96

Table 1.5: Share of early high school leavers. Source: Author's calculations based on EHM

HS Dropout	2005-1	2005-2	2006-1	2006-2	2007-1	2007-2
No	90.63	90.88	90.35	89.93	91.50	90.40
Yes	9.37	9.12	9.65	10.07	8.50	9.60

Figure 1.4: High school dropouts by type of activity and by semester (as % of total dropouts). Source: Author's calculations based on EHM

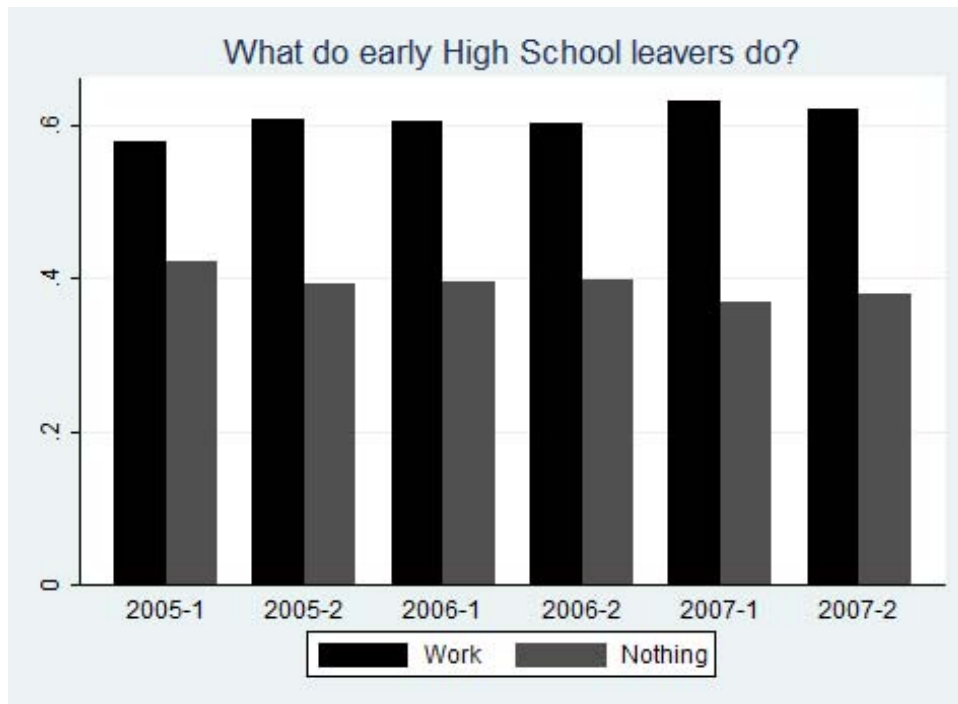


Figure 1.5: High school dropout rate across age and by semester. Source: Author's calculations based on EHM

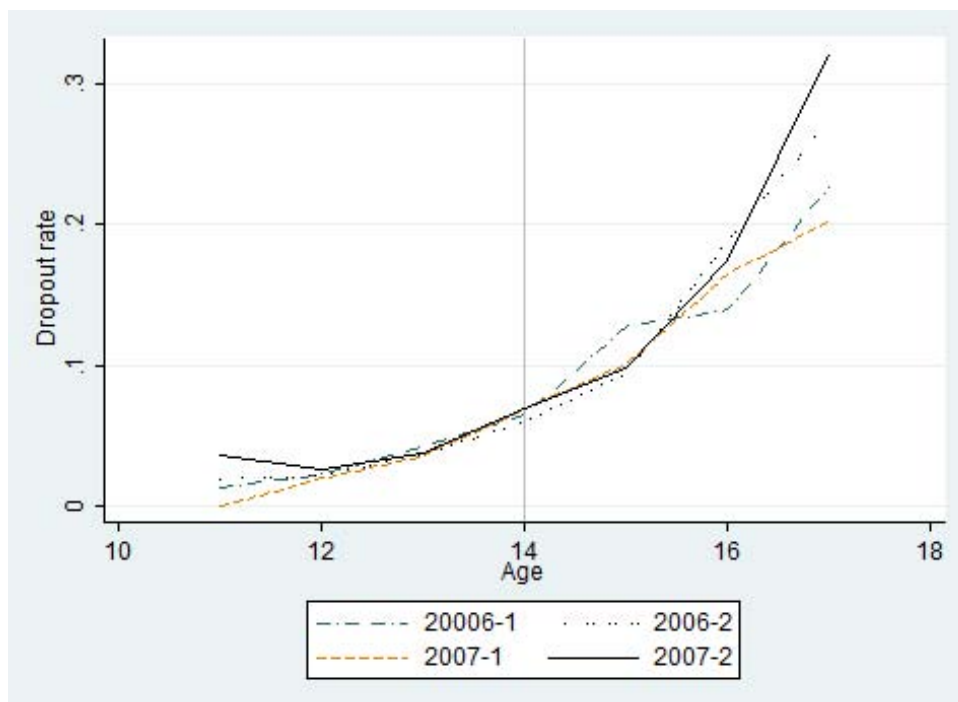
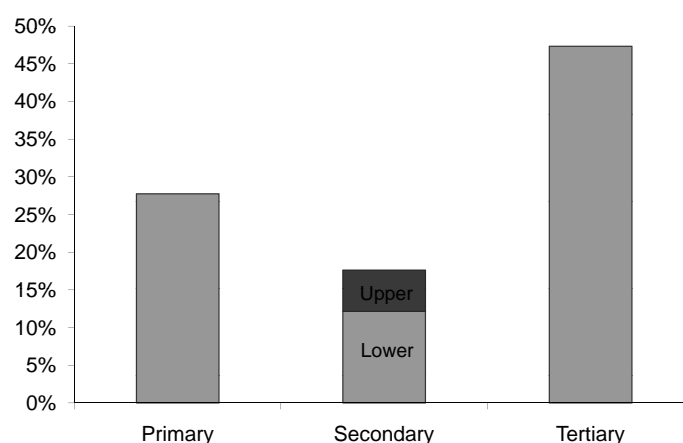


Table 1.6: Reasons for not attending school (as % of total dropouts, 2005-2007 average). Source: Author's calculations based on Encuesta de Hogares por Muestreo (EHM)

Why did they leave school?	Percent
Do not want to study	37.45
Work	17.46
Handicap	5.49
No vacancies	5.42
Cannot afford	4.53
Finished school	4.35
Just married or pregnant	2.84
Behavioral/physical problems	1.85
Housekeeping	1.77
Attending training	1.32
Moved	0.96
No further grades	0.71
Overaged	0.14
Other	15.70

Figure 1.6: Education expenditure in primary, secondary, and tertiary level as % of total education expenditure (%) in 2006. Source: UNESCO



Chapter 2

Errors in Variables in Pseudo-Panel Data: A Simple Correction

2.1 Introduction

Many studies in applied microeconometrics are carried out using cross-sectional data since panel data are often not available. This is especially true in the case of developing countries. This situation translates into a significant loss of information and results in the usual biases stemming from uncontrolled for individual heterogeneity. In response to this, Deaton [1985] suggested a way of taking advantage of the information available in a repeated series of cross-sections. His approach consists in aggregating groups of individuals (cohorts) sharing common time-invariant characteristics. While individuals cannot be followed over time, cohorts are by construction easily observed, which allows one to estimate a model with cohort-specific effects.

A number of papers have examined the consistency of estimates using such “pseudo-panel” data, with a particular focus on the ensuing errors in variables (EV) problem (e.g., McKenzie, 2004, Verbeek and Nijman, 1993, Moffitt, 1993). The EV problem arises from the construction of the cohorts, given that a finite number of observations are available in the sample and each pseudo-individual or cohort is measured with error in relation to the true cohort population values.

This study proposes an alternative to the standard measurement error model approach to pseudo-panels that has received *no* attention in the literature: the use of consistent instrumental variables estimators based on sample moments of order higher than two (early work on this topic includes Geary [2002], Drion [1951], Pal [1980]). Here, we propose a simple solution based on the higher moments estimator of Dagenais and Dagenais [1997].

The rest of the chapter is organized as follows. In the section 2, we present the different solutions to the EV problem that arises when estimating over pseudo-panel data. We begin with the basic notation of Verbeek and Nijman’s errors-in-variables estimator, as well as Deaton’s estimator. We then introduce Dagenais and Dagenais’s estimator as an alternative method. In section 3, Montecarlo experiments are carried out in order to compare the performance of these estimators in terms of bias and mean-squared error (MSE), under various assumptions. Section 4 applies these different approaches to the estimation of a Mincerian wage equation

using Venezuelan data. Section 5 concludes.

2.2 Estimation strategy

2.2.1 Basic notation

We adopt the notation of Verbeek and Nijman [1993].¹ Consider the following panel regression where $i = 1, \dots, N$ indexes individuals and $t = 1, \dots, T$ indexes time periods:

$$y_{it} = x_{it}\beta + \theta_i + \varepsilon_{it}. \quad (2.1)$$

Our purpose is to consistently estimate the $r \times 1$ vector of parameters β . We assume that the individual effects θ_i are correlated with the right-hand-side (RHS) explanatory variables x_{it} :

$$E[x'_{it}\theta_i] \neq 0. \quad (2.2)$$

Given that one cannot follow individuals over time in a time series of independent cross-sections, one constructs C “pseudo-individuals” or cohorts, indexed by c , based on some time-invariant characteristic denoted by z_i . In order to simplify notation, we assume that all cohorts are of equal size $n_c = N/C$ in each period. Aggregating over cohorts, and letting $\bar{x}_{ct} = \frac{1}{n_c} \sum_{i \in c} x_{it}$, the model in terms of observed cohort aggregates can be written as:

$$\bar{y}_{ct} = \bar{x}_{ct}\beta + \bar{\theta}_{ct} + \bar{\varepsilon}_{ct}. \quad (2.3)$$

Since x_{it} and θ_i are correlated, it follows that $E[\bar{x}'_{ct}\bar{\theta}_{ct}] \neq 0$ as well. In contrast to the model based on individual observations, however, $\bar{\theta}_{ct}$ is *not* time-invariant. This implies that the within estimator applied to (2.3) will, in general, be inconsistent.

As in Verbeek and Nijman [1993], consider the cohort population version of (2.3):

$$y_{ct}^* = x_{ct}^*\beta + \theta_c + \varepsilon_{ct}^*, \quad (2.4)$$

and where one assumes the following structure on the measurement errors:

$$\begin{pmatrix} u_{vt} \\ v_{vt} \end{pmatrix} = \begin{pmatrix} \bar{y}_{ct} - y_{ct}^* \\ \bar{x}_{ct} - x_{ct}^* \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_{00} & \sigma' \\ \sigma & \Omega_2 \end{pmatrix} \right). \quad (2.5)$$

Let the variables in the model, expressed in terms of deviation with respect to the within-cohort means, be given by $\tilde{x}_{ct} = \bar{x}_{ct} - \frac{1}{T} \sum_{s=1}^T \bar{x}_{cs} = \bar{x}_{ct} - \bar{x}_c$ (and similarly for \tilde{y}_{ct}), and consider the following empirical moments:

$$M_{xx} = \frac{1}{CT} \sum_{c=1}^C \sum_{t=1}^T \tilde{x}_{ct}' \tilde{x}_{ct}, m_{xy} = \frac{1}{CT} \sum_{c=1}^C \sum_{t=1}^T \tilde{x}_{ct}' \tilde{y}_{ct}. \quad (2.6)$$

Then the class of estimators proposed by Verbeek and Nijman [1993], and indexed by the

¹ See also Verbeek and Nijman [1992]. Note that Devereux [2006] suggests using bias-adjusted instrumental variables estimators, but confines his attention to Jackknife Instrumental Variables (JIVE).

parameter α , is given by:

$$\tilde{\beta}(\alpha) = (M_{xx} - \alpha\Omega_2)^{-1} (m_{xy} - \alpha\sigma). \quad (2.7)$$

Performing a simple within estimation on the pseudo-individuals corresponds to $\alpha = 0$, whereas setting $\alpha = 1$ corresponds to Deaton [1985]'s estimator. Letting $\tilde{x}_{ct}^* = x_{ct}^* - \frac{1}{T} \sum_{s=1}^T x_{cs}^* = x_{ct}^* - x_c^*$, and posing:

$$\Omega_1 = \text{p} \lim_{C \rightarrow \infty} \frac{1}{CT} \sum_{c=1}^C \sum_{t=1}^T \tilde{x}_{ct}^* \tilde{x}_{ct}^{*'}, \Omega_2 = \text{p} \lim_{C \rightarrow \infty} \frac{1}{C(T-1)} \sum_{c=1}^C \sum_{t=1}^T v_{vt} v_{vt}', \quad (2.8)$$

it follows that:

$$\text{p} \lim_{C \rightarrow \infty} \tilde{\beta}(\alpha) = \left[\Omega_1 - \left(\frac{T-1}{T} - \alpha \right) \Omega_2 \right]^{-1} \left[\Omega_1 \beta - \left(\frac{T-1}{T} - \alpha \right) \sigma \right]. \quad (2.9)$$

Verbeek and Nijman [1993] then show how to optimally choose the parameter α in order to minimize mean squared error (MSE).

2.2.2 An instrumental variables approach

This work proposes an alternative to the standard measurement error model approach to pseudo panels sketched above that has received *no* attention in the literature: the use of consistent instrumental variables estimators based on sample moments of order higher than two [Pal, 1980, Drion, 1951, Geary, 2002].

The problem with these estimators is that their behavior is substantially more erratic than the corresponding least squares estimators (see e.g., Malinvaud, 1978, Kendall and Stuart, 1963). One possible solution to this endemic instability is to use a higher moments estimator suggested by Dagenais and Dagenais [1997] which is essentially a linear matrix-weighted combination of third and fourth moment estimators. Additional developments of this approach are provided in Lewbel [1997]. As Dagenais and Dagenais [1997] point out, all higher moments estimators can be considered as a special type of IV estimators where the instruments are given by functions of the original variables raised to some power.

The estimator used here is inspired by Dagenais and Dagenais [1997], where the matrix of feasible instruments, denoted by $Z_{ct} = (z_{1ct}, z_{4ct})$, is given by:

$$z_{1ct} = \tilde{x}_{ct} * \tilde{x}_{ct}, \quad (2.10)$$

$$z_{4ct} = \tilde{x}_{ct} * \tilde{x}_{ct} * \tilde{x}_{ct} - 3\tilde{x}_{ct} \left(\frac{\tilde{x}_{ct}' \tilde{x}_{ct}}{N} * I_r \right), \quad (2.11)$$

and where the symbol $*$ designates the Hadamard element-by-element matrix multiplication operator, I_r is an r -dimensional identity matrix, and ι_r is an $r \times 1$ vector of ones. Detailed proofs of the orthogonality of these instruments with respect to the disturbance term are provided in Dagenais and Dagenais [1997].²

²We maintain the (z_{1ct}, z_{4ct}) notation of Dagenais and Dagenais. In their paper, a broader set of seven higher moment IVs is also considered. Their Montecarlo evidence, as well as our own, shows that it is preferable to

Our preferred implementation of this instrument set is given by a Fuller [1977] modified IV estimator, with the “Fuller constant” set equal to 1. This estimator possesses finite moments for all values of the “concentration parameter” associated with the reduced forms, as well as good small sample properties. Moreover, Hahn, Hausman, and Kuersteiner [2004] have provided extensive Montecarlo experiment results that show that this estimator performs well when compared to other prominent IV estimators, under weak instruments, an issue we shall address below.

The resulting “higher moments” estimator is consistent when there are EV and is also much less erratic than other estimators based on sample moments of order higher than two heretofore suggested in the literature. Note that various implementations of the proposed instrument set are possible, apart from the Fuller-estimator chosen here. These include GMM (the road taken in an earlier paper by Dagenais and Dagenais, 1994), Nagar (or bias-adjusted 2SLS, see Donald and Newey, 2001), or general k -class estimation. In the Montecarlo experiment results that follow, we report results based both on standard 2SLS and the Fuller approach.

2.3 Montecarlo evidence

Montecarlo simulations are carried out in order to evaluate the asymptotic properties of the Verbeek and Nijman [1993] and Fuller [1977] estimators and the higher moments IV estimator based on Dagenais and Dagenais [1997] using pseudo-panel data, as a function of the number of individuals in each cohort n_c and the number of cohorts C .

Individual panel data is generated using the following simplified model:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 z_i + \beta_3 d_{it} + \theta_i + \varepsilon_{it}, t = 1, \dots, T. \quad (2.12)$$

The explanatory variable x_{it} is generated as follows:

$$x_{it} = m_t + z_i + \xi_{it}. \quad (2.13)$$

The first component of the right-hand-side (RHS) of equation (2.13), m_t , is a uniformly distributed variable and identical for all individuals in a given time period t . The second component is a time-invariant variable z_i , which is used as the cohort-identifying variable. For the sake of simplicity, z_i is assumed to be distributed according to the uniform density, with 0 mean and standard deviation equal to 1, such that individuals are split into C intervals, each with equal probability mass. We shall assume that z_i is correlated with the time-varying explanatory variable x_{it} . Let d_{it} be a second regressor representing the control variables which is drawn from the standard normal distribution. The third component is given by $\xi_{it} \sim N(0, 10)$, with $E[\xi_{it}|z_i] = 0$.

Our second assumption relates to the existence of correlation between the individual effects θ_i and the explanatory variable x_{it} . Let $w_i \sim N(0, 1)$. Then we construct the individual effects as:

$$\theta_i = x_{i\bullet} + w_i, \quad (2.14)$$

restrict one's attention to the more limited set of higher moment IVs.

where $E[w_i|z_i] = 0$ for all $t = 1, \dots, T$ and $x_{i\bullet} = \frac{1}{T}(x_{i1} + \dots + x_{iT})$.

Our previously generated panel data is then used to construct a pseudo-panel. We now assume that individuals are not the same over time. Note that equation (2.14) is no longer applicable since individuals change over time and individual effects are not constant. Consequently, a time-varying variable v_{ct} is created, which corresponds to aggregated individual effects. We assume that $v_{ct} \sim N(0, 10)$, and that it is correlated with x_{it} in the following way:

$$\rho[\xi_{it}, v_{ct}] = 0.7. \quad (2.15)$$

The results of the Montecarlo experiments for the various estimators are presented in Tables 2.1 to 2.5. We consider:

1. within estimation on the true panel data,
2. within estimation on the pseudo-panel (grouped) data, with no correction for the EV problem,
3. within estimation on the pseudo-panel (grouped) data, using Fuller approach to correct for the EV problem,
4. the higher moments IV estimator applied to pseudo-panel data, estimated by two-stage least squares (2SLS), with no correction for cohort-specific effects (i.e. a pooling estimator),
5. the higher moments IV estimator applied to pseudo-panel data, estimated by two-stage least squares (2SLS), and controlling for cohort-specific effects,
6. the Fuller version of the higher moments IV estimator applied to pseudo-panel data (with cohort-specific effects), and
7. Verbeek and Nijman's correction combined with the higher moments IVs (with cohort-specific effects).

In these experiment, we use 1,000 replications and set $\beta_1 = 1$. In comparing the performance of the different estimators, it is important to note that we generally keep n_c fixed (instead of N), while modifying T or C .

Given our experimental design in which $\rho[\xi_{it}, v_{ct}] > 0$, we find as expected that all estimators applied to pseudo-panel data are biased upward and perform worse than within estimation over the true panel data. This is due to the individual effects not being completely eliminated when estimating using the pseudo-panel data and to the EV problem that arises as a result of the construction of the cohorts. As N or n_c increases, *ceteris paribus*, the bias of the estimators decreases. This pattern is constant across all of our specifications. For a given n_c , the bias of all estimators usually decreases as the number of time periods T increases, though, when n_c is small, this result does not always hold. For the higher moments estimator (be it in 2SLS or Fuller form), and for a fixed combination of n_c and T , the magnitude of the bias decreases as the number of cohorts C increases (this is apparent in the results presented in Tables 2.1, 2.2

and 2.3). The performance of the other estimators does not always improve in terms of bias as C increases.

If we assume a negative correlation between the aggregated individual effects v_c and x_{it} ($\rho[\xi_{it}, v_{ct}] = -0.7$), the estimators will be biased downward (see Table 2.5). We also analyze the case in which the correlation is relatively weak: $\rho[\xi_{it}, v_{ct}] = 0.2$ (see Table 2.5). In this scenario, the magnitude of bias is, as expected, lower than in the case where $\rho[\xi_{it}, v_{ct}] = 0.7$. In short, the bias of all estimators decreases as the correlation, either negative or positive, decreases.

The upshot of these Montecarlo experiments is that the higher moments IV estimator with cohort-specific effects generally performs better than the others in term of bias when using pseudo-panel data. Moreover, this estimator presents the smallest bias when the correlation is relatively strong (see Table 2.5). On the other hand, the Fuller implementation of the higher moment IVs, as well as accounting for the Verbeek and Nijman's correction yields a smaller MSE (the second result being what one would expect, based on the results in Verbeek and Nijman, 1993). Our preferred estimators are therefore the Dagenais and Dagenais higher moments IV estimator estimated by 2SLS with cohort-specific effects, and its Fuller implementation since they both combine a relatively low level of bias with a small MSE.

2.4 An application to a large Venezuelan pseudo-panel

2.4.1 Data

The source of our data is the *Encuesta de Hogares por Muestreo* (EHM) from the Venezuelan national institute of statistics (INE). This survey contains information about the structure and the evolution of the labor market, as well as the socioeconomic characteristics of individuals. We consider 11 years of this survey, covering the 1994-2004 period, using every second half-year wave. The sample is nationally representative, with the exception of the indigenous population (less than 1% of the total), the inhabitants of small islands and people who are imprisoned, in hospitals or living in hotels. Our sample is restricted to workers between 15 and 67 years of age.³ Households are not followed over time, given that one sixth of the sample is replaced every semester.

We transformed the individual data into pseudo-individuals by constructing cohorts that we then follow over the 11 years of the surveys. Grouping into cohorts is based upon two time-invariant variables: date of birth and region of residence. In the pseudo-panel literature it is very common to find applications in which educational attainment is used to construct cohorts. However in the present context, in which our purpose is to estimate the returns to education, the use of such a grouping variable would be inappropriate in that it would needlessly reduce the variance of our key explanatory variable, thereby rendering identification more difficult.

We construct eight cohorts based on an individual's date of birth using five-year intervals (the exception being the first cohort), namely 1937-1944, 1945-1949, 1950-1954, 1955-1959, 1960-1964, 1965-1969, 1970-1974 and 1975-1979. Similarly, we have six regional-cohorts defined

³We include both females and males given that an Oaxaca decomposition that we carried out did not reveal any significant discrimination against females. These results are, of course, available upon request.

as functions of the geographic-economic division of country that is illustrated in Table 2.6 and in Figure 2.1. Overall, the number of groups, or pseudo-individuals, per year is 48 (8 date of birth cohorts \times 6 regions) as illustrated in Table 2.7. The total number of observations is thus 528 (48 groups \times 11 periods).

The size of our date of birth/region cohorts varies between 35 and 3013 individuals, with 685 individuals on average. According to McKenzie [2004], the size of cohorts should be greater than 100 individuals in order to ensure consistent estimates using pseudo-panel techniques. In our grouped dataset, only 4.6% of the cohorts contain less than 100 individuals. Table 2.8 summarizes the main characteristics of our pseudo-individuals.

Note that, in constructing our cohorts, we weight each observation by its representative proportion in the entire population: this goes a long way towards dealing with any heteroskedasticity issues that might arise as a consequences of our grouping procedure.⁴

A typical example of applying these techniques to data is provided by Gardes et al. [2002], who suggest that cross-sectional estimates, based on a pseudo-panel specification, are less biased than estimates based on individual data. They analyze the bias in the estimated expenditure-elasticity of food consumption. Using Polish and American panel data, they construct six cohorts based on each individual's age at ten years interval and three cohorts based on the level of schooling (incomplete secondary schooling, completed secondary schooling, and a completed university degree). Though their average cohort size is 65 individuals, which is insufficient according to the rules of thumb proposed by McKenzie [2004], their contribution lies in the comparison among different estimators based on panel and pseudo-panel data. Interestingly, their findings are similar whether based on a panel or pseudo-panel specification.

2.4.2 Specification and existing estimates

We estimate a standard Mincerian wage equation [Mincer, 1974] given by:

$$\ln w_{ct} = a + \beta_1 S_{ct} + \beta_2 Exp_{ct} + \beta_3 Exp_{ct}^2 + \beta_4 Controls_{ct} + \varepsilon_{ct} \quad (2.16)$$

where $c = 1, \dots, C$ indexes cohorts, $t = 1, \dots, T$ indexes time periods, $\ln w_{ct}$ is the natural logarithm of hourly earnings of each pseudo-individual c in period t , deflated by the consumer price index (base 1997), S_{ct} is the number of years of schooling, Exp_{ct} represents the years of experience, and $Controls_{ct}$ is a vector of control variables. The coefficient β_1 can be interpreted as the private rate of return to one additional year of schooling, averaged across all pseudo-individuals, at all levels of education and for the entire time period. In what follows, we abstract from issues of non-linearity in the returns to education, which have been analyzed in the Venezuelan context by Patrinos and Sakellariou [2006] using quantile regression techniques in a cross-sectional framework.

Research on the returns to education in Venezuela is scant. Given that no panel datasets are currently available, existing studies are restricted to cross-sectional techniques. Ortega [2003] examines the determinants of income inequality in Venezuela while focusing on the role played by education, and includes estimates of the returns to education in his analysis. Contrary to

⁴An alternative approach, pursued by Gardes, Duncan, Gaubert, and Starzec [2002], is to estimate over the pseudo-individuals while weighting by the relative size of cohorts in the total sample.

the specification given in equation (2.16), he restricts his explanatory variables to educational attainment and experience, and neither includes additional controls nor considers issues of unobserved heterogeneity and measurement error. Using the same EHM data as is used here (though restricted to the 1995-2002 period), his cross-section regressions yield estimates of the returns to education that lie between 8% and 10%.

Patrinos and Sakellariou [2005], again using the EHM surveys in a cross-sectional framework, provide instrumental variables estimates of the returns to education. The authors restrict their sample to individuals between ages 15 and 65 who have a household head in their family, so as to be able to use parental education as an excluded instrument. Given that this IV is probably correlated with unobservable ability, the authors then focus on a supply-side instrument given by the natural experiment furnished by a change in compulsory school attendance laws in 1980 (compulsory schooling went from 6 to 9 years). Their preferred point estimate for the returns to education is 12% in 2000.

Patrinos and Sakellariou [2006] use quantile regression methods in order to estimate heterogeneity in the returns to education while abstracting from unobservable ability or measurement error problems. For males, they show that the returns to education increase as one moves up the income quantiles. For females there is no clear pattern, although their results suggest that the phenomenon is reversed. The returns to education vary between 6.9% and 13.8%. They argue that the pattern of the returns to education for males is explained by complementarity between unobserved ability and educational attainment. As such, it is difficult to take their point estimates as being unbiased.

2.4.3 Results

Recall that standard pseudo-panel techniques, such as Verbeek and Nijman [1993], attempt to correct for the EV problem by eliminating a proportion α of the measurement error from the empirical moments, whereas our alternative approach uses the higher moments instrumental variables estimator in order to handle the same EV problem. Our results are presented in Table 2.9, and largely mirror the patterns uncovered in our Montecarlo experiments.

On the basis of our Montecarlo experiments, and considering the results presented in Table 2.9, it would appear reasonable to infer that the correlation between the aggregated individual effects and the educational attainment variable is positive ($\rho > 0$). This is because the point estimate obtained by applying the within estimator to the pseudo-panel data (column 1) is larger —16.6%— than the estimates based on our IV procedure (columns 3, 4 and 5). In our Montecarlo experiments, we found that the within estimator generally yielded significantly greater upward bias than our IV alternatives.⁵

Our preferred specification, on the basis of our Montecarlo results, is given by the results presented in columns 4 and 5. We find that the higher moments IV estimator yields a point estimate for the returns to education that is equal to 15.5%. The validity of our IV-based approach is confirmed both by the results of the Sargan tests of the overidentifying restrictions,

⁵ Note that the higher moments estimator applied to pooled pseudo-panel data without controlling for cohort-specific effects (column 2) yields a point estimate of the returns to education that is significantly lower than in all other cases and it is not statistically significant. We are currently seeking an explanation for this result, which did not emerge in our Montecarlo experiments.

as well as by the Hahn and Hausman [2002] test of the joint null of instrument orthogonality and instrument strength, which never reject.⁶ In Table 2.9, we also present the coefficients associated with the higher moment instruments from the first-stage reduced forms corresponding to education attainment. As should be obvious, and in conformity with the results of the Hahn-Hausman test, there is no evidence for instrument weakness.

The upshot of our results is that, once ability bias and the EV issue have been addressed using our IV-based approach, the returns to education for Venezuelan workers have been significantly underestimated, by a factor of at least 30%, when derived by means of cross-section regression methods. As expected, the experience has an inverted-U relationship with the salary. It means that an additional year of working experience increases the logarithm of the salary, on average, by 6%. When an average worker is over 40 years-old, the impact of an additional year of experience on the salary becomes negative yet small (on average, -0.09%).

For the rest of the control variables, we find that if the worker is the household head his or her salary tend to be lower than other members of the family. One explanation may be that, after controlling for experience, gender, and marital status, this variable is probably picking up some “type of work” effect. For instance, those household heads who work on low skilled jobs such as a seamstress or a taxi driver. Females are likely to earn less than males, however the effect is not statistically significant.⁷ Patrinos and Sakellariou [2006] show that, for most quantiles regressions, returns to education are higher for females than for males. Yet the authors do not perform any test to check if the difference is statistically significant. Married workers seem to receive a higher salary than those who are single, cohabiting or divorced. This result confirms what it is largely demonstrated in other empirical studies. According to Chun and Lee [2001], Neumark and Korenman [1988], once endogeneity is accounted for, marriage enhances men’s labor market productivity.

2.5 Concluding remarks

In this chapter, we have studied the performance of various estimators using pseudo-panel data stemming from a time series of repeated cross-sections. In contrast to the standard approach to the EV problem generated by the construction of pseudo-individuals, and in which a proportion of the measurement errors from the empirical moments is eliminated, as in Deaton [1985] and Verbeek and Nijman [1993], we have considered an alternative approach based on an instrumental variables estimator. Our estimator presents the added advantage that the underlying orthogonality conditions can be assessed using standard test statistics, an option that is not available using traditional approaches.

Our Montecarlo results demonstrate that our proposed higher moments estimator performs better than other available estimators, as illustrated by its lower level of bias. In particular, the reduction in bias associated with our estimator increases as the degree of correlation between the aggregated individual effects and the explanatory variable rises. In general, as sample or cohort size increases, the bias of our estimators decreases. Concomitantly for a given cohort

⁶See also Hahn et al. [2004] for Montecarlo evidence on the performance of the test.

⁷We also perform the Oaxaca decomposition to check for a wage gap by sex (not shown). The result shows that although female’s earnings are lower than male’s, gender discrimination is not statistically significant.

size, the bias of our estimators decreases as the number of time periods or the number of cohorts increases.

Our empirical illustration, based on a large Venezuelan pseudo-panel, demonstrates that the returns to education for Venezuelan workers have been significantly underestimated in other studies. Once ability bias and the errors in variables problem are accounted for using our estimator, we find that an additional year of schooling is associated with an increase in wages of 15.5%.

Aside from the methodological contribution, our empirical results suggest that improvements in educational attainment may have been neglected as a route towards greater income equality in Venezuela. Future research will consider means of dealing with sample selection bias as well as potential non-linearities in the returns to education within a pseudo-panel framework.

REFERENCES

- H. Chun and I. Lee. Why do married men earn more: Productivity or marriage selection? *Economic Inquiry*, 39(2):307–19, April 2001.
- M. Dagenais and D. Dagenais. Gmm estimators for linear regression models with errors in the variables. Cahiers de recherche 9404, Centre interuniversitaire de recherche en économie quantitative, CIREQ, 1994. URL <http://ideas.repec.org/p/mtl/montec/9404.html>.
- M. G. Dagenais and D. L. Dagenais. Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics*, 76(1-2):193–221, 1997.
- A. Deaton. Panel data from time series of cross-sections. *Journal of Econometrics*, 30(1-2):109–126, 1985.
- P. J. Devereux. Improved errors-in-variables estimators for grouped data. Working Papers 200602, School Of Economics, University College Dublin, Jan. 2006. URL <http://ideas.repec.org/p/ucn/wpaper/200602.html>.
- S. G. Donald and W. K. Newey. Choosing the number of instruments. *Econometrica*, 69(5):1161–91, September 2001.
- E. F. Drion. Estimation of the parameters of a straight line and of the variance of the variables, if they are both subject to error. *Indagationes Mathematicae*, 13:256–260, 1951.
- W. A. Fuller. Some properties of a modification of the limited information estimator. *Econometrica*, 45(4):939–53, May 1977.
- F. Gardes, G. J. Duncan, P. Gaubert, and C. Starzec. Panel and pseudo-panel estimation of cross-sectional and time series elasticities of food consumption : The case of american and polish data. Working Papers 2002-02, Centre de Recherche en Economie et Statistique, 2002. URL <http://ideas.repec.org/p/crs/wpaper/2002-02.html>.
- R. C. Geary. Inherent relations between random variables. Technical Report 47, Proceedings of the Royal Irish Academy, 2002.
- J. Hahn and J. Hausman. A new specification test for the validity of instrumental variables. *Econometrica*, 70(1):163–189, January 2002.
- J. Hahn, J. Hausman, and G. Kuersteiner. Estimation with weak instruments: Accuracy of higher-order bias and mse approximations. *Econometrics Journal*, 7(1):272–306, June 2004.
- M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*. Charles Griffin and Company, London, UK, second edition, 1963.
- A. Lewbel. Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and r&d. *Econometrica*, 65(5):1201–1214, September 1997.

- E. Malinvaud. *Méthodes Statistiques de l'Econométrie*. Dunod, Paris, France, third edition, 1978.
- D. McKenzie. Asymptotic theory for heterogeneous dynamic pseudo-panels. *Journal of Econometrics*, 120(2):235–262, June 2004.
- J. Mincer. *Schooling, Experience and Earnings*. Columbia University Press for NBER, New York, NY, 1974.
- R. Moffitt. Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics*, 59(1-2):99–123, September 1993.
- D. Neumark and S. D. Korenman. Does marriage really make men more productive? Finance and Economics Discussion Series 29, Board of Governors of the Federal Reserve System (U.S.), 1988. URL <http://ideas.repec.org/p/fip/fedgfe/29.html>.
- D. Ortega. *Descripción y Perfiles de Desigualdad de Ingresos en Venezuela, 1975-2002. Documento para la Discusión, Programa de las Naciones Unidas para el Desarrollo*, 2003.
- M. Pal. Consistent moment estimators of regression coefficients in the presence of errors in variables. *Journal of Econometrics*, 14(3):349–364, December 1980.
- H. Patrinos and C. Sakellariou. Schooling and labor market impacts of a natural policy experiment. *LABOUR*, 19(4):705–719, December 2005.
- H. Patrinos and C. Sakellariou. Economic volatility and returns to education in venezuela: 1992-2002. *Applied Economics*, 38(17):1991–2005, 2006.
- M. Verbeek and T. Nijman. Can cohort data be treated as genuine panel data? *Empirical Economics*, 17(1):9–23, 1992.
- M. Verbeek and T. Nijman. Minimum mse estimation of a regression model with fixed effects from a series of cross-sections. *Journal of Econometrics*, 59(1-2):125–136, September 1993.

2.6 Tables and Figures

Figure 2.1: Geographic division of Venezuelan territory by regions.
Source: made by the author

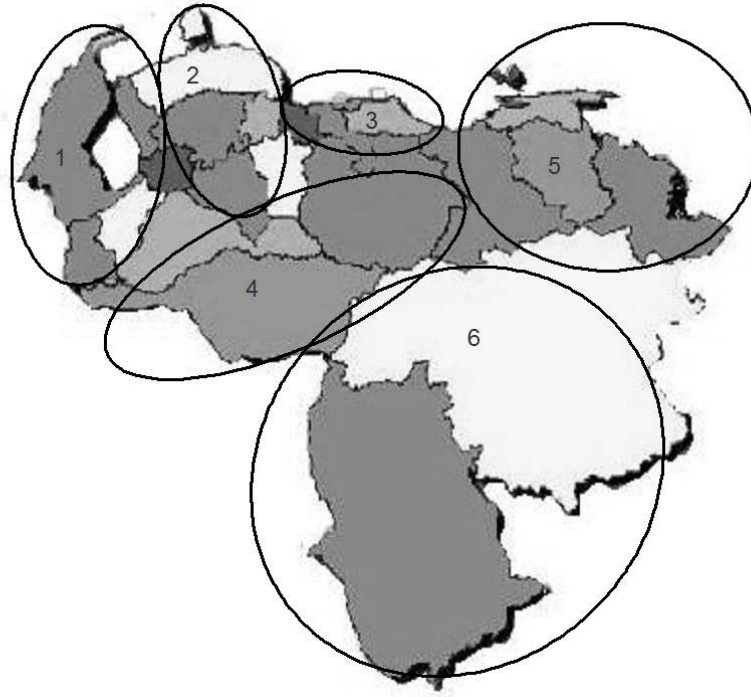


Table 2.1: Montecarlo experiment results: a comparison of the bias and mean-squared error (MSE) of five different estimators as applied to pseudo-panel data. $T = 5, C = 8, \rho = 0.7$.

Estimator	Panel data	Pseudo-panel data					
	Within	Within	Within by Fuller	Dagenais higher moment IVs			
				by 2SLS (pooling)	by 2SLS	by Fuller	with Verbeek -Nijman correction
Cohort-specific effects		yes	yes	no	yes	yes	yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$N = 160, n_c = 20$							
Bias	0.000	0.592	0.592	0.575	0.626	0.603	0.592
MSE	0.250	1.178	1.660	2.098	2.066	1.812	1.680
$N = 400, n_c = 50$							
Bias	0.000	0.485	0.485	0.476	0.457	0.466	0.478
MSE	0.250	1.135	1.099	1.362	1.298	1.193	1.111
$N = 800, n_c = 100$							
Bias	0.000	0.382	0.382	0.285	0.345	0.336	0.361
MSE	0.250	0.852	0.825	1.058	0.987	0.894	0.834
$N = 1600, n_c = 200$							
Bias	0.000	0.279	0.279	0.184	0.223	0.217	0.250
MSE	0.250	0.630	0.610	0.758	0.692	0.657	0.618
$N = 3200, n_c = 400$							
Bias	0.000	0.185	0.185	0.089	0.108	0.119	0.150
MSE	0.250	0.464	0.449	0.547	0.493	0.479	0.455
$N = 6000, n_c = 750$							
Bias	0.000	0.120	0.120	0.032	0.060	0.065	0.090
MSE	0.250	0.346	0.334	0.405	0.360	0.352	0.339

Table 2.2: Montecarlo experiment results: a comparison of the bias and mean-squared error (MSE) of five different estimators as applied to pseudo-panel data. $T = 5, C = 30, \rho = 0.7$.

Estimator	Panel data	Pseudo-panel data					
	Within	Within	Within	Dagenais higher moment IVs			
			by Fuller	by 2SLS (pooling)	by 2SLS	by Fuller	with Verbeek -Nijman correction
Cohort-specific effects		yes	yes	no	yes	yes	yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$N = 600, n_c = 20$							
Bias	0.000	0.597	0.597	0.501	0.565	0.587	0.595
MSE	0.250	1.751	1.736	2.290	2.160	1.929	1.740
$N = 1500, n_c = 50$							
Bias	0.000	0.493	0.493	0.445	0.450	0.444	0.482
MSE	0.250	1.152	1.142	1.404	1.478	1.261	1.146
$N = 3000, n_c = 100$							
Bias	0.000	0.387	0.387	0.277	0.260	0.267	0.353
MSE	0.250	0.855	0.848	1.059	1.010	0.942	0.853
$N = 6000, n_c = 200$							
Bias	0.000	0.290	0.290	0.123	0.147	0.150	0.243
MSE	0.250	0.632	0.627	0.749	0.720	0.687	0.633
$N = 12000, n_c = 400$							
Bias	0.000	0.190	0.190	0.052	0.058	0.066	0.139
MSE	0.250	0.466	0.462	0.533	0.505	0.496	0.467
$N = 22500, n_c = 750$							
Bias	0.000	0.120	0.120	0.004	0.027	0.029	0.077
MSE	0.250	0.349	0.346	0.403	0.368	0.364	0.350

Table 2.3: Montecarlo experiment results: a comparison of the bias and mean-squared error (MSE) of five different estimators as applied to pseudo-panel data. $T = 5, C = 80, \rho = 0.7$.

Estimator	Panel data	Pseudo-panel data					
	Within	Within	Within	Dagenais higher moment IVs			
			by Fuller	by 2SLS (pooling)	by 2SLS	by Fuller	with Verbeek -Nijman correction
Cohort-specific effects		yes	yes	no	yes	yes	yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$N = 1600, n_c = 20$							
Bias	0.000	0.597	0.597	0.577	0.554	0.574	0.595
MSE	0.250	1.753	1.747	2.162	2.161	1.961	1.748
$N = 4000, n_c = 50$							
Bias	0.000	0.493	0.493	0.371	0.362	0.370	0.477
MSE	0.250	1.156	1.152	1.436	1.490	1.308	1.154
$N = 8000, n_c = 100$							
Bias	0.000	0.391	0.391	0.205	0.196	0.219	0.357
MSE	0.250	0.855	0.852	1.051	1.035	0.965	0.856
$N = 16000, n_c = 200$							
Bias	0.000	0.282	0.282	0.068	0.082	0.083	0.229
MSE	0.250	0.635	0.633	0.773	0.729	0.708	0.638

Table 2.4: Montecarlo experiment results: a comparison of the bias and mean-squared error (MSE) of five different estimators as applied to pseudo-panel data. $n_c = 100$ and $\rho = 0.7$. N , T and C vary.

Estimator	Panel data	Pseudo-panel data					
	Within	Within	Within	Dagenais higher moment IVs			
			by Fuller	by 2SLS (pooling)	by 2SLS	by Fuller	with Verbeek -Nijman correction
Cohort-specific effects		yes	yes	no	yes	yes	yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$N = 800, C = 8, T = 5$							
Bias	0.000	0.382	0.382	0.285	0.345	0.336	0.361
MSE	0.250	0.852	0.852	1.058	0.987	0.894	0.834
$N = 800, C = 8, T = 10$							
Bias	0.000	0.369	0.369	0.309	0.310	0.309	0.341
MSE	0.250	0.861	0.861	1.064	1.011	0.935	0.862
$N = 800, C = 8, T = 20$							
Bias	0.000	0.355	0.355	0.250	0.271	0.276	0.316
MSE	0.250	0.868	0.868	1.050	1.021	0.939	0.876
$N = 3000, C = 30, T = 5$							
Bias	0.000	0.387	0.387	0.277	0.260	0.267	0.353
MSE	0.250	0.855	0.855	1.059	1.010	0.942	0.853
$N = 3000, C = 30, T = 10$							
Bias	0.000	0.370	0.370	0.234	0.218	0.233	0.327
MSE	0.250	0.863	0.860	1.073	1.030	0.959	0.867
$N = 3000, C = 30, T = 20$							
Bias	0.000	0.361	0.361	0.158	0.155	0.183	0.305
MSE	0.250	0.870	0.869	1.045	1.052	0.968	0.878
$N = 8000, C = 80, T = 5$							
Bias	0.000	0.391	0.391	0.205	0.196	0.219	0.357
MSE	0.250	0.855	0.852	1.051	1.035	0.965	0.856
$N = 8000, C = 80, T = 10$							
Bias	0.000	0.366	0.366	0.112	0.115	0.128	0.310
MSE	0.250	0.866	0.865	1.046	1.042	0.986	0.872
$N = 8000, C = 80, T = 20$							
Bias	0.000	0.362	0.362	0.085	0.086	0.090	0.296
MSE	0.250	0.868	0.867	1.042	1.026	1.005	0.876

Table 2.5: Montecarlo experiment results: a comparison of the bias and mean-squared error (MSE) of five different estimators as applied to pseudo-panel data. $T=10$, $C=30$, $N=1500$, $n_c=50$. ρ varies.

Estimator	Panel data	Pseudo-panel data					
	Within	Within	Within	Dagenais higher moment IVs			
			by Fuller	by 2SLS (pooling)	by 2SLS	by Fuller	with Verbeek -Nijman correction
Cohort-specific effects		yes	yes	no	yes	yes	yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\rho = -0.7$							
Bias	0.000	-0.478	-0.478	-0.417	-0.394	-0.408	-0.461
MSE	0.250	1.163	1.158	1.451	1.451	1.289	1.163
$\rho = -0.5$							
Bias	0.000	-0.342	-0.342	-0.299	-0.280	-0.286	-0.332
MSE	0.250	1.299	1.294	1.559	1.586	1.418	1.298
$\rho = -0.2$							
Bias	0.000	-0.139	-0.139	-0.095	-0.145	-0.117	-0.134
MSE	0.250	1.408	1.403	1.709	1.730	1.527	1.405
$\rho = 0.2$							
Bias	0.000	0.133	0.133	0.139	0.141	0.121	0.130
MSE	0.250	1.410	1.404	1.677	1.686	1.523	1.407
$\rho = 0.5$							
Bias	0.000	0.339	0.339	0.363	0.292	0.292	0.328
MSE	0.250	1.302	1.298	1.702	1.584	1.418	1.301
$\rho = 0.7$							
Bias	0.000	0.476	0.476	0.389	0.383	0.404	0.457
MSE	0.250	1.167	1.163	1.468	1.440	1.288	1.167

Table 2.6: States by region. Source: Plan Nacional de Desarrollo Regional 2001, Ministerio de Desarrollo y Planificacin.

Identifier	Region	States
1	West	Zulia, Táchira, Mérida y Trujillo
2	North west	Falcón, Lara, Portuguesa, Yaracuy y Cojedes
3	North central	Carabobo, Aragua, Miranda, Distrito Capital y Vargas
4	Central	Barinas, Apure y Guárico
5	East	Anzoátegui, Monagas, Sucre, Nueva Esparta y Delta Amacuro
6	South	Amazonas y Bolivar

Table 2.7: Construction of pseudo-individuals. Sample restricted to employees who are between 15 and 65 years of age. Source: authors' calculations based on the 1994 to 2004 EHM household surveys.

Identifier	Region	Date of birth	Pseudo indiv.	Identifier	Region	Date of birth	Pseudo indiv.
1,1	West	01/75-12/79	1	4,1	Central	01/75-12/79	25
1,2	West	01/70-12/74	2	4,2	Central	01/70-12/74	26
1,3	West	01/65-12/69	3	4,3	Central	01/65-12/69	27
1,4	West	01/60-12/64	4	4,4	Central	01/60-12/64	28
1,5	West	01/55-12/59	5	4,5	Central	01/55-12/59	29
1,6	West	01/50-12/54	6	4,6	Central	01/50-12/54	30
1,7	West	01/45-12/49	7	4,7	Central	01/45-12/49	31
1,8	West	01/37-12/44	8	4,8	Central	01/37-12/44	32
2,1	Northwest	01/75-12/79	9	5,1	East	01/75-12/79	33
2,2	Northwest	01/70-12/74	10	5,2	East	01/70-12/74	34
2,3	Northwest	01/65-12/69	11	5,3	East	01/65-12/69	35
2,4	Northwest	01/60-12/64	12	5,4	East	01/60-12/64	36
2,5	Northwest	01/55-12/59	13	5,5	East	01/55-12/59	37
2,6	Northwest	01/50-12/54	14	5,6	East	01/50-12/54	38
2,7	Northwest	01/45-12/49	15	5,7	East	01/45-12/49	39
2,8	Northwest	01/37-12/44	16	5,8	East	01/37-12/44	40
3,1	North central	01/75-12/79	17	6,1	South	01/75-12/79	41
3,2	North central	01/70-12/74	18	6,2	South	01/70-12/74	42
3,3	North central	01/65-12/69	19	6,3	South	01/65-12/69	43
3,4	North central	01/60-12/64	20	6,4	South	01/60-12/64	44
3,5	North central	01/55-12/59	21	6,5	South	01/55-12/59	45
3,6	North central	01/50-12/54	22	6,6	South	01/50-12/54	46
3,7	North central	01/45-12/49	23	6,7	South	01/45-12/49	47
3,8	North central	01/37-12/44	24	6,8	South	01/37-12/44	48

Table 2.8: Descriptive statistics on grouped pseudo-panel data: 528 observations. Source: 1994 to 2004 EHM household surveys.

Variable	Mean	Std. dev.	Min.	Max.
Hourly earnings	729.18	195.43	261.04	2178.59
Log(Hourly earnings)	6.55	0.26	5.56	7.68
Years of education	8.39	1.03	5.21	10.27
Experience	25.18	12.54	3.95	50.84
% household head (HH)	0.52	0.21	0.00	0.86
% females	0.36	0.05	0.11	0.46
% married	0.35	0.12	0.01	0.62
Region index	3.50	1.70	1	6
Date of birth index	4.50	2.29	1	8

Table 2.9: The returns to education in Venezuela, 1994-2004. Pseudo-panel regressions using various estimators. The number of observations = 528. Standard errors in parentheses.

Pseudo panel estimator	Within	Dagenais higher moment IVs			
		by 2SLS (pooling)	Verbeek -Nijman correction	by Fuller	by 2SLS
Cohort-specific effects	yes	no	yes	yes	yes
	(1)	(2)	(3)	(4)	(5)
Education	0.166 (0.022)	0.083 (0.046)	0.158 (0.039)	0.155 (0.045)	0.155 (0.063)
Experience	0.059 (0.014)	0.010 (0.013)	0.058 (0.014)	0.058 (0.014)	0.058 (0.014)
Experience ²	-0.0009 (0.0002)	-0.0001 (0.0001)	-0.0009 (0.0002)	-0.0009 (0.0002)	-0.0009 (0.0002)
% of HH	-0.922 (0.231)	-0.064 (0.207)	-0.914 (0.000)	-0.910 (0.236)	-0.910 (0.236)
Female	-0.287 (0.294)	0.127 (0.549)	-0.252 (0.353)	-0.238 (0.375)	-0.237 (0.376)
Married	0.710 (0.247)	1.211 (0.244)	0.732 (0.275)	0.741 (0.286)	0.742 (0.286)
Coefficients associated with the higher moment IVs in the first stage reduced forms for educational attainment					
z_{1ct}		-0.170 (0.024)	-0.227 (0.020)	-0.227 (0.020)	-0.227 (0.020)
z_{4ct}		-0.051 (0.013)	-0.035 (0.01)	-0.035 (0.01)	-0.035 (0.01)
k -class			0.9		
Sargan test of the OID restrictions: p -value		0.54	0.60	0.61	0.61
Hahn-Hausman m_3 test statistic: p -value					0.52

Chapter 3

High School Dropouts and the Informal Sector

3.1 Introduction

School dropouts and child labor are widely studied subjects in socioeconomic literature [Basu, 1999, Becker, 1991] and continue to be a concern for economists and social scientists [Cardoso and Dorte, 2006, Fafchamps and Wahba, 2006, Rosati and Tzannatos, 2006]. Most papers have shown a high propensity for child labor related to rural activities in less developed countries, e.g., India, Vietnam, and Brazil [Kruger, Soares, and Berthelon, 2007, Fafchamps and Wahba, 2006, Rosati and Tzannatos, 2006, Rosati and Deb, 2004]. The broadly accepted argument is that rural activities need a less qualified labor force; therefore children can easily participate in these types of activities. However, not much has been done with respect to child labor and the informal sector [World Bank, International Labor Organization, and United Nations Children's Fund, 2010, World Bank, 2007b, 2005].

According to Almeida and Carneiro [2005], informal activities are a result of structural economic problems, labor market failures, and lack of law enforcement. The predominant informal sector, particularly in Latin America, has been an area of active policymaking. Even though some mechanisms for eradicating informality have been implemented (e.g., incentives to register micro and small enterprises and microcredits), this issue is still far from being addressed. Informal workers in Venezuela accounted for almost 50% of all workers in 2003 [Gasparini and Tornarolli, 2007]. Generally, the skills needed to work in informal activities are relatively low.¹ Hence, it is easy and sometimes enticing for children to become informal workers. Alternatively, if parents are working in the informal sector, they are more vulnerable to negative income shocks,² and this can force them to keep children out of school [World Bank, 2005]. Thus, the informal economy could negatively affect not only business productivity but also children's educational attainment.

The Venezuelan economy is based mostly on urban activities. In fact, the rural population

¹According to Gasparini and Tornarolli [2007] the share of Venezuelan informal workers is higher in primary sector activities, construction, trade, utilities and transportation, and domestic workers.

²For instance, an informal working parent who is unprotected by the labor law and, suddenly, is dismissed, he will not receive a compensation or severance pay.

is low (7% of the total population in 2007) and has been decreasing over time.³ We believe that the effect of parents working in the informal sector can be compared to that of parents working in farming activities. Fafchamps and Wahba [2006] suggest that in poor countries, households tend to rely on self-employment. Moreover, as children are more productive in the family farm or business than in more market-oriented activities, they may assist the family or do household chores while parents work outside the home. Fafchamps and Wahba [2006] also consider that the activities in which a family engages shape the demand for child work within the household.

In this regard, evidence from Venezuela is scarce. There are two descriptive analyses of child labor in Venezuela as part of the research project called Understanding Children's Work [Blanco, 2009, Blanco and Valdivia, 2006]. The authors argue that due to the major economic downturn and political instability during 2002-2003, more children were sent to work, to some extent, to cope with the negative household income shocks. Poor families tend to cut back on investing in their children's education and either put their children to work or combine work with study. The said research project's authors also mention that the consequences of the crisis may be permanent, given that the incidence of child labor remained higher than it was before the negative shock. Additionally, they find that the most common activities in which children work are in agriculture and the trade sector. Both sectors are characterized by being traditionally informal or by hiring informal employees.

Psacharopoulos [1997] has an interesting paper about child labor and educational attainment in Venezuela. Similar to the results of related empirical papers, the author finds that the probability of being a working child increases with age, with being a male, and having a female household head. However, he also suggests that the decision to work is associated with a higher household income, which is the opposite of what wealth-based models demonstrate [Basu, 1999]. The author shows that contributions of working children to the total household income are very significant (27%). Nevertheless, income growth is not the only factor explaining child labor. According to Basu [1999], changes in technology, access to education, and particularly, conditions in the adult labor market are as important determinants of child labor as household income.

Table 1.6 shows that the main reasons in Venezuela for not attending school are: "*does not want to study*" (37.45%), "*work*" (17.46%), "*handicap*" (5.49%), "*no vacancies*" (5.42%), and "*cannot afford it*" (4.53%). As we can see, the lack of enthusiasm for attending school is possibly a major factor associated with dropouts. This illustrates that factors other than income may also be important when deciding whether to keep children out of school. At the same time, a child's interests may be affected by other factors, not always measured in the survey, e.g., ability, school quality, and access to school.⁴

The contribution of this study is twofold. First, the inclusion of informality (informal worker parents) as one of the main factors affecting the likelihood of children completing high school has not been fully explored in the available literature on the subject. Second, the use of recent panel data in assessing the problem of early high school dropouts, provides new and more robust insight into Venezuela's educational bottlenecks.

³WB indicators database.

⁴However, Table 1.6 may shed light on some of these other factors; for instance, "no vacancies" is a proxy for issues related to access to school.

The rest of the chapter is divided into five sections. First, we present the theoretical model. Second, we explain our empirical strategy. The third and fourth sections refer to the data and results, respectively. Finally, we discuss our conclusions.

3.2 Theoretical Model Review

Among the rich literature about child labor, the models of Basu and Van [1998] and Basu [1999] capture this phenomenon in a straightforward way. Basically, they conceive the household wealth as the determinant of child labor. We focus on the basic model, as well as on two enhanced versions.

In the basic model, the authors' main assumptions are simple and not very restrictive. As the first assumption, keeping children out of work is a luxury item. In other words, they suppose that parents send their children to work when the household's income is low. In the opposite case, parents withdraw their children from the labor force as soon as their income can afford the household's consumption. The household preference relation is a function of (c, e) and is given by:

$$\begin{cases} (c, 0) \succ (c + \delta, 1) & \text{if } c \geq s, \\ (c + \delta, 1) \succ (c, 0) & \text{if } c < s. \end{cases} \quad (3.1)$$

where c is the consumption by each family member and e is the child's work effort, in which 0 means *no work* and 1 *work*. The authors also assume that adults always work, irrespective of wage level⁵ for all $\delta > 0$; the family prefers to keep the child out of work if and only if each individual's consumption is above a certain exogenously fixed subsistence level, s .

Regarding the household's characteristics, the authors assume that there are N identical families, and each one is made up of one adult and one child. They attribute just one utility function to each household, even though they admit this idea abstracts from reality. It means that decisions of child labor-supply are taken by parents. The authors also assume that child and adult consumptions are equal.

The household budget constraint is as follows:

$$2c \leq ew_C + w_A, \quad (3.2)$$

where w_C and w_A are the market wages for child and adult, respectively. Hence, the maximization problem is solved in the following way:

$$c(w_A) = \begin{cases} \frac{w_A}{2} & \text{if } w_A \geq 2s, \\ \frac{w_A + w_C}{2} & \text{if } w_A < 2s. \end{cases} \quad (3.3)$$

$$e(w_A) = \begin{cases} 0 & \text{if } w_A \geq 2s, \\ 1 & \text{if } w_A < 2s. \end{cases} \quad (3.4)$$

⁵A more realistic approach is developed by Basu [1999], in which he takes into account unemployment in the model.

The labor supply is composed of adults, S^A , and children, S^C , and defined by:

$$S^A = N \quad (3.5)$$

$$S^C(w_A) = \begin{cases} 0 & \text{if } w_A \geq 2s, \\ N & \text{if } w_A < 2s. \end{cases} \quad (3.6)$$

The second main assumption is defined as the substitution axiom. It considers that adult labor and child labor are substitutes, subject to an adult-equivalent scaling, denoted by γ , where $0 < \gamma < 1$. It means that a child works a proportion γ of what an adult works in a specific time. The authors assume there are n identical firms with a maximization problem, as follows:

$$\max_{A_i, C_i} f(A_i + \gamma C_i) - A_i w_A + C_i w_C, \quad f' > 0, f'' < 0, \quad (3.7)$$

where A_i and C_i are respectively the numbers of adults and children working for the firm i , and $A_i + \gamma C_i = L_i$ is the “effective labor employment” measured in adult-equivalents. The firm is a wage taker.

Given that firms want to employ the cheapest labor force possible, they will employ adults only if $w_A < w_C/\gamma$. In the opposite case, firms will hire children if $w_A > w_C/\gamma$. The “effective child wage”, w_C/γ , is the market child wage per adult-equivalent. The demand for adults and children are defined by $D^A(w_A, w_C)$ and $D^C(w_A, w_C)$, respectively. Thus, the aggregate demand for adults and children leads to:

$$\begin{cases} w_A > w_C/\gamma, & \text{then } D^A = 0 \text{ and } f'(\frac{\gamma D^C}{n}) = \frac{w_C}{\gamma}, \\ w_A < w_C/\gamma, & \text{then } D^C = 0 \text{ and } f'(\frac{D^A}{n}) = w_A, \\ w_A = w_C/\gamma, & \text{then } f'(\frac{D^A + \gamma D^C}{n}) = w_A = \frac{w_C}{\gamma}. \end{cases} \quad (3.8)$$

The equilibrium in the labor market is depicted by a pair of wages, as follows:

$$\begin{aligned} D^A(w_A^*, w_C^*) &= N \quad \text{and} \\ D^C(w_A^*, w_C^*) &= S^C(w_A^*). \end{aligned} \quad (3.9)$$

In the case of $w_A = \frac{w_C}{\gamma}$, the authors show that it is possible to have multiple equilibria in the model. However, households are better off in the no working-children equilibrium than the one in which children work. Depending on the labor force productivity, the two-equilibria model might turn into the one-equilibrium model, that is, if a country’s labor force increases its productivity so that the wages are high enough to keep children in school, then there would be no working-children equilibrium. This could be the case in industrialized countries. The opposite would occur when wages are so low due to an unproductive labor force, as is the case in poor countries, that there would only be a working-children equilibrium.

It is worth mentioning the complementary contributions of Rosati and Tzannatos [2006]. They present an altruistic model where the decisions concerning child labor also enter the

parent's utility function. This function depends on the parent and his or her offspring's consumption in two periods of life. The authors corroborate the negative income effects on child labor, similar to Basu's findings. Moreover, they suggest that high returns to education could induce parents to choose more schooling for their children, while high returns to child labor could make it less likely for children to continue studying. According to Baland and Robinson [2000], child labor is inefficient even if parents are altruist, because of its negative effects on the social returns to education, as well as the lower ability to capture private returns. This would occur in the case of poverty or imperfect capital markets. On the other hand, the authors have found that child labor is efficient only when the returns to human capital investment, in terms of income, is equal to its opportunity cost.⁶

Child labor and family size. Basu and Van [1998] show that an increase in the number of children in families can generate child labor. Considering "*number of children*" as a new variable in the model, the new budget constraint will be:

$$c + m\beta c = mew_C + w_A, \quad (3.10)$$

where $m (\geq 1)$ is the number of children in a household and $\beta (< 1)$ is the fraction of a child's consumption in terms of an adult's consumption c . Thus, the children's effort will be determined by:

$$e(w_A, w_C, m) = \begin{cases} 0 & \text{if } s + sm\beta + mw_C \leq w_A, \\ 1 & \text{if } s + sm\beta - mw_C \geq w_A, \text{ and} \\ \frac{mw_C - w_A + s + sm\beta}{2mw_C} & \text{otherwise.} \end{cases} \quad (3.11)$$

The aggregate labor market will be modified as follows:

$$\begin{aligned} S^A &= N \quad \text{and} \\ S^C(w_A, w_C, m) &= mNe(w_A, w_C, m). \end{aligned} \quad (3.12)$$

$$\begin{aligned} D^A(w_A^*, w_C^*) &= N \quad \text{and} \\ D^C(w_A^*, w_C^*) &= S^C(w_A^*, w_C^*, m). \end{aligned} \quad (3.13)$$

Compared to that of the basic model, the equilibrium in the labor market depends also on the number of children within the household. As the number of children increases, the parents are more likely to send them to work. Similar to the basic model, the authors show that this enhanced version has two stable equilibria, assuming that $w_A = \frac{w_C}{\gamma}$. It consists of a high-wage equilibrium in which children do not work, and a low-wage equilibrium where child labor comes up. Assuming a bad scenario where the population's age structure is strongly biased toward

⁶This seems not to be the case in Venezuela, where returns to education are high, and in fact, one of the main sources of inequality [Arcand and Carpio, 2007, Patrinos and Sakellariou, 2005].

young people (high m), the parents cannot afford to pay for the entire household consumption. Therefore, they send their children to work, increasing the effective labor supply. This scenario can generate the existence of just one stable equilibrium, in which children and adults work, and wages are very low.

Child labor and parent's education. Basu (1999) introduces dynamics to the child labor analysis in order to study its long-term consequences. The outstanding result of this model is what he calls “child labor trap” equilibrium. Roughly, this equilibrium depicts the case of a father who earns very little because of the low level of education he acquired as a child, and so he has to send his child to work. Hence, his child might not be able to go to school in turn and is likely to follow the same path as his father's, i.e., earning a low wage.

The author assumes the existence of overlapping generations. It means that a child in the first period will become an adult in the second period, and will have a child. In a daytime period denoted by 1, a child can be at school or at work. If e is the time spent at work, then $1 - e \equiv h$ is the time spent at school acquiring human capital. As the productivity of an adult depends on the amount of human capital h , the number of labor units produced (L) is given by:

$$L_t = L(h_{t-1}) \quad L' > 0, L'' < 0. \quad (3.14)$$

Assuming a perfectly elastic labor demand, the wage per unit of labor will be fixed at \bar{V} . The adult's income in period t depends on the human capital acquired as a child and the wage rate. It is defined as follows:

$$\bar{V}L(1 - e_{t-1}) \equiv W_t. \quad (3.15)$$

The amount of work that a child could do in period t is a function of the parent's wage in period t :

$$e_t = e(W_t), \quad (3.16)$$

$$e(W) = \begin{cases} 0 & \text{if } W \geq \bar{W}, \\ 1 & \text{if } W \leq \underline{W}, \text{ and} \\ \alpha & \text{if } \underline{W} < W < \bar{W}, \end{cases} \quad (3.17)$$

where \underline{W} and \bar{W} represent the wage limits on which an adult has to decide between sending a child to work or to school. Let α be the portion of the day that a child works, $0 < \alpha < 1$.

From (3.16) and (3.17), the function of child's effort is as follows:

$$e_t \equiv e(\bar{V}L(1 - e_{t-1})). \quad (3.18)$$

Given the assumptions in (3.17), the latter function can be written as $e_t \equiv \Phi(e_{t-1})$, where Φ is upward sloping. Hence, the father's time spent at work in $t-1$ could lead to the children

spending their time also at work. The author considers the case of two stable equilibria. The bad equilibrium, called “child labor trap”, describes a poor unskilled parent with a child working full-time. On the other hand, the good equilibrium or the virtuous cycle depicts a skilled father who earns enough to send his child to school.

The author suggests that an intervention policy just needs to focus on educating one generation to get the family out of the “child labor trap” and positioning the economy at the virtuous equilibrium.

3.3 Empirical Approach

We build our empirical model on some key variables mentioned in the theoretical models presented above, i.e., household income, parents’ education, and family size. Besides other standard covariates such as child age and gender, we also include a dummy for whether parents work as informal workers. The latter is one of our contributions to the subject of school dropouts. Additionally, we introduce several control variables related to the parents’ work status, marital status, and whether the child is already a parent. Another important control that measures mobility is a dichotomous variable indicating whether the family has moved recently (less than 4 years). It is important to point out that the set of household characteristics and parents’ education can be correlated with the household income. For that reason, we do not include the latter in our main regressions. However, we repeat all the regressions, including *per capita* household income as a control to reduce the risk of our informality indicator picking up any income effect; the results are quite similar to those shown in this section.

We first estimate a reduced-form equation for school dropouts. Since we use panel data, our model for binary outcomes is treated as the unobserved effects logit model. The binary response model is given by:

$$P(D_{it} = 1|X_{it}, c_i) = \Phi(X_{it}\beta + c_i), \quad t = 1, \dots, T, \quad (3.19)$$

where D_{it} is equal to 1 when child i is not attending school in period t , and 0 otherwise, X_{it} represents the explanatory variables, and the existence of additive random effects is denoted by c_i . Following Wooldridge [2002], the model relies on three assumptions. First, we suppose X_{it} is strictly exogenous, conditional on c_i . The second hypothesis is the independence of D_{it} across t . The presence of the correlation between the composite latent error $c_i + e_{it}$ across time may generate a serial dependence or within-panel correlation, therefore violating this hypothesis. We account for the presence of a serial correlation by means of a generalized estimating equations (GEE) approach [Wooldridge, 2002, p. 487]. Accordingly, this alternative is more efficient than pooled probit but less efficient than random effects maximum likelihood estimate (MLE) under the full set of assumptions. The last assumption implies that c_i and X_{it} are independent, and that the additive random effects follow a normal distribution.

We believe that the existence of unobserved heterogeneity affecting the decision of high school dropouts is likely to be correlated with X_{it} . For instance, a child’s ability or motivation could determine an early parenthood. On the other hand, the parents’ ability could have an effect on their children dropouts and their own work status. In order to account for omitted

factors, some studies adopt instrumental variables; for instance, Cardoso and Dorte [2006] use the ideal age for first-time sexual intercourse as an instrument for early parenthood. Since there is neither a good instrument nor can fixed-effects estimators be used in this study because many relevant variables do not vary over time, Mundlak corrections are applied [Mundlak, 1978] in our random effects model to deal with the potential endogeneity problem. Following Wooldridge [2002], unobserved effects c_i are allowed to be correlated with X_i by assuming a conditional normal distribution with a linear expectation and a constant variance. The Mundlak [1978] assumption is:

$$c_i|X_i \sim Normal(\psi + \bar{X}_i\epsilon, \sigma_a^2), \quad (3.20)$$

where \bar{X}_i is the average of X_{it} in $t=1, \dots, T$, and σ_a^2 is the conditional variance of c_i in equation $c_i = \psi + \bar{X}_i\epsilon + a_i$. The random effects model with Mundlak corrections takes the following form:

$$D_{it} = \psi + X_{it}\beta + \bar{X}_i\epsilon + a_i + e_{it}, \quad t = 1, \dots, T, \quad (3.21)$$

where a_i is a vector of unobserved effects, which are not correlated with X_i , and e_{it} is the idiosyncratic error term. The advantage of this specification is that once we introduce the averages of all explanatory variables \bar{X}_i , we are able to control for the correlation between our explanatory variables and time-invariant unobserved heterogeneity. In other words, this would take out the part of the variables that is correlated with the time-invariant unobserved heterogeneity. Thus, β should draw near to fixed-effects estimators.⁷ For comparison purposes, we also modeled a pooled ordinary least squares (OLS), pooled probit, GEE model, and random effects panel probit.

Finally, we estimate a multinomial logit model (MNL) with unobserved random effects in order to distinguish among idle children, working children, and those combining study and work. We use four categories: study only, work only, study and work, and idle [Rosati and Tzannatos, 2006, Cigno, Rosati, and Tzannatos, 2002]. This model allows us to understand better the likelihood that children fall under one of the four categories, given changes in some household and individual characteristics.

3.4 Data

The source of our data is the household survey called *Encuesta de Hogares por Muestreo* (EHM) from the Venezuelan National Institute of Statistics (INE). This survey contains information about socioeconomic characteristics of individuals, including intra-household relationships and several questions about employment issues. The sample is nationally representative, with the exception of the indigenous population (less than 1% of the total), the inhabitants of small islands, and people who are imprisoned, in hospitals, or living in hotels. We consider six rounds of the survey, from 2005 to 2007, using half-year waves. Some households are followed over time, given that one-sixth of the sample is replaced every semester. Hence, if we include more waves of the survey, we would lose the observations needed to construct our panel.

⁷Mundlak correction is also used when the incidental parameters problem is present; see Green and Hensher [2010].

The unit of analysis is high school-age children (11-17 years old). After eliminating individuals by their age and because of missing information, the number of children left is 49,161. Since the survey asks questions about the relationships among people in the household, we have relevant information about the children's parents, e.g., level of education, marital status, and work status. Besides the household head's children, the data allows us to identify the children from other families living in the same household, e.g., there might be a household head's sister with her son living with her and her family. The share of households with more than one nuclear family is about 10%. We use a dummy to identify the children who are part of the household head's nuclear family. The number of children with both parents living in the household is 60% of the sample. The construction of our dependent dummy variable, "high school dropout" is based on the question whether the high school-age child attends an education center.⁸ Since not all children have both parents in the household,⁹ we first generate the set of explanatory variables related to parents as an average of the data between them, i.e., parent's education is the average between the years of education of both parents or just the years of education of one of them, in case there is only one. For the variable "work", the value 0 means none of the parents works, 0.5 means one of them has a remunerated job, and 1 means both have remunerated work, or either the mother or the father works, in case of a single-parent. This could be read as the share of remunerated working parents living in the household. Additionally, we include dummies for the activity in which the household head works. An alternative sample would be based on children with both parents, either married or cohabiting.¹⁰ This sample would allow us to distinguish the effect between the mother and the father. However, we could face a sample selection problem, if we take into account only these types of families. We may be leaving out single-parent families in which their children are associated with higher dropouts rates (as shown later in the results). For that reason, we prefer to keep the complete sample. Table 3.1 shows descriptive statistics of the main variables for the whole sample.

3.4.1 Informality: What are we measuring?

The definition of informality is complex due to the many faces of this phenomenon. According to International Labour Organization. ILO Department of Statistics [2011], the conceptual framework can be divided into two major concepts. One is the Informal Sector which is an enterprise-based concept and Informal Employment that refers to a job-based concept. The informal sector consists of "*An unincorporated enterprise is a production unit that is not constituted as a separate legal entity independently of the individual (or group of individuals) who owns it, and for which no complete set of accounts is kept. An enterprise is unregistered when it is not registered under specific forms of national legislation (e.g., factories' or commercial acts, tax or social security laws, professional groups' regulatory acts).*" Informal Employment refers to "*those jobs that generally lack basic social or legal protections or employment benefits and may be found in the formal sector, informal sector or households. Informal employment includes*

⁸The question is *Asistencia a un centro de educación*. Answers are: a) *Sí*, b) *No*, c) *No aplica*, d) *No sabe*, and e) *No declarado*.

⁹According to data from the household survey, 40% of children live in single-parent households.

¹⁰Lerman [2002] points out that the gain in economic well-being depends on marital status; he particularly distinguishes married from cohabitation. See also Ginther and Pollak [2003].

the following types of jobs: (i) own-account workers employed in their own informal sector enterprises; (ii) employers employed in their own informal sector enterprises; (iii) contributing family workers, irrespective of whether they work in formal or informal sector enterprises; (iv) members of informal producers cooperatives; (v) employees holding informal jobs in formal sector enterprises, informal sector enterprises, or as paid domestic workers employed by households; (vi) own-account workers engaged in the production of goods exclusively for own final use by their household, if considered employed given that the production comprises an important contribution to total household consumption.” See more detail in Hussmanns [2004] and Gasparini and Tornarolli [2007]. According to the World Bank [2007a], direct approaches to measurement are based on microeconomic data and can be separated into two categories, productive and legalistic definitions. The International Labor Organization (ILO) notes the traditional productive definition referring to nonprofessional self-employed workers and microfirms with no more than five employees. Then, in an attempt to capture evasion and/or coverage in large firms, the legalistic perspective of informal employment emerged. Following the ILO definition related to this category: *“employees are considered to have informal jobs if their employment relationship is, in law or in practice, not subject to national labor legislation, income taxation, social protection or entitlement to certain employment benefits (advance notice of dismissal, severance pay, paid annual or sick leave, etc.)”*.

Given the richness of our data, different measures of informality are shown in Figure 3.1. For these variables, we just create a dummy where 0 indicates neither the father nor the mother works in the informal sector, and 1 denotes at least one of them works in the informal sector.¹¹ Informality is clearly high in Venezuela, and according to the World Bank [2007a], it is also increasing. The dynamics that lead people to be self-employed might be different from those of workers who accept jobs without benefits. However, as also mentioned by the World Bank [2007a], these measures are positively correlated (about 0.40), and the main reason is because many self-employed are unprotected workers (without benefits). Since our interest is to measure informality in a direct and unambiguous way, we use the legalistic view, which basically refers to unprotected workers.

It is important to point out that we do not try to explain the determinants of informality. However, to provide a better understanding of who the unprotected workers are, we present a simple regression of informal employees in Table 3.2. Consistent with the literature on informality, we find that informal employees are identified as low-skilled and low-income workers. Also, they are mostly women with more children than formal workers have, and more likely to have moved recently. Unprotected workers work in small firms and mainly in activities such as agriculture, transportation, and communications.

3.5 Results

Table 3.3 shows our regression results across several methods mentioned above, always controlling for region, time, and type of activity dummies. For the logit model with pooled data,

¹¹Regarding the legalistic definition, we use a survey question asking about employment rights and benefits. The variable shows whether or not the employee enjoys the benefits established in the Venezuelan labor law, i.e., pension funds, holiday bonus, and Christmas bonus.

similar to the results reported in previous literature, we find that being a boy and older are associated with a higher probability of leaving high school. Interestingly, other studies show the opposite effect in terms of gender, e.g., in Africa and India, girls are more likely to leave school [Rosati and Deb, 2004]. Although gender inequalities still exist in Venezuela, in terms of access to school there is no significant difference between girls and boys. In fact, girls' enrollment rate in secondary school is higher than that of boys. Moreover, as mentioned in the World Bank [2012], girls belonging to low-income families tend to stay longer in school than boys do. There are two main explanations for this matter; one concerns the significant efforts made by governments for decades to offer equal education for girls and boys, and we can see this pattern in the Latin American region. Second, the reduced number of young males attending high school may be associated with violence and crime. Briceño, Villaveces, and Concha [2008] show that 39% of all homicide deaths in Venezuela involve young males from 15 to 24 years old; the country's homicide rate was on average 37.7 per 100,000 inhabitants from 2000 to 2007 (over three times the global average rate).

As Cardoso and Dorte [2006] note, we find that early parenthood seems to be the most important reason for dropping out of high school. When a child between 11 and 17 years old becomes a parent, the probability of leaving school increases by almost 19%. Additionally, a high number of siblings and dwellers in a household is also associated with an increase in the likelihood of children's dropouts [Fafchamps and Wahba, 2006, Rosati and Tzannatos, 2006].

Less educated parents are more likely to withdraw their children from high school. One more year of education for parents is correlated with a 0.7% decrease in the probability of their children dropping out of school.¹² Although working parents tend to keep children in school, after controlling for parents' education, those who work specifically in precarious jobs (or as unprotected workers) have the opposite effect on children's dropouts. The fact that at least one of the parents has a job without any of the labor rights established by law is correlated with a 0.6% increase in the likelihood of his or her child leaving secondary school (conditional on the mean of the rest of the variables). One of the possible explanations might be that because unprotected jobs are by nature more unstable and demand low skills only most of the time, parents may keep their children out of school, either to get them to work or to take care of household chores in belt-tightening times. As Fafchamps and Wahba [2006] argue, the labor market conditions under which parents work can shape their decision to withdraw their children from school.

Another interesting finding involves the nonlinear effect of parents' ages. Children are more likely to drop out of school when their parents are very young or very old. One can think of young and inexperienced parents with planning and monetary problems who end up deciding to withdraw their children from school, as well as older parents who could have health problems and need their offspring to help with household chores or the family business. Regarding marital status, married parents are more likely to keep their children in school. Results are basically the same for the random effects and GEE panel, but it seems that coefficients in the random effects model are overestimated, partly due to serial correlation corrected by using GEE.

As mentioned in the methodological strategy, we introduce Mundlak corrections, assuming

¹²Similar results are found by Cigno et al. [2002] in India, Vietnam and in urban areas of Morocco.

the existence of endogeneity for some of the explanatory variables (Table 3.4). Once we control for endogeneity, the covariates that seem to have a real impact on high school dropouts are early parenthood, children's age, and parents' education. When a child between 11 and 17 years old becomes a parent, the probability of leaving school increases by almost 0.32%, compared to the case of a child with no offspring. Furthermore, as children grow old, they are more likely to leave school, possibly because the cost of opportunity of being in school increases as well. The probability increases by 0.07% for each year of a child's age. In addition, as mentioned in the literature, the likelihood of withdrawing a child from school decreases by 0.01% when his or her parent have one more year of education.

In an attempt to figure out whether the effect of informal jobs is nonsignificant because of the lack of variation in the variable or because of nonlinear effects, we add some interactions of this variable with other covariates (Table 3.4, in columns 2,3 and 4). We find that parents who have precarious jobs in terms of employee benefits are more prone to withdraw their children from school, by 0.18%, than parents with formal jobs. This result comes out after adding the interaction between an unprotected worker and his or her child's age. The effect of being a youth, nearly 17 years old, and having a parent that changed from a formal to an informal job, decreases the probability of dropping out of school. One of the explanations could be that parents may decide to become unprotected workers during their midlife because this type of employment offers them a better situation than the former job did. Even if they do not have all the employee benefits, the World Bank [2007a] mentions that it is more suitable for some experienced and entrepreneurial workers to be self-employed than to be formal employees.

Another interesting result is that the probability of dropping out of school is higher for a child who lives in a household with a large number of adult dwellers and whose parent becomes an unprotected worker. For each additional adult living in the household, the likelihood of leaving school for the child of an informal employee increases by 0.01%. This can be the case of extended families living in the same household, which may present a high degree of residential mobility; therefore, the prospect of changing schools may result in dropouts [Astone and McLanahan, 1994]. Delgado [2005] also suggests that overcrowded households can negatively affect child development because parents are unable.

Finally, the probability of a child leaving school rises by 0.07% when at least one of the parents is working as an unprotected worker in the trade sector. This result is expected in the context of Venezuela, where trade is the second most important non-oil sector activity after manufacturing (11% of GDP).¹³ A noticeable case is that of street vendors; policymakers have been trying to formalize their work status, but efforts have failed.

Although the effect of having an unprotected working parent on the probability of high school dropouts is not as significant as the effect of early parenting, it is worth mentioning that parents' choice of working in informal activities has a larger effect than parents' education.

¹³Available at the Central Bank of Venezuela: <http://www.bcv.org.ve/c2/indicadores.asp>.

3.5.1 Children's activities

Table 3.5 presents the results of the multinomial logit regressions.¹⁴ These results allow us to see the differentiated impact across categories, which are: “work only”, “study and work”, and “idle”, “school only” being the base category. We divide the table into four panels, each one showing different specifications with emphasis on informality variables. The effects of age and early parenthood remain strongly significant and with expected signs across regressions and categories. Comparing among categories, older children seem to be more attracted to leave high school and begin to work than to leave school and remain idle, or work/study at the same time.

As the results in Table 3.5 show, younger children of unprotected working parents are more prone, by 1.766, to leave high school and remain idle than to stay in school (panel A). However, this informal working condition of parents seems not to have a significant effect on the probability of being a “working” or a “working and studying” child. The same pattern occurs when we introduce an interaction between the number of individuals in the household and informal working parents (panel B). The likelihood of a child dropping out of school and being idle versus keeping studying increases by 0.115 when his or her parent becomes an informal worker and they live in an extended-family household (e.g., uncles, aunts, cousins, and grandparents). Nevertheless, the probability of being a child who leaves school for work or does both work and study, is not affected by having informal worker parents and living with their extended family.

As expected, we also find that unprotected farmer parents tend to have children who work and study at the same time, rather than just studying (panel C). However, being an unprotected parent working in activities other than agriculture decreases the probability of having a working/studying child by 0.379, compared to having a student child. Additionally, results show that becoming an unprotected farmer does not translate into having a higher probability of putting one's children to work only or in an idle condition than keeping them studying. Agricultural activities are performed mostly by low-skilled workers. Thus, it seems highly reasonable that a farmer's child is assigned some tasks to help his parents once he arrives from school.

The opposite effect is shown when we emphasize the case of parents working in the construction sector. The probability of being an idle child increases by 0.259 when at least one parent is working in a precarious job in any activity other than construction. Accordingly, those parents working in construction informally are more prone to keep their children in school. This result is consistent with the growth path that the construction sector experienced between 2003 and 2008. In fact, it was the second fastest-growing sector in the Venezuelan economy during that period.

3.6 Concluding remarks

Child labor and school dropouts are still major problems in Venezuela. Although there has been an increase in gross enrollment ratio at the secondary school level during recent decades, the shares of school dropouts and child labor remain significant. Almost 10% of secondary school-

¹⁴Following Small and Hsiao [1985], the test of independence of irrelevant alternatives (IIA) indicates that our four alternatives are independent at the 1% level.

age children are out of the classrooms, and even more at the upper secondary level, making Venezuela one of the worst performing countries in Latin America in terms of high school attendance. These children and youth dropouts end up entering the labor market without any technical or vocational training. Most likely, as they grow up, the cost of human capital investment will increase. Therefore, they will notably be low-skilled, low-wage workers.

In this section, we find that early parenthood has a significant effect on the probability of dropping out. Being pregnant or having a baby increases the likelihood of dropping out of school. Additionally, we point out the effects of labor market conditions, for instance, the role of informality. We find that if parents are unprotected workers, their children are more likely to leave high school, either to work or to do nothing. Moreover, as children grow up, they are prompted to leave high school, either because the household needs additional income or they lose interest in studying. Parents' education also contributes in the probability of their children dropping out from school early. Parents with low levels of education provide less motivation for their children to keep studying.

We find other results that corroborate the findings of previous applied research. However, these results cannot be interpreted as causalities. For instance, families with few children and few dwellers in their households are associated with a low probability of withdrawing their children from school. In the same way, married parents and families who have not recently moved are positively correlated with children who attend school.

We go in depth and also study how the different activities in which children are involved are affected by the informal working condition of their parents. The activities are: "work only", "study and work", "study only" and "idle". We show that the probability of dropping out of school and remain idle increases when the parent of a child work as an informal employee, compared to the probability of continuing studying. Nevertheless, the transition of leaving school and becoming a worker or does both work and study, is not affected by having informal working parents.

In terms of policymaking, this study sheds light on some strategic directions to take in order to promote high school attendance. First, to avoid high school dropouts, more resources should be provided for teen pregnancy prevention programs or existing programs should be implemented more efficiently. Second, as it is not just a problem of economic resources, the government must identify parents' work conditions when implementing social programs to promote education. Educational programs must differentiate between protected and unprotected workers. For instance, conditional cash transfer programs might not be effective in terms of school attendance, when parents prefer to include their children as part of the labor force in their informal business, instead of hiring non-family workers. Sectoral policies should ensure that children's participation in family businesses does not interfere with school completion.

Third, policymakers should pay particular attention to youth in upper secondary schools, who are at a high risk of being prompted to leave school. Programs to promote education at this level must assure the supply of upper secondary schools throughout the country. Authorities should also reassess the content of the curriculum to adapt it to the country's needs, particularly the appropriateness of general and vocational programs. For instance, in Finland, 50% of youth enroll in general programs in the upper secondary level; the other 50% enroll in

vocational/technical programs [UNESCO, 2007]. Further research has to be done regarding the youth's lack of interest in finishing high school.

Fourth, another vulnerable and easily targeted group comprises children of low-skilled parents. In order to avoid the “child labor trap” issue mentioned by Basu [1999], policymakers must also convey more support to this group of children by better targeting existing or new educational programs.

Fifth, authorities should also pay attention to those children who work at the farms, to guarantee that they do not drop out of school. The curriculum in rural high schools may be evaluated and designed to allow children to both work and study in a suitable way.

Finally, idle children are in a more precarious situation than children working and studying or working only. Policies aimed at decreasing high school dropout rates must differentiate between these two groups of children. The number of idle children is alarming, and this issue should be addressed promptly because of its linkages to criminality. According to the United Nations Office on Drugs and Crime [2008], Venezuela is among the five worst performers in Latin America and the Caribbean in terms of homicide rates. Illegal activities could be linked to why so many children (33%) do not attend school and report “other situation” when asked about their activities the previous week. The repercussions of dropouts on the accumulation of human capital may, in turn, affect economic growth and the level of crime in Venezuela. In other words, the phenomenon of school dropouts is, ultimately, economically and socially harmful for the country.

REFERENCES

- R. Almeida and P. Carneiro. Enforcement of labor regulation, informal labor and firm performance. *World Bank Policy Research Working Paper*, (3756), 2005.
- J.-L. Arcand and S. Carpio. Errors in variables in pseudo-panel data: A simple correction. CERDI. Université d'Auvergne. (MIMEO), 2007.
- N. Astone and S. McLanahan. Family structure, residential mobility and school dropout: A research note. *Demography*, 31(4), 1994.
- J. Baland and J. Robinson. Is child labor inefficient? *The Journal of Political Economy*, 108(4):663–679, 2000.
- K. Basu. Child labor: Cause, consequence, and cure, with remarks on international labor standards. *Journal of Economic Literature*, 37(3):1083–1119, 1999.
- K. Basu and P. H. Van. The economics of child labor. *The American Economic Review*, 88(3):412–427, 1998.
- G. Becker. *A Treatise on the Family*. Harvard University Press, Cambridge, MA, 1991. Enlarged.
- F. Blanco. *Trabajo infantil en Venezuela: 1998-2007. Understanding Children's Work Project*, 2009.
- F. Blanco and C. Valdivia. Child labour in venezuela: children's vulnerability to macroeconomic shocks. *Understanding Children's Work Project*, 2006.
- R. Briceño, A. Villaveces, and A. Concha. Understanding the uneven distribution of the incidence of homicide in latin america. *International Journal of Epidemiology*, 37(4):751–7, 2008.
- A. R. Cardoso and V. Dorte. School drop-out and push-out factors in brazil: The role of early parenthood, child labor, and poverty. *IZA Discussion Paper*, (2515), 2006.
- A. Cigno, F. Rosati, and Z. Tzannatos. Child labor handbook. Technical Report 0206, World Bank, 2002.
- Y. Delgado. *Venezuela y Colombia. Trabajo Infantil: Camino y Frontera*. Technical report, Ministerio de Ciencia y Tecnología de Venezuela. FONACYT., 2005.
- M. Fafchamps and J. Wahba. Child labor, urban proximity, and household composition. *Journal of Development Economics*, 79(2):374–397, 2006.
- L. Gasparini and L. Tornarolli. Labor informality in latin america and the caribbean: Patterns and trends from household survey microdata. Technical Report 46, *Centro de Estudios Distributivos, Laborales y Sociales*, 2007.
- D. Ginther and R. Pollak. Does family structure affect children's educational outcomes? *NBER*, (9628), 2003.

- W. Green and D. Hensher. *Modeling Ordered Choices: A Primer*. Cambridge University Press, Cambridge, 2010.
- R. Hussmanns. Measuring the informal economy: From employment in the formal sector to informal employment. ILO Working Papers 375000, International Labour Organization, 2004. URL <http://ideas.repec.org/p/ilo/ilowps/375000.html>.
- International Labour Organization. ILO Department of Statistics. Statistical update on employment in the informal economy. http://www.ilo.org/public/libdoc/ilo/2011/111B09_241_eng1.pdf, 2011.
- D. Kruger, R. Soares, and M. Berthelon. Household choices of child labor and schooling: A simple model with application to brazil. *IZA Discussion Paper*, (2776), 2007.
- R. Lerman. Marriage and the economic well-being of families with children: A review of the literature. Technical report, Urban Institute and American University, 2002.
- Y. Mundlak. On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85, 1978.
- H. Patrinos and C. Sakellariou. Schooling and labor market impacts of a natural policy experiment. *Labour*, 19(4):705–719, 2005.
- G. Psacharopoulos. Child labor versus educational attainment. some evidence from latin america. *Population Economics*, 10(4):377–386, 1997.
- F. Rosati and P. Deb. Determinants of child labor and school attendance: The role of households unobservables. *Understanding Children’s Work (UCW) Working Paper*, 2004.
- F. Rosati and Z. Tzannatos. Child labour in vietnam. *Pacific Economic Review*, 11(1):1–31, 2006.
- K. A. Small and C. Hsiao. Multinomial logit specification tests. *International Economic Review*, (26):619–627, 1985.
- UNESCO. Institute for statistics. data centre. http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=136&IF_Language=eng&BR_Topic=0, 2007. Accessed: 2009-09-10.
- United Nations Office on Drugs and Crime. Statistics. <http://www.unodc.org/unodc/en/data-and-analysis/homicide.html>, 2008. Accessed: 2008-12-09.
- J. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, Massachusetts, 2002.
- World Bank. *World Development Report. A Better Investment Climate For Everyone*. Washington, DC, 2005.
- World Bank. *Informality. Exit and Exclusion*. Washington, DC, 2007a.

World Bank. *World Development Report. Development and The Next Generation*. Washington, DC, 2007b.

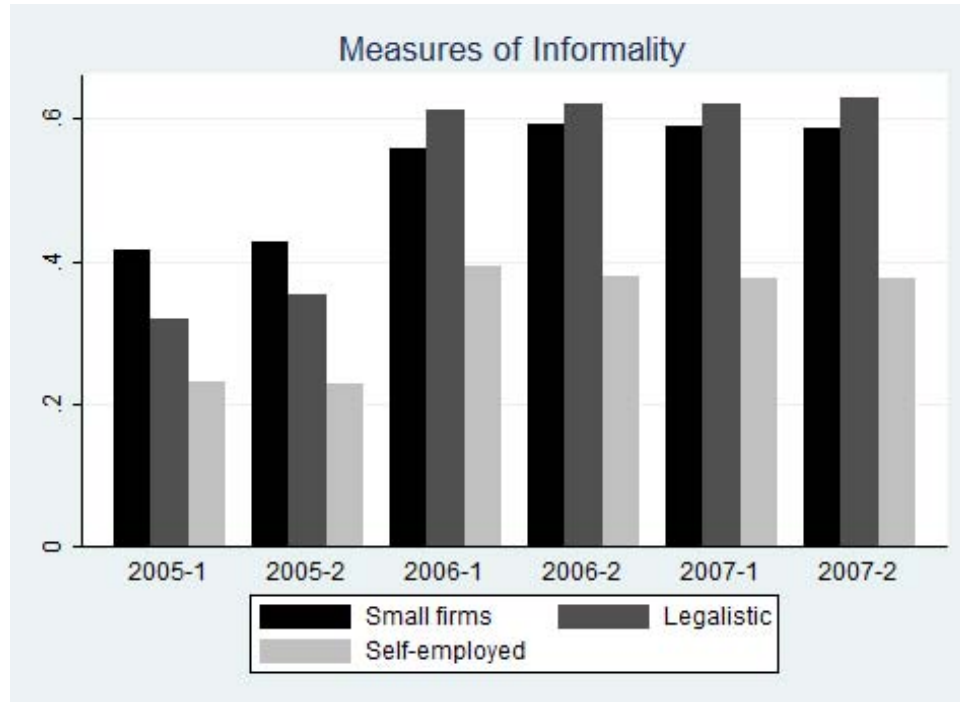
World Bank. *World Development Report. Gender Equality and Development*. Washington, DC, 2012.

World Bank, International Labor Organization, and United Nations Children's Fund. *Joining Forces Against Child Labour. Inter-agency Report for the Hague Global Child Labour Conference of 2010*, 2010.

3.7 Tables and Figures

Table 3.1: Descriptive statistics 2005-2007. Per capita income is measured in *Bolivares* at constant prices, base 2007. Source: Author's calculations based on EHM

Variables	Mean	Std. Dev.	N
Child characteristics			
Share of high school dropouts	0.0940	0.2918	107828
Gender: Girl	0.4870	0.4998	107828
Age	13.8808	1.9767	107828
Years of education	6.9709	2.2231	106607
Share of household head's child	0.8929	0.3091	107828
Share of child-parent	0.0113	0.1059	107828
Number of siblings	2.6801	1.4234	107828
Average income in the region	4.0742	1.1510	107828
Family characteristics			
Number of individuals	5.8206	2.2496	107828
Parent's age	41.1376	7.0358	107828
Parent's years of education	8.4516	3.3613	107828
Single	0.0821	0.2745	107828
Married	0.3540	0.4782	107828
Cohabiting	0.4242	0.4942	107828
Separated	0.1457	0.3528	107828
Work	0.7077	0.3401	107828
Unprotected worker	0.5576	0.4966	107828
Family moved	0.0211	0.1439	107828
Per capita income	240.7877	264.6095	107828
Piped water supply at home	0.9055	0.2925	95646

Figure 3.1: Share of households with at least one informal working parent by three definitions of informality. Source: Author's calculations based on EHM**Table 3.2:** Random effects logit model on informal employment. Dummies for state, activity, and time are included. Agriculture is the omitted activity. Asterisks refer to the t-statistics significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Variables	Unprotected workers
Individual characteristics	
Gender: Girl	0.2572 ***
Age	0.0081 ***
Years of education	-0.1232 ***
Married	-0.0114
Number of children	0.1219 ***
Number of individuals	-0.0033
Family moved recently	0.4919 ***
Employment characteristics	
Real income	-0.5127 ***
Small firm(< 5)	2.4004 ***
Oil/mining	-2.4025 ***
Manufacturing	-1.2288 ***
Electricity/water	-2.0325 ***
Construction	-0.1869 ***
Trade	-0.7177 ***
Transport/communication	0.2125 ***
Financial services	-1.4577 ***
Other services	-1.7786 ***
Number of observations	194244

Table 3.3: Dependent variable: Child Dropouts - pooled logit, random effects logit and panel GEE - 2005-2007. Dummies for state, activity, and time are included. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1

Variables	Pooled Logit	RE Logit	Panel GEE
Coefficients as marginal effects			
Child characteristics			
Gender: Girl	-0.0283***	-0.0034***	-0.0264***
Age	0.0257***	0.0033***	0.0230***
Child is parent	0.1881***	0.0696***	0.0972***
Family characteristics			
Unprotected worker parent	0.0061***	0.0006***	0.0044***
Number of siblings	0.0026***	0.0004***	0.0033***
Number of individuals	0.0027***	0.0003***	0.0023***
Parents' age	-0.0025***	-0.0003***	-0.0003
Squared parents' age	0.0000***	0.0000***	0.0000
Parents' years of education	-0.0071***	-0.0009***	-0.0058***
Married	-0.0132***	-0.0017***	-0.0103***
Work	-0.0090***	-0.0010***	-0.0075***
Family moved recently	0.0242***	0.0033***	0.0112**
Number of obs.	107724	107724	68319
Number of children	57834	57834	22817

Table 3.4: Dependent variable: Child Dropouts - random effects logit with Mundlak correction - 2005-2007. Dummies for state, activity, and time are included. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1

Variables	1	2	3	4
Informality variables				
Unprotected worker parent(UW)	0.0001	0.0018 **	-0.0009	-0.0000
UW x age		-0.0001 **		
UW x number of individuals			0.0001 **	0.0002
UW x agriculture				0.0007 **
UW x trade				0.0003
UW x construction				0.0000
UW x transportation				-0.0001
UW x cleaning				
Child characteristics				
Gender: Girl	-0.0005	-0.0005	-0.0005	-0.0005
Age	0.0007 ***	0.0007 ***	0.0007 ***	0.0007 ***
Child is parent	0.0032 ***	0.0032 ***	0.0032 ***	0.0032 ***
Family characteristics				
Number of siblings	-0.0001	-0.0001	-0.0001	-0.0001
Number of individuals	-0.0000	-0.0000	-0.0001	-0.0000
Parents' age	0.0000	0.0000	0.0000	0.0000
Squared parents' age	-0.0000	-0.0000	-0.0000	-0.0000
Parents' years of education	-0.0001 **	-0.0001 **	-0.0001 **	-0.0001 **
Married	-0.0002	-0.0002	-0.0002	-0.0002
Work	-0.0003	-0.0003	-0.0003	-0.0003
Family moved recently	0.0003	0.0003	0.0003	0.0003
Number of obs.	107724	107724	107724	107724
Number of children	57834	57834	57834	57834

Table 3.5: Multinomial logit model (MNL) for three alternatives - Work Only, Study & Work, Idle - 2006-2007. Dummies for state, activity, and time are included. Coefficients as odd ratios. Asterisks refer to the t-statistics significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Base group: School only. Only statistically significant coefficients are shown.			
Panel A	Work	Study & Work	Idle
Informality variables			
Unprotected worker parent(UW)	2.6395	0.2531	1.7663**
UW x age	-0.1837	-0.0358	-0.1053*
Child characteristics			
Age	1.2279***	0.5215***	0.7904***
Child is parent	1.2424***	1.0201*	1.4812***
Family characteristics			
Number of individuals	-0.1668**	-0.0689	-0.0010
Parents' years of education	0.0355	0.0752*	-0.0342
Work	0.4180	0.4653**	-0.2937
Number of obs.	131768	131768	131768
Panel B			
Informality variables			
Unprotected worker parent(UW)	-0.5191	-0.2654	-0.5197*
UW x number of individuals	0.0524	-0.0002	0.1155***
Child characteristics			
Age	1.0901***	0.4977***	0.7173***
Child is parent	1.0211**	0.0284**	1.4882***
Family characteristics			
Number of individuals	-0.2029**	-0.0675	-0.0810
Parents' years of education	0.0350	0.0753*	-0.0360
Work	0.4152	0.4648**	-0.2945
Number of obs.	131768	131768	131768
Panel C			
Informality variables			
Unprotected worker parent(UW)	-0.2678	-0.3793**	0.2011
UW x agriculture	0.0045	0.8043**	0.0038
Child characteristics			
Age	1.0859***	0.4861***	0.7194***
Child is parent	1.2259**	1.0420*	1.4794***
Family characteristics			
Number of individuals	-0.1621**	-0.0649	0.0054
Parent's years of education	0.0349	0.0787*	-0.0021***
Work	0.4169	0.4801**	-0.3048*
Number of obs.	131768	131768	131768
Panel D			
Informality variables			
Unprotected worker parent(UW)	-0.2088	-0.2365	0.2598*
UW x construction	0.1222	-0.2029	-0.4258*
Child characteristics			
Age	1.0865***	0.4974***	0.7220***
Child is parent	1.2259**	1.0168*	1.4928***
Family characteristics			
Number of individuals	-0.1639**	-0.0683	-0.0001
Parents' years of education	0.0348	0.0757*	-0.0324
Work	0.4060	0.4675**	-0.2835
Number of obs.	131768	131768	131768

Chapter 4

Youths in the Informal Sector: Evidence from Brazil

4.1 Introduction

Bosch and Maloney [2010] suggest that young Brazilian employees work involuntarily in the informal sector. In the literature about the shadow economy, this phenomena is known as the queuing view. World Bank [2007] argues that informal jobs may offer an entry point to the labor market for young and unskilled workers. They would hold informal jobs until they get the skills needed to find a better job. Another motivation for working in the informal sector is related to the help of children/youths in the family business. For instance, a son working in his father's small business to contribute to the household income. World Bank [2007] shows that Dominican unpaid workers are family of the owner, and the main reason for working without monetary compensation (62%) is because they have to help their family.

There are several risks that young workers may face if their first job is informal salaried. Ultimately, this may affect the firm's productivity where they work. First, applicable to all young workers, work keeps them from allocating time to their studies. Therefore, the process of human capital accumulation is undermined and consequently, they fail to get a better job. Second, in the event of a crisis, unregistered workers are more likely to become unemployed. According to Bosch, Goni, and Maloney [2007], the probability of transition to unemployment is three times higher and more volatile for informal salaried workers than their formal counterparts in Brazil. The authors suggest that one explanation may be the lack of protection of these workers against firing. Youth are the first workers to be laid off given their lack of experience. Another mechanism could be related to the high exposure and vulnerability of small firms during economic recessions. The third risk related to informal jobs is that young workers who drop out of school before completing secondary education are less likely to migrate to formal jobs in the future. Arias and Bustelo [2007] and Arias, Landa, and Yáñez [2007] in World Bank [2007], show that the percent of workers that never held a job other than informal salaried is 60% and 33%, for Dominican Republic and Argentina, respectively. Thus, the mobility from informal to formal sector can be quite difficult. Fourth and the focus of this study, hiring young workers whose characteristics are low skills and lack of experience can negatively affect productivity

of the business. One mechanism can be the fact that young workers have less professional experience and lack the “soft” skills that older workers may have improved throughout their professional life. Another explanation may be that young workers are more likely to study and have a part time job. Hence, it would be difficult for them to be fully committed and highly productive.

In this chapter, we analyze the effect of the share of young workers on small firms’ performance using a survey from Brazil. Several studies show a negative effect of informality on firms’ productivity [Fajnzylber, Maloney, and Montes-Rojas, 2011, World Bank, 2007] and also on productivity growth at the aggregate level [Farell, 2004]. Yet, the effect of age composition of informal workers on firms’ performance has not been studied.¹ There are some works using European data. For instance, Lallemand and Rycx [2009] show that increasing the share of young workers has a positive effect on productivity in Belgian firms. They argue that one explanation can be that young workers are healthier, more mobile and motivated individuals. In contrast, Mahlberg, Freund, and Prskawetz [2011] find that there is a negative relationship between the share of young workers and the productivity of the firm. They use an Austrian matched employer-employee panel data. Vandenberghe and Waltenberg [2010] analyze the effect of aging on firm’s productivity using Belgian panel data. They find that younger workers between 18-29 are less productive than prime-aged workers. Yet the effect is not statistically different from zero. There is a potential endogeneity problem that most of the literature on this topic does not account for. Changes in workers’ age composition and changes in productivity can be jointly determined by unobserved heterogeneity. For instance, a bad manager may induce low levels of productivity, and also may not make good decisions when hiring workers. He may choose a cheaper labor force (i.e., young or old workers). Another source of endogeneity is the simultaneity or reverse causality between productivity and age composition. For instance, owners dealing with low performance firms are not able to pay a highly skilled labor force, thus they hire younger workers. The article by Ours and Stoeldraijer [2010] is one of the most comprehensive work in this topic. They use a matched worker-firm level dataset from Dutch manufacturing companies to estimate the effects of age structure on the pay-productivity gap for the period 2000-2005. They account for endogeneity bias using panel data and also the instrumental variable (IV) approach. Ours and Stoeldraijer [2010] show that the lowest productivity is associated with workers under 25 years-old. One of the weaknesses of their analysis is the lack of educational attainment variables. If the level of education for the CEOs or workers varies over the period 2000-2005, then their estimates may be biased. Unlike our data, the studies mentioned above use datasets from medium or large firms and are mostly in the manufacturing industry. One of our contributions consists in providing insight on how younger workers may affect the level of productivity of a small firm.

As mentioned above, the model that we want to estimate is likely to generate misleading estimates because of the double causality (i.e., the share of young workers in a firm may be a function of the firm performance). Since we have cross-sectional data, we deal with the endogeneity bias using the IV method. We suppose that the sex of children are likely to be randomly assigned within a family group. However, whether the child is sent to work in the

¹Informal workers’ definition is mentioned in Chapter 3.

family business may depend on gender. There is an extensive body of literature dedicated to explain why parents, particularly fathers, tend to favor boys [Bertrand, Johnson, Samphantharak, and Schoar, 2008, Bennedsen, Nielsen, Pérez-González, and Wolfenzon, 2006, Dahl and Moretti, 2004, Angrist and Evans, 1998]. According to Dahl and Moretti [2004], boys have more opportunities for work (e.g., they help their parents in rural activities). The authors analyze the effect of child sex composition on divorce, fertility and shotgun marriage. They show that there is a parental gender bias towards boys in US and several developing countries. They suggest three explanations, not mutually exclusive, for this preference. One is simply related to the taste of parents. Boys could generate more utility to their parents. Second, parents believe that a father plays a more important role in the development of his son than of his daughter. The authors called this hypothesis *role model*. Third, parents can perceive a *differential cost* between raising a boy and raising a girl, either monetary or in terms of time resources. The authors mention that, according to the December 2000 Gallup poll of 1,026 American adults, most think that girls are harder to raise. The work done by Angrist and Evans [1998] show that parents prefer mixed-sex children. However, the families with two boys are less likely to have a third child compared to families with two girls.

In a family business context, the gender bias is also present. Several studies analyze the effect of succession decisions on firm performance [Bertrand et al., 2008, Bennedsen et al., 2006]. According to Bertrand et al. [2008], using data from large firms in Thailand, the number of family members working in the business increases with the number of sons of the firm founder. Bennedsen et al. [2006] use the gender of the first-born child as an instrument for family successions in Danish firms. They argue that fathers favor boys when it comes to choosing someone to work and manage the firm. On the other hand, Carpio [2011] shows that in the Venezuela case, dropout girls tend to stay at home and take charge of household chores, whereas dropout boys tend to work out of the home.

This chapter is divided as follows. First, we present the main characteristics of the informal sector in Brazil. Next, in Section 2 we describe the data. Section 3 explains the econometric approach. Finally, in Sections 4 and 5 we present our findings and conclusions.

4.2 Informality in Brazil

Informality in Latin America is an economic phenomenon that has been identified as a result of high inequality, weak rule of law and recurrent episodes of macroeconomic instability. As mentioned above, this phenomenon can be measured in different ways given that it is a very heterogeneous sector. In the case of Brazil, the size of the informal sector was equivalent to almost 40% of the country's GDP. About 30% of the labor force does not have the right to a pension linked to employment when retired, and more than 30% of sales were not reported for tax purposes [World Bank, 2007]. Although Brazil does not have the biggest informal sector in Latin America, many studies are devoted to this topic and it is a major concern for the government. According to International Labour Organization. ILO Department of Statistics [2011] five countries, including Brazil (India, Mexico, Vietnam and Pakistan), concentrate three-fourths of the total informal employment estimated for a group of 46 countries. Between

1992-2002, the percent of salaried workers without pension coverage increased by 6.2% in the Brazilian metropolitan areas, whereas the same statistic decreased by 0.8% at the national level. This means that it is a growing problem that takes place mostly in the main cities of Brazil. According to Botelho and Ponczek [2011], some of the reasons that explain the great participation of unregistered workers in the economy are the increase of mandatory fringe benefits in 1988 and the increased trade openness during 1990s. They argue that there is little room for negotiation between firms and employees since labor contracts are highly rigid and costly. In order to understand the rigidity of their labor contracts, it is worth mentioning the main benefits established in the labor laws. Employers, with respect to their employees, agree to comply with clauses of the standard labor contract written in 1943. This contract states that the worker is entitled to receive a thirteenth salary by the end of the year, one-month paid vacation, severance payment for unjustified dismissal, a work week of 44 hours, a food and commuting subsidy and a four-month paid pregnancy leave for women. There is also a minimum wage to be complied with by employers.

It is worth noting that in the 1990s and early 2000s, the macroeconomic situation was characterized by episodes of economic stagnation and an acute increase of unemployment. One of the structural adjustment policies that the Brazilian government introduced was trade liberalization. Menezes-Filho and Scorzafave [2009] suggest that trade openness during the 1990s played a decisive role in workers' reallocation into the informal sector. Mainly those workers from comparative advantage sectors and export industries became informal salaried or self-employed. However, Bosch et al. [2007] argue that trade openness had a small impact on the increase of informality compared to the impact of the very rigid labor laws in Brazil. In the same vein, Heckman and Pagés [2004] and Almeida and Carneiro [2005] show that changes in labor legislation have a considerable impact on the size of the formal sector and productivity.

Botelho and Ponczek [2011] analyze the informal sector in Brazil, more specifically their hypothesis is whether there is segmentation in the Brazilian labor market.² They define unregistered workers as those who do not possess the work booklet ("carteira nacional de trabalho"). Using data from the six largest metropolitan areas, the authors found that unregistered workers accounted for 27% of total workers in 2000, without including self-employed. According to their findings, the labor market is segmented since there is a wage differential of around 7.8%. Segmentation is greater for young workers (under 30 years old) and for less-educated workers. For instance, the wage differential for a 25-year-old worker with 2 years of schooling is 10.8 percentage points higher than a 60-year-old university graduate worker. They also found that the segmentation decreases as the level of income increases. The wage premium for low-wage workers is around 15%, while there is no evidence of segmentation between workers in the top of the wage distribution.³

Authors also mention that in the literature of informality there are two views of informal work. One of them consists of a temporary state during recessions in which informal workers are not able to find a job in the formal sector. It is also defined as a "*transitory phase*" for new entrants in the labor market in early stages of their life cycle. These informal workers would

²The presence of segmentation would be reflected by a wage premium for workers in the formal sector.

³Informal workers in the top of the distribution are the richest employees and it does not include self-employed.

prefer to have formal jobs, but there are barriers that hamper their access to them. Later, once they gain some professional skills and experience, they will be able to get a formal job. The other view considers informal jobs as an attractive alternative since it provides more flexibility and workers can evade taxes and cumbersome regulations. Authors suggest that the case of Brazil falls into the first view. World Bank [2007] also mentions that most of the Brazilian workers in the informal sector did not choose voluntarily to work in that sector. Instead, they describe the phenomena as an outright exclusion. However, they also argue that youth may choose to work in the informal sector. Either because they place more value on having more flexibility at work or because, for instance, they are committed to work in family small businesses. In the same vein, the majority of independent or self-employed workers have voluntary motives to exit the formal sector and choose to work in the informal sector. The principal reason for choosing informal work is to avoid taxes and cumbersome paperwork. The difference between age groups is clear; young workers are more vulnerable and their interests and motivation can be very different from the prime-aged and older population. As this work focuses on young informal workers, the next section describes in detail the youth situation in the informal sector.

4.2.1 Youth in the informal sector

Young workers⁴ may prefer to start their professional careers in the formal sector in order to receive all the benefits established in labor laws, and perhaps a higher salary. In fact, about 75% of Brazilian young informal salaried between 15-18 years prefer a signed work contract.⁵ This corroborates the theory of exclusion from the formal sector for most of young informal workers. The same response of dissatisfaction among young informal salaried is exhibited in Colombia, Bolivia, Dominican Republic and Argentina. On the other hand, 25% of Brazilian young informal workers prefer an unprotected job. Most of them (74%) declare that they are happy in their current unprotected job. One of the reasons can be related to the flexibility offered by the informal sector (e.g., they may look at informal jobs as an opportunity to work and do some other activities at the same time). Another reason may be that they believe the informal job is a good match for them given their low skills.

In order to understand how important the proportion of youths in the informal sector is, we highlight some labor force statistics presented by Menezes-Filho and Scorzafave [2009]. Most young workers, between 15-24 years, find their first job in the informal sector. The high incidence of unregistered working youths reached above 60% of total working youths from 1991 to 2004. Since then, it has been slightly declining to arrive at approximately 55% in 2007. The incidence of informality for the other groups, 25-50 and 51-65 years, is lower than the one shown by workers between 15-24 years. The latter group is clearly in the most vulnerable and precarious position. According to Menezes-Filho and Scorzafave [2009], young workers perceive the informal sector as a point of entry to the labor market. In that way, they would acquire experience and skills needed to be able to find a formal job. Many studies favor this hypothesis such as Bosch and Maloney [2010], Cunningham, McGinnis, García-Verdú, and Tesliuc [2007], World Bank

⁴Age range identifying young people can vary depending on each study. However most papers, including ours, use between 15-24 years old.

⁵See World Bank [2007], estimates based on the national household survey in 1990.

[2007] and Lucas [1978]. Cunningham et al. [2007] show that Argentina, Brazil and Mexico have the same life-cycle pattern across sectors (self-employed, formal and informal). In these countries, youths represent around 20% of informal salaried employment, which means the highest participation among all unregistered salaried workers. The opposite happens when analyzing self-employed. Older workers (between 35 and 55 year) are largely represented in this sector.

Youth participation rate in the labor force, between 15-24, decreased during the 1990s and then stabilized around 60%. Menezes-Filho and Scorzafave [2009] argue that the phenomenon of aging in the Brazilian population and the great efforts for keeping youths in school may be the main reason of an unchanging participation rate of this group in the 2000s. They show that the share of young people in the labor force has also been falling over the period 1985-2007. At the same time, the unemployment rate for this age group increased in the 1990s, reaching 20% in 2003.⁶ Since then, the youth unemployment rate has not fallen back. Thus, although many young people are spending more years at school as mentioned by Menezes-Filho and Scorzafave [2009], there are still serious problems of unemployment and participation in the informal sector for this group. In fact, youths present high levels of mobility between unemployed status and informal salaried [World Bank, 2007]. The authors argue that some of the reasons leading to this pattern can be the difficulty of dismissing a new worker with a formal contract, or exclusionary factors as a result of deficient schooling.

Workers in informal firms tend to be related to the owner of the firm. In Mexico, most workers in micro-firms are directly related to the owner of the firm [World Bank, 2007]. These workers probably receive part of their payment in food or shelter. Almost all unpaid workers are family members (97%). An interesting fact is that among workers, family members are younger than non-relatives, and unpaid relatives are even younger. Another characteristic of small firms in Mexico is that few have registered workers with the social security administration, while most firms with 10 or more workers have at least one registered worker. The fact that small firms are composed of family members and that most of them are unregistered can be linked to the owner's lack of managerial skills. Therefore, he or she is unable to hire qualified workers or expand the business. Authors suggest that the more family members work in the firm, the fewer formal contracts there will be. The same authors show that in the Dominican Republic unpaid workers are asked the reason for working without monetary compensation, and the main motivation (62%) is that they want to help their families.

4.3 Data

We use firm and individual level data from the Brazilian Institute of Geography and Statistics (IBGE). The survey is called *Economia Informal Urbana 2003* (ECINF). The appealing feature of the data is that it has both household and firm level information. The questionnaires look at issues on individual characteristics as well as on firm attributes.

ECINF-2003 was constructed using the probability sampling method and its design consisted of a two-stage stratification process. Using the Demographic Census 2000, primary units (urban

⁶The unemployment rate for the other age groups are much lower, 7% and 4% for 25-50 and 51-65, respectively.

sectors) were stratified by geographic areas and then, each geographic stratum was stratified by the level of household income.

IBGE used “number of owners in the productive unit” as the factor that determines the sample size in each geographic region, with a coefficient of variation of 5%. Then, they identified the households to be surveyed (secondary units), which were households occupied by self-employed or employers with less than five paid employees, dedicated to non-agricultural activities. The target households were classified according to economic activity declared by those surveyed. Finally, from this list they selected some household businesses with a probability proportional to the total of units in this census (this sampling method is known as probability proportional to size). In this sense, the sample is representative of Brazilian urban small firms, ensuring all income categories. For ECINF 2003, IBGE classified firms in the informal sector as those formed by self-employed workers and employers with less than five employees (irrespective of the number of non-remunerated employees or owners). They take into account firms in urban areas only. Firms classified as *Sociedades Anonimas* are considered as formal; firms that declared income taxes in 2003 are considered as formal (specifically those who filled in the *lucro real form*). Agricultural and illegal activities are excluded from the sample, as well as domestic services.

Most firms in the survey are managed by self-employed workers (73%). This means that unregistered firms have few employees. However, we limit our sample to firms with more than one worker. The reason is because in order to include important variables and the instrument, we need to be able to identify the relationship between owners and workers.

Only 12% of employees from unregistered firms have a work permit. It is worth noting that half of the work force, from firms that did declare income tax, does not have a work permit. In other words, although the firm follows the legal framework, their employees are not legally registered. This shows that there are serious problems of enforcement [Almeida and Carneiro, 2005]. Interestingly, high bureaucracy and high costs are the main drawbacks declared by owners to register their businesses.

Regarding the unit used in our regressions, we collapse the information to have firm-level observations. In other words, each observation is a firm. The reason is because many firms have more than one owner and their responses are different. Hence, the average will serve to mitigate potential measurement errors. The variable called “firm’s productivity” is measured by the firm’s value added, that is, sales minus intermediate costs. The variable measuring capital includes all, owned or rented, properties, buildings, machines, equipment, and vehicles declared in the survey. The age structure of the workforce consists of three groups. We define all employees or owners who are less than 24 years old as young workers. Prime-age workers are those with ages between 24 and 60, and older workers are represented by people over 60 years old. These cut-off points were chosen in order to compare with other studies. Education, gender and skin tone are included as control variables.

4.3.1 Descriptive statistics

Table 4.1 presents the summary statistics of each variable used in the econometric model. The share of young workers, our variable of interest, represents 23.7% of the workers in the firm.

It is clear that most workers belong to the prime-age group (72.6%). The age of the firm is on average 10 years, and the average number of workers is approximately 3. The nearest bank is located less than half an hour from the business (on foot). Regarding owner characteristics, most of them have achieved primary or secondary level education. This means that most owners in small businesses are low-skilled workers. It is worth noting that owners began to work when they were very young (14 years old). This finding may reflect the high likelihood of young people ending up in the informal sector when they stop schooling and begin to work. Additionally, more than half of the workers are white and 33% are females.

Table 4.1 also shows, in the last two columns, the difference between the two types of firms across several firm characteristics. Firms with at least one young employee (less than 24 years) seem to be statistically different from firms with only an adult labor force, in terms of size, gender composition, and value added per worker. Young employees work in larger and newer firms. Owners who hire young workers tend to be younger and began to work earlier than owners from firms with adult employees only. Also, firms with young workers show a slightly lower value added than those with no young employees.

Table 4.2 presents the differences between young and older informal workers. It shows that roughly one-third of unregistered workers under 24 years have achieved primary education, while unregistered workers with no education tend to be in the older age-group. Differences between these two groups intensify when we compare the number of hours worked. Older workers work more hours than young ones. Interestingly, it seems that young owners of firms search for young employees. Additionally, the proportion of young workers without remuneration is higher than the share of older workers with no remuneration (24.8% versus 19.2%). As we mention above, this fact confirms that youths are the most vulnerable group. As expected, a significant share of unregistered youths are sons or daughters of the firms' owners.

In Brazilian small firms, youths tend to work in certain industries given their lack of experience (see Table 4.3). The mining sector and other services have the highest participation of young workers in Brazil in 2003. They are less present in the transportation and communication sectors, as well as education and health services, lodging and restoration, and building.

4.4 Econometric strategy

We estimate the effect of young workers on firm's productivity. The dependent variable is measured as the value added of the firm per worker. In small firms, particularly if they are not registered, contracts are verbal and these may be based on informal arrangements. Workers may not work for the same hours. For example, youths work less hours than adults (see Table 4.2). As a robustness check, instead of per worker, we use an alternative definition of firm's productivity as the value added per hours worked. The reason is that we want to make sure our variable of interest "share of young workers" is not picking up the variation in hours worked by youths. First, we present a simple linear least square regression for the following equation:

$$P_i = \alpha + \beta_1 A_{i1} + \beta_2 A_{i2} + \beta_3 A_{i3} + \delta X_i + \epsilon_i, \quad i = 1, \dots, N \quad (4.1)$$

where the dependent variable is P_i , the logarithm of value added per worker of firm i is a function of the age structure defined by three variables representing the share of each age group of workers (A_{1i} , A_{2i} , A_{3i}), indicating youth, prime-aged and older, respectively. In addition, several firm characteristics represented by X_i are related to the firm's performance.

Potential endogeneity arises from two sources. First, non-observable characteristics of the owner (e.g., managerial skills, ability, and effort) may be correlated to both low-skilled young workers and firm's productivity. For instance, Mahlberg et al. [2011] use panel data, solving this problem by removing the fixed effects. There is also a potential simultaneity bias between productivity and the share of young workers. Owners of low productivity firms may not be able to pay for a skilled workforce, therefore hire cheaper workers such as youths or unpaid family members. If we do not take into account these sources of endogeneity our estimates could be biased. Due to the lack of panel data in our survey, we use instrumental variables. Thus, we need a variable uncorrelated with firm's productivity, but correlated with the access of young workers to the firm. A good candidate is the gender of the owner's children. We use the sum of young sons living in the household of the firm's owner as our instrument. The strength of our instrument is based on the fact that the sex of children at birth is biologically determined. It is very unlikely that parents could control the sex of their child.⁷

There is extensive research on gender bias towards sons, particularly concerning the allocation of time between work, school, and household chores for children within the family. Boys are more likely to work than girls. Guarcello, Henschel, Lyon, Rosati, and Valdivia [2006] analyze child involvement in work activities in several countries. In the case of Brazil, they use the household survey of 2003. The report shows that the proportion of boys in economic activity is greater than that of girls (9.5% versus 4.6%, respectively) for the 7-14 age group. They mention that the sex differential increases with age. The same finding is shown in a later report by ILO [Rosati, Manacorda, Kovrova, Koseleci, and Lyon, 2011]. One of the explanations of this pattern in the context of a family business is that boys are supposed to be the successors, thus they need to gain working experience [Bertrand et al., 2008, Bennedsen et al., 2006]. A second explanation mentioned above is that the intra-household decision of hiring a son is driven by the belief that boys have more labor opportunities than girls. A third explanation may be related to crime in Brazil; parents may believe that boys are less vulnerable than girls in risky situations. Thus, parents are more prone to send boys to out-of-home work than girls. As mentioned above, World Bank [2007] points out the high probability in small firms of hiring family members. We

⁷According to Arnold [1997], Brazil does not show any pattern of sex-selection abortion techniques. The author argues that in countries where son preference is strong, such as China and Korea, male/female ratio of children at birth tends to increase over time. He shows that Brazil is not among the countries with strong gender bias. Sex ratio of Brazilian children at birth does not increase or decrease monotonically during the period reported retrospectively by respondents (1971-1991). The author analyzes the gender bias using other indicators, such as desire for pregnancy, intention to use a contraceptive method, use of contraception, use of sterilization, contraceptive discontinuation rates, percentage of pregnant women, school attendance and some health indicators as well. He demonstrates that countries such as India, Egypt, Nepal or Bangladesh have a strong son bias. In Brazil, most of the results show an erratic behavior in parents' responses. For instance, the percentage of women with two children that have an intention to use a contraceptive method is 46% for those having all sons versus 31% for those having all daughters. However, the opposite result is shown in the case of women with three children. We think that the desired number of children might be affecting the results. For instance, if the ideal number of children in Brazil is three and there is a gender bias towards boys, then couples that have three boys could no longer have a strong gender preference.

show, as well, in Table 4.1, the overwhelming increase of the share of family members in firms with at least one young worker compared to businesses without youth participation.

Gender preference may also be observed when analyzing the effect of parents educational attainment. Emerson and Souza [2007] find that there is an intra-household gender bias in the allocation of resources in the child labor context. By using the Brazilian household survey for 1998, they show that fathers favor boys, while mothers favor girls. The negative impact of father's education on child labor is higher for sons than for daughters. The authors mention that one of the explanations is the differences in parental preferences related to a child's gender.

4.5 Results

Table 4.4 shows the results of our estimations. As expected, the share of young workers has a negative effect on a firm's performance and it is statistically significant. It shows that if the proportion of youths in the firm increases by 1%, then the value added per worker would decrease by 0.35% compared to an increase of the share of prime-age workers (omitted variable). Since we control for education, this effect is entirely due to age and not to the low level of education inherent in youth. This effect can be explained by the fact that the young have less professional experience and lack the "soft" skills that older workers may have developed over time.⁸ Soft skills are perhaps the most important characteristics affecting labor market outcome and success in life as mentioned by Heckman and Kautz [2012]. Another explanation may be that young workers are more likely to study and have a part time job. Hence, it would be difficult for them to be fully committed and be highly productive. Similarly, firms with an older workforce tend to perform worse than firms with mostly workers between 24 and 60 years old. Our results of the age composition estimates confirm what was shown in the literature [Mahlberg et al., 2011, Vandenberghe and Waltenberg, 2010, Lallemand and Rycx, 2009]. It is worth noting that the results are not strictly comparable to the existing literature due to the differences in the size of firms. They use firm-level data composed by medium and large firms (more than 20 employees). Whereas in our dataset, there is a maximum of 30 workers per firm.

For the rest of the covariates, most of them show correlations with the expected signs and are statistically significant. The capital stock and the age of the firm are positively related to a firm's productivity. There is potentially a significant selection bias driven by the inclusion of the variable age of the firm. Presumably, the firms that were really unproductive are no longer in the sample because they went out of business. Therefore, the marginal effect could be biased. Particularly, small firms are more vulnerable and tend to close down. We assume this variable as a control and we should be careful with its interpretation.

Like most papers that analyze the relationship between productivity and the size of the firm, we find that an increase in the number of workers is associated to a higher value added; however, this effect is not statistically significant. Since firms are very small, size does not vary much. Firms with three workers or less account for 74% of our sample. Nevertheless, we include the number of owners because it seems that this kind of firm tends to be composed by a group of owners, instead of one owner and several workers. The results show that productivity increases

⁸Soft skills refer to personality traits, such as communicational skills, self-esteem, punctuality, and team-working abilities.

with the number of owners, yet the relationship is non-linear. Firms with many owners and many employees are likely to perform worse. This is reasonable if we imagine a firm with ten owners and five employees trying to make decisions about reducing costs or putting a price on a service. Hierarchy and responsibilities may not be well defined; hence, making decisions can be complicated.

The disaggregated effect of education on firm performance is very straightforward. The more educated the workforce, the better performance the firm will show. Furthermore, the effect is even bigger at the college level of education than for the primary level. This is true for owners or for employees. The share of females, whether employees or owners, is negatively correlated with a firm's performance. However, the increase of white workers may be associated with a rise of the firm's productivity. A peculiar finding concerns the proximity to a bank. Firms located far from a bank display higher levels of productivity than firms located close to a bank. This result may be driven by the fact that only 8% of firms in our sample uses a public or private bank as a first source of loans. The rest ask family members or providers for money. Thus, proximity to a bank may not affect firm performance. Also, since it is a urban survey, banks may not be that far. In the sample, the average distance to a bank is 30 minutes on foot.

The age of the owner and his or her age at their first job has a positive impact on a firm's performance. However, the age squared has a negative effect. That means that as the owner postpones his or her entry into the labor market, he or she is more likely to better manage the firm and, therefore, achieve higher productivity levels. Productivity increases with the owner's age. However, the relationship is non-linear, it has an inverted-U shape. Firms composed of mostly older owners are likely to exhibit low performance. The same results are found in other studies, and the reason may be that, as for workers in general, older owners are less willing to invest in training programs to acquire new skills. Furthermore, problem-solving and mathematical abilities tend to decline with age [Lallemand and Rycx, 2009].

The problem with the previous estimates is that we have not resolved the potential endogeneity bias due to the reverse causality or omitted variables. The second and third columns in Table 4.4 show regressions that use the instrumental variable approach. In both columns, the share of young workers has a negative effect on firm's productivity. If youth participation increases by 1%, the value added per worker will decrease on average by 1.72% (column 2). It seems that the endogeneity problem causes an underestimation of the effect of young workers on a firm's performance. Similarly, firms where most workers are older are likely to show low productivity levels.

The size of the firm still shows a positive relationship but becomes statistically significant. The rest of the results are roughly the same. There are noticeable differences with respect to OLS estimations. The variable female owners become positive. One of the reasons could be that female owners tend to hire younger workers because, for example, it is a male chauvinist society where prime-age males do not respect female owners. The correlation between these two variables may disappear after instrumenting our endogenous variable. On the other hand, variables related to the age of the owner can be correlated to the age of workers. For instance, young owners tend to hire young workers. Thus, as a robustness check, we drop out the variables related to age of the owner and the results do not change much (see the first column in Table

4.5).

Finally, we use an alternative measure of productivity, value added per hours worked (third column in Table 4.4). The hours worked varies considerably among workers in small firms, even more among youths. In our sample, youths work less hours per month than other workers. Therefore, in order to avoid that our variable of interest pick up changes in hours worked, we include the alternative measure of value added. For that reason, this is our preferred estimation. Results are very similar to the regression in the second column.

An additional robustness check is shown in Table 4.5. In columns 2, 3 and 4, we drop out some variables which have many missing values; hence, it reduces quite a bit our sample. We want to check if those missing variables are not randomly distributed and may generate selection bias. Similarly in column 5, we test if the results change when dropping out the variable age of the firm, which is suspected to generate selection bias. The estimates of our variable of interest does not vary much. In the last column, we want to test if the sum of young daughters instead of the sum of young sons would yield the same results. However, the effect of the share of young workers on firm's productivity becomes statistically insignificant.

In order to verify the validity of our estimators, we perform the Angrist-Pischke test to verify whether the endogenous variable is "weakly identified". We reject the null hypothesis of weak identification with an $F\text{-stat}(1, 5593) = 56.77$ and $\text{Prob} \geq F = 0.0000$. Table 4.6 shows the first stage estimation.

4.6 Concluding remarks

This study has sought to show the impact of youth participation on small firms' productivity in Brazil. More than half of Brazilian young workers are in the informal sector. Although most of them would have preferred a formal job, the informal sector is an entry point to the labor market for this age group. In addition, they tend to take their first job very early in life, on average at the age of 14. Thus, it is clear that young workers not only lack experience, but also have low qualifications when starting to work. In our analysis, we find that an increase in the share of young informal workers will reduce a firm's productivity. This effect is considerably important compared to other characteristics linked to the firm. However, a variation in the share of older workers does not seem to impact productivity differently from the share of prime-age workers. As documented in the literature, another key factor in determining a firm's performance is the education of the owner(s). The use of technology and if the owner is male have a positive and significant impact on a firm's performance.

From our results, two main risks are identified for the young. One is related to the low level of education attained by young workers and the other is the low level of firm's productivity. Regarding the first risk, most young workers do not complete secondary education, thus disrupting human capital accumulation. Policy makers should address education programs for youths at risk in order to avoid school dropouts. In this case, youths at risk means those whose relatives work in the informal sector. The vulnerable situation of youths does not only affect their school attainment and their future returns to education, but also the productivity of the business where they work. We find that firms that have a younger workforce are less productive

than the ones comprising workers between 24-60 years old. Consequently, owners of less productive firms are unable to hire a skilled workforce, to expand their business, and to protect the firm against eventual negative shocks. Informal firms are already characterized by presenting lower productivity than formal ones. In addition to this, informal firms with a considerable participation of youths have the worst performance among informal firms. Therefore, firms with informal young workers fill the bottom of the firm's productivity distribution of the universe of enterprises in Brazil. Ultimately, the existence of these disadvantaged firms affects economic growth.

Some young people cannot stop working because of their extremely poor economic situation. Hence, poor families cannot afford to send young members to school; instead, they send them to work. In this case, it is difficult to motivate them to leave work and restart school. A more realistic policy recommendation could be to invest in training programs to improve the employability of young workers, similar to the labor training program put in place in the Dominican Republic called *Juventud y Empleo*. Ibarraran, Ripani, Taboada, Villa, and Garcia [2012] show the positive impacts of this program on the probability of getting a formal job and also on different measures of non-cognitive skills. Furthermore, training apprenticeship programs seem to reveal successful school to work transitions. Private-public partnerships are involved to train youths in locally demanding and more relevant areas [Steedman, 2012].

Further research needs to focus on the youth's decision of choosing informal jobs, and whether it is a voluntary selection or a consequence of the labor market exclusion. In order to shed light on this, it is crucial to have data on informal and formal enterprises. Additionally, it is important to understand the long-term effects of being an informal worker on job quality and salary. This analysis would require panel data and also a richer set of variables, including information about soft skills.

REFERENCES

- R. Almeida and P. Carneiro. Enforcement of labor regulation, informal labor and firm performance. *World Bank Policy Research Working Paper*, (3756), 2005.
- J. Angrist and W. Evans. Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 88(3), 1998.
- O. Arias and M. Bustelo. Profiles and dynamics of informal employment in latin america. Photocopy, World Bank, 2007.
- O. Arias, F. Landa, and P. Yáñez. *Movilidad Laboral e Ingresos en el Sector Formal e Informal de Bolivia*. Documento de trabajo, UDAPE. La Paz, Bolivia, 2007.
- F. Arnold. Gender preferences for children. Technical Report 23, Calverton, MD: Macro International., 1997.
- M. Bennedsen, K. Nielsen, F. Pérez-González, and D. Wolfenzon. Inside the family firm: The role of families in succession decisions and performance. *National Bureau of Economic Research*, (12356), July 2006.
- M. Bertrand, S. Johnson, K. Samphantharak, and A. Schoar. Mixing family with business: A study of thai business groups and the families behind them. *National Bureau of Economic Research*, (13738), January 2008.
- M. Bosch and W. F. Maloney. Comparative analysis of labor market dynamics using markov processes: An application to informality. *Labour Economics*, 17(4):621–631, August 2010.
- M. Bosch, E. Goni, and W. F. Maloney. The determinants of rising informality in brazil: Evidence from gross worker flows. IZA Discussion Papers 2970, Institute for the Study of Labor (IZA), August 2007. URL <http://ideas.repec.org/p/iza/izadps/dp2970.html>.
- F. Botelho and V. Ponczek. Segmentation in the brazilian labor market. *Economic Development and Cultural Change*, 59(2):437 – 463, 2011.
- S. Carpio. High school dropouts and the informal sector: Evidence from venezuela. CERDI. Université d'Auvergne. (MIMEO), 2011.
- W. Cunningham, L. McGinnis, R. García-Verdú, and C. Tesliuc. The promise of youth: Policy for youth at risk in latin america and the caribbean. Technical report, World Bank, 2007.
- G. Dahl and E. Moretti. The demand for sons: Evidence from divorce, fertility, and shotgun marriage. *National Bureau of Economic Research*, (10281), February 2004.
- P. M. Emerson and A. P. Souza. Child labor, school attendance and intra-household gender bias in brazil. Technical Report 2, World Bank, 2007.
- P. Fajnzylber, W. Maloney, and G. Montes-Rojas. Does formality improve micro-firm performance? evidence from the brazilian simples program. *Journal of Development Economics*, 94(2), 2011.

- D. Farrell. The hidden dangers of informal economy. *McKinsey Quarterly*, (3):26–37, July 2004.
- L. Guarcello, B. Henschel, S. Lyon, F. Rosati, and C. Valdivia. Child labour in the latin american and caribbean region: A gender based analysis. Technical report, International Labour Organization, 2006. Understanding Children’s Work (UCW). An Inter-Agency Research Cooperation Project with ILO, World Bank and Unicef.
- J. J. Heckman and T. Kautz. Hard evidence on soft skills. IZA Discussion Papers 6580, Institute for the Study of Labor (IZA), May 2012.
- J. J. Heckman and C. Pagés. Introduction to law and employment: Lessons from latin american and the caribbean. In *Law and Employment: Lessons from Latin America and the Caribbean*, NBER Chapters, pages 1–108. National Bureau of Economic Research, Inc, September 2004.
- P. Ibarraran, L. Ripani, B. Taboada, J. M. Villa, and B. Garcia. Life skills, employability and training for disadvantaged youth: Evidence from a randomized evaluation design. IZA Discussion Papers 6617, Institute for the Study of Labor (IZA), June 2012.
- International Labour Organization. ILO Department of Statistics. Statistical update on employment in the informal economy. http://www.ilo.org/public/libdoc//ilo/2011/111B09_241_engl.pdf, 2011.
- T. Lallemand and F. Rycx. Are young and old workers harmful for firm productivity? IZA Discussion Papers 3938, Institute for the Study of Labor (IZA), Jan. 2009. URL <http://ideas.repec.org/p/iza/izadps/dp3938.html>.
- R. Lucas. On the size distribution of business firms. *The Bell Journal of Economics*, 9(2), 1978.
- B. Mahlberg, I. Freund, and A. Prskawetz. Ageing, productivity and wages in austria: Evidence from a matched employer-employee data set at the sector level. ECON WPS - Vienna University of Technology Working Papers in Economic Theory and Policy 02/2011, Vienna University of Technology, Institute for Mathematical Methods in Economics, Research Group Economics (ECON), 2011.
- N. Menezes-Filho and L. Scorzafave. Employment and inequality of outcomes in brazil. Ibme Working Papers 200, Insper Instituto de Ensino e Pesquisa, October 2009.
- J. Ours and Stoeldraijer. Age, wage and productivity. IZA Discussion Papers 4765, Institute for the Study of Labor (IZA), February 2010.
- F. Rosati, M. Manacorda, I. Kovrova, N. Koseleci, and S. Lyon. Understanding the brazilian success in reducing child labour: empirical evidence and policy lessons. Technical report, International Labour Organization, 2011. Understanding Children’s Work (UCW). An Inter-Agency Research Cooperation Project with ILO, World Bank and Unicef.
- H. Steedman. Overview of apprenticeship systems and issues. ILO contribution to the G20 Task Force on Employment (9789221267416), International Labor Organization (ILO), November 2012. URL http://www.ilo.org/wcmsp5/groups/public/---ed_emp/---ifp_skills/documents/genericdocument/wcms_190188.pdf.

V. Vandenberghe and F. D. Waltenberg. Ageing workforce, productivity and labour costs of belgian firms. Discussion Papers (IRES - Institut de Recherches Economiques et Sociales) 2010003, Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES), 2010.

World Bank. *Informality: exit and exclusion*. Washington, DC, 2007.

4.7 Tables

Table 4.1: Descriptive Statistics 2003. Remuneration is measured in *Reais* at constant prices. The last two columns present the mean of each variable by age group. N=5594. The asterisks show the statistical significance of differences: *** p<0.01, ** p<0.05, * p<0.1. Source: Author's calculations based on ECINF 2003.

Variables	Mean	Std. Dev	No young	At least 1 young
Firm characteristics				
Share of young workers (<24 yrs)	0.2370	0.2530	0	0.4575***
Share of prime-age workers (24-60 yrs)	0.7264	0.2649	0.9339	0.5208***
Share of older workers (>60 yrs)	0.0364	0.1282	0.0671	0.0217***
Log of capital	8.6683	1.9091	8.6425	8.6913
Age of the firm (years)	10.1918	8.7979	10.7571	9.6884**
Number of workers (including owners)	2.8789	1.2050	2.5984	3.1287***
Number of owners	1.1125	0.3656	1.1085	1.1161
Primary	0.3148	0.4098	0.2574	0.3660***
Secondary	0.2849	0.4123	0.2752	0.2937
College	0.0609	0.2205	0.0651	0.0573
Female	0.3388	0.3384	0.3545	0.3250*
White	0.6190	0.4985	0.6170	0.6209
Time to nearest bank (hours)	0.3959	1.0798	0.3686	0.4203
Log of value added per worker	6.3913	1.0446	6.4414	6.3468**
Log of value added per hours worked	1.8446	1.3110	1.8497	1.8402
Sum of owner's young sons	0.5310	0.7658	0.4059	0.6424***
Owner characteristics				
Primary	0.2151	0.4045	0.2074	0.2220
Secondary	0.2513	0.4290	0.2344	0.2664
College	0.1917	0.4069	0.1885	0.1946
Owner's age	43.0400	10.7800	44.6364	41.6328***
Owner's age at first job	13.9523	5.1901	14.2708	13.6688**
Female	0.2058	0.3889	0.1805	0.2284**

Table 4.2: Differences between young (<24 years) and not young (>24 years) workers (measured as % of each age group unless mentioned otherwise). Source: Author calculation based on ECINF 2003.

Variables	<24 yrs	>24 yrs
No Education (%)	29.27	35.85
Primary (%)	35.74	26.31
Secondary (%)	28.71	30.20
College (%)	6.27	7.62
No remuneration (%)	24.76	19.17
Hours worked per month	149.88	172.18
Owner is family (%)	47.61	47.77
Owner is his/her father (%)	31.06	5.67
Owner's age (years)	35	42

Table 4.3: Share of Brazilian young workers by industry (2003). Source: Author calculation based on ECINF

Industry	N	Mean (%)
Mining (Oil/Coal excluded)	6	30.55
Manufacturing	800	24.08
Building	494	22.60
Trade/Maintenance	2038	26.27
Lodging/Restaurant	707	21.17
Transportation/Communication	240	19.08
Banking/Insurance	24	24.09
Real estate/Vehicle and Equipment rental	109	25.93
Computing/Provisioning services	464	23.40
Education/Health/Social services	711	20.52
Others	1	50

Table 4.4: Dependent variable: Log of value added per worker - 2003. Dummies for industry and region are included. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1

Variables	small firms		
	OLS(1)	IV(2)	IV(3)
Coefficients as marginal effects			
Firm characteristics			
Share of young workers	-0.3526***	-1.7226***	-1.5701**
Share of older workers	-0.2260	-0.5915**	-0.7120**
Ln(capital)	0.1801***	0.1629***	0.1315***
Age of the firm	0.0055***	0.0054*	0.0073**
Number of workers	0.0445	0.0796**	0.0405
Number of owners	0.1666*	0.0146	0.6940***
Owners X workers	-0.0151	-0.0108	-0.0758***
Share of employees with primary education	0.1415***	0.2932***	0.3277***
Share of employees with secondary education	0.2286***	0.3257***	0.2814***
Share of employees with college	0.3434***	0.4136***	0.3958***
Share of females	-0.2952***	-0.5031***	-0.7786***
Share of whites	0.1317***	0.1356***	0.0709
Time to nearest bank (hours)	-0.0368*	-0.0343	-0.0149
Owner characteristics			
Share of female owners	-0.0504	0.1132	0.3916***
Age of owners	0.0092	-0.0178	-0.0260
Age of owners ²	-0.0001	0.0002	0.0003
Age of owners at first job	0.0076*	0.0033	0.0030
Share of owners with primary education	0.0034	-0.0258	0.0905
Share of owners with secondary education	0.2719***	0.2605***	0.4574***
Share of owners with college	0.5057***	0.4941***	0.8300***
Number of obs.	5594	5594	5594

Table 4.5: Dependent variable: Log of value added per hours worked - 2003. Dummies for industry and region are included. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1

Variables	Dropped out variables				Other IV	
	Age(1)	Capital(2)	Education(3)	Nearest bank(4)	Firm age(5)	∑daughters as IV(6)
Coefficients as marginal effects						
Firm characteristics						
Share of young workers	-1.7993*	-1.5555**	-1.4947**	-1.4703**	-1.5565**	-1.1760
Share of older workers	-0.6299**	-0.6485**	-0.8459***	-0.7803***	-0.7292**	-0.6668*
Ln(capital)	0.1277***		0.1437***	0.1402***	0.1354***	0.1365***
Age of the firm	0.0059	0.0114***	0.0070**	0.0074**		0.0073**
Number of workers	0.0456	0.1198***	0.2078***	0.0082	0.0406	0.0404
Number of owners	0.6664***	0.8173***	-1.0410***	0.5945***	0.6966***	0.7378***
Owners X workers	-0.0751***	-0.0980***	0.2012***	-0.0531**	-0.0762***	-0.0770***
Share of employees with primary	0.3482***	0.3587***		0.2959***	0.3239***	0.2841**
Share of employees with secondary	0.2911***	0.2959***		0.3061***	0.2832***	0.2535**
Share of employees with college	0.3890***	0.5188***		0.4266***	0.3930***	0.3756***
Share of females	-0.8120***	-0.8253***	-0.7332***	-0.7693***	0.7941***	0.7188***
Share of whites	0.0771	0.1195**	0.0498	0.0612	0.0730	0.0697
Time to nearest bank (hours)	-0.0145	-0.0172	-0.0195		-0.0146	-0.0157
Owner characteristics						
Share of female owners	0.4241***	0.3293***	0.3793***	0.2802***	0.3894***	0.3446**
Age of owners		-0.154	-0.0274	-0.0318	-0.0243	-0.0183
Age of owners^2		0.0002	0.0003	0.0003	0.0003	0.0002
Age of owners at first job		0.0077	0.0021	0.0047	0.0022	0.0043
Share of owners with primary	0.1007	0.1169	0.1598**	0.1428**	0.0908	0.0989
Share of owners with secondary	0.4728***	0.5639***	0.5607***	0.4522***	0.4528***	0.4607***
Share of owners with college	0.8504***	0.9656***	0.9783***	0.7895***	0.8286***	0.8384***
Number of obs.	5594	6048	6826	6833	5594	5594

Table 4.6: First-stage. Dependent variable: Share of young workers - 2003. Dummies for industry and region are included. Asterisks refer to the t-statistics significance: *** p<0.01, ** p<0.05, * p<0.1

Variables	First Stage
Excluded instrument	
Sum of young sons	0.0574***
Firm characteristics	
Share of older workers	-0.2276***
Ln(capital)	-0.0109***
Age of the firm	-0.0000
Number of workers	0.0255***
Number of owners	-0.1006***
Owners X Workers	0.0024
Share of employees with primary education	0.1058***
Share of employees with secondary education	0.0688***
Share of employees with college	0.0447
Share of whites	0.0066
Share of females	-0.1255***
Time to nearest bank (hours)	0.0031
Owner characteristics	
Share of female owners	0.1061***
Age of owners	-0.0240***
Age of owners ²	0.0002***
Age of owners at first job	-0.0031**
Share of owners with primary education	-0.0124
Share of owners with secondary education	-0.0022
Share of owners with college	-0.0013
Number of obs.	5594

Conclusion

In this last section, we present a brief review of each chapter's conclusions and discuss general policy implications.

Chapter 1: The Venezuelan Context: Economy and Education System. This chapter aims at providing a wide background of the Venezuelan context, including a brief review of its economy and the history of the education system.

Venezuela is an upper-middle income country with challenges normally linked to low and low-middle income countries. Compared to Latin America, Venezuela performs quite well in terms of GDP, poverty and inequality. However, when we take a deeper look at the education sector, the country ranks amongst the bottom three in the region.

Regarding the labor market, the informal sector has grown during the 1980's and 1990's and currently is stagnant at 50% of the economy. The labor market imbalances might affect the returns to education. For instance, the premium wage for skilled workers can be explained not only by the higher level of education of those workers, but also by the lack of high skilled positions offered by the public and private sector.

Since Venezuela became an oil exporter country in the early 1900s, public education spending has been at the mercy of oil price fluctuations. However, throughout the years different governments have achieved good results on eradicating illiteracy and increasing the enrollment rate in primary education. At the present, they have not succeeded in improving the low enrollment rate in upper secondary school. The distribution of the education spending reveals that upper secondary public institutions receive much less than public primary and tertiary level. Because of the lack of information on costs of public education and the demand-supply gap of education by levels, we cannot argue that the education budget increase for public secondary institutions should be relatively more than for primary or tertiary institutions. In other words, this does not necessarily mean that either the public primary or tertiary institutions are relatively in better conditions than the secondary ones, or the secondary education has a bigger budget deficit than the two other levels. Yet, this might be a lead of the lack of investment at the secondary level. Definitely, it is worth studying further the efficiency of the public spending on education.

Regarding the quality of education, the only reliable study was performed by UNESCO-OREALC in 1990. They found that Venezuela occupied the 28th position among 31 countries; just above Nigeria, Zimbabwe and Botswana. They argued that a lack of capacity and quality among teachers and strong centralization of the Venezuelan education system.

Concerning the legal framework of the education system, the Organic Law of Education, pro-

mulgated in 1980, established 10 years of compulsory education instead of 7 years, which means that children are obliged to finish basic school. That covers kindergarten (1 year), primary (6 years), and lower secondary education (3 years). Also, a commonly used complementary policy to compulsory schooling, is a ban on child labor. According to the law for Child and Youth Protection and the Labor Law in Venezuela, the minimum age for admission to a non-hazardous employment or work is 14 years-old. In some cases, children under the minimum age are allowed to work as long as it does not jeopardize education activities. Since a 14 years-old child is, on average, about to finish lower secondary, this regulation can be linked to the fact that children abandon school after having achieved lower secondary level (10 years of education). Laws offer children the same rights than adult workers in terms of benefits like vacations and social security. Unlike eight hours working day for adults, children are not allowed to work more than six hours a day. Assuming the same conditions at work, the salary must be the same as for an adult, either in rural or urban work.

Chapter 2: Errors in variables in pseudo-panel data: a simple correction. The contribution of this chapter is twofold; one is related to the econometric approach, and the second concerns a more empirical question.

In the absence of panel data, we take advantage of the repeated cross-sections and construct pseudo-individuals. This means that individuals with specific characteristics, in this case region and age, are gathered together. Each group is called as a pseudo-individual. Our cohorts have 685 individuals, on average. The errors in variables problem generated by this type of construction is traditionally addressed by eliminating a proportion α of the measurement error from the empirical moments. We find that higher moments instrumental variables estimator is an alternative method to deal with errors in variables. Our estimator presents the added advantage that the underlying orthogonality conditions can be assessed using standard test statistics. By means of montecarlo experiments, we demonstrate that our higher moments estimator performs better than other available estimators, in terms of low level of bias and mean-squared error.

Our empirical application of the higher moments estimator is on the mincer equation. We find that one additional year of education in Venezuela increases wages by 15% on average. This figure is higher than those found in the empirical literature, even in the previous Venezuelan estimates. The difference between our estimates and those in the previous literature may be associated to the ability bias and errors in variables which are not accounted for in the latter.

Chapter 3: High school dropouts and the informal sector: Evidence from Venezuela. The contribution of this chapter relies on the role of the informal sector as a determinant of the school dropouts.

Venezuela, together with Mexico and Ecuador, ranks last among Latin American countries in terms of high school attendance. We find that a significant group of young people decide to abandon school by the end of the lower secondary education. Eventually, most children and youth dropouts will end up being low-skilled, low-wage workers.

This study sought to answer two questions. First, what are the determinants of the high school dropouts? Second, does a parent working as an informal employee have a impact on

the probability of dropping out of school? We find, as in previous literature, that there is a significant effect of the early parenthood on the probability of dropping out. Pregnant teenagers are more likely to drop out of high school. We also show that parents with low levels of education are more likely to take their children out of school.

An interesting finding and the contribution of this chapter is the significant effect of parents' type of work, in terms of labor protection, on the probability of dropping out. We show that if parents are unprotected workers or informal employees, their children are more likely to leave high school, either to work or to do nothing. We go in depth and also show that the probability of dropping out of school and remain idle increases when the parent of a child works as an informal employee, compared to the probability of continuing studying.

The results of most control variables corroborate the findings of previous applied research. For instance, families with few children and few dwellers in their households are associated with a low probability of withdrawing their children from school. In the same way, married parents and families who have not recently moved are positively correlated with children who attend school.

Chapter 4: Youth working in the informal sector in Brazil. The highlight of this chapter is to show the negative impact of hiring young workers on firms' productivity.

In Brazil, young workers are mainly absorbed by the informal sector, which is an entry point to the labor market for this age group. We find that unprotected workers start to work very early in life, on average at the age of 14. The aim of this chapter is to identify the effect of the age structure, particularly the impact of young workers, on a firm's performance. The key finding in this study is that an increase in the share of young informal workers will reduce a firm's productivity. Nevertheless, the share of older workers does not seem to have a different effect from the share of prime-age workers. In other words, the fact of hiring workers over 60 years old does not have a significant effect on a firm's performance in the informal sector.

Another main finding in this study, and well documented in previous literature, is the positive impact of the education of the owner on a firm's performance. We also find that the use of technology and if the owner is male have a positive and significant effect on a firm's productivity.

Summing up:

- Attending to school pays back; however, children continue to drop out of school.
- Parents working in the informal sector have a positive effect on school dropouts.
- Most young workers work in the informal sector. Furthermore, the probability of youth working in the informal sector increases when their parents own a small firm.
- Productivity is lower in small firms with a higher share of young workers.

Policy implications.

- The Venezuelan government should keep making efforts to improve access to school, as one of the main strategies to fight against income inequalities. Moreover, public education spending should be distributed efficiently between primary, secondary and tertiary level. More resources must be assigned to enhance high school access. Particularly, policies that aim at the construction or reconstruction of secondary level institutions, as well as at ameliorating the choices of public transport to school for those young people traveling long distances between home and school.
- Quality of education should be one of the priority of Venezuelan public policy. Authorities should reassess the content of the curriculum, particularly that of secondary level, to adapt it to the country's needs, and also to motivate youth to attend school. It means that they have to reevaluate the appropriateness of general and vocational programs, in order to avoid dropouts and, at the same time, facilitate youth's labor market insertion.
- Additionally, policy-makers should create, or enhance, performance-based incentives pay for teachers and for schools. The use of incentives mechanisms are needed to improve the quality of teaching and learning, such as a performance-based bonus. At the same time, teachers' evaluation should have a real impact on their salary, instead of being just another annual paperwork to fill out. Teachers should receive training courses in order to update skills and learn new and more effective teaching's methods.
- It is necessary that the government invest in national campaigns to prevent teenage pregnancy in order to avoid dropout. Early pregnancy prevention programs should be disseminated, particularly in public high schools which bring together young students from low-income families. Sexuality and family-life education programs have positive impact on reducing the risk of a pregnancy.
- This study does not intent to make any specific recommendations on how to deal with the informal sector. In a succinct manner, however, we must encourage the Brazilian and the Venezuelan governments to continue the fight against the phenomenon of informality. Policy makers should design an incentives mechanism for small firms to register in the social security institutions. Nonetheless, they must be careful not to discourage the creation of bigger firms. This measure will have, in the long-term, positive externalities on unprotected workers' children who are at higher risk of dropping out.
- The Labor and Education Ministries should join efforts to enforce child labor laws and to collect information on young dropouts. Governments should develop "Back to school" programs and use the information collected to target young workers, which are the potential beneficiaries, in a more efficient way.
- For those high school dropouts who really need to work to subsist and cope with life, we suggest that the governments offer training courses to increase their hard skills and, in turn, their employability. The content of the training courses must be pertinent and relevant to the needs of the labor demand. It should be designed by the employer and

the vocational training institute. This guarantees that the use of monetary resources in youth training will be spent in a more efficient way.

Further research

- It would be useful to find a way of dealing with sample selection bias as well as potential non-linearities in the study of returns to education within a pseudo-panel framework.
- It is necessary to determine the costs of education in order to properly calculate the returns to education.
- Similarly, it is key to understand what the idle children do; how many are handicapped? What do they do at home? How much time do they spend at home? Do they have a criminal record?
- There is not much evidence on the factors that affect the youth's decision of choosing informal jobs. Whether it is a voluntary decision or a consequence of the labor market exclusion. Variables such as previous school grades of the young worker and soft skills' measures must also be included in the future research.
- The study of the relationship between productivity and the age structure should be replicated for firms bigger than 5 employees. That means, it should contain the universe of firms, registered and unregistered, in the country.
- Additionally, it is important to understand the long-term effects of being an informal worker on the quality of the job and the salary.