IDB WORKING PAPER SERIES No. IDB-WP-319

Mobility and Entrepreneurship in Ecuador:

A Pseudo-Panel Approach

Xavier Ordeñana
Ramón Villa

July 2012

Inter-American Development Bank
Department of Research and Chief Economist

Mobility and Entrepreneurship in Ecuador:

A Pseudo-Panel Approach

Xavier Ordeñana*
Ramón Villa**

* ESPAE Graduate School of Management,
Escuela Superior Politécnica del Litoral
** Centro de Investigaciones Económicas,
Escuela Superior Politécnica del Litoral

Inter-American Development Bank

2012

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Ordeñana, Xavier.
Mobility and entrepreneurship in Ecuador: a pseudo-panel approach / Xavier Ordeñana, Ramón Villa.
p. cm. (IDB working paper series ; 319)
Includes bibliographical references.
1. Entrepreneurship—Ecuador. 2. Social mobility—Latin America. I. Villa, Ramón. II. Inter-American Development Bank. Research Dept. III. Title. IV. Series.

http://www.iadb.org

Corresponding author: Xavier Ordeñana (email: xordenan@espol.edu.ec)

Abstract[1]

Does entrepreneurship contribute to improving social mobility in Ecuador? This paper constructs a pseudo-panel to analyze the dynamic effect of entrepreneurship on Ecuadorian household incomes during the period 2002-2010. Using three estimation scenarios, the paper finds a significant level of unconditional mobility and an important effect of entrepreneurship (conditional mobility).

JEL Classification: J16, L26, M13
Keywords: Mobility, Pseudo-panel, Entrepreneurship, Ecuador

[1] This paper is part of the research project "Strengthening Mobility and Entrepreneurship: A Case for The Middle Classes" financed by the Inter-American Development Bank (IDB) Research Department. The authors thank Hugo Ñopo, Virginia Lasio, Gustavo Solorzano, and the participants in the Discussion Seminars for helpful comments.

1. Introduction

There seems to be a consensus among policymakers in Latin America that promoting entrepreneurship is a way to achieve economic development. Different programs around the region, such as *Emprende Ecuador* and Start-Up Chile, exemplify this idea. However, the economic effect of policies that promote entrepreneurship at the country level is still unclear.

In countries like Ecuador, where this study is focused, about one in five people is engaged in entrepreneurial activities, according to the Global Entrepreneurship Monitor Ecuador Report 2010. However, most entrepreneurial activity is highly ineffective at creating jobs; in fact, 98 percent of entrepreneurs created fewer than five jobs. Shane (2009) suggests that these activities are not contributing to economic growth and thus should not be promoted by the government. However, Amorós and Cristi (2010) find a positive effect on poverty reduction, which remains an important issue in Latin America, particularly in Ecuador.

This paper focuses on the effect of entrepreneurship on one economic variable: social mobility. Is there evidence that entrepreneurship increases a person's relative income? In order to determine whether such a correlation exists, we studied the evolution of household income over time, using panel data. Unfortunately, in Ecuador, attempts to build rotating panels have only recently begun to be undertaken, and we encountered several problems when attempting to construct a database using this information. We found many statistical inconsistencies, and only short time spans are available for the construction of the data series. Techniques have been developed to remedy these limitations, and several authors have established that panel data are not necessary for many commonly estimated dynamic models (Heckman and Rob, 1985; Deaton, 1985; and Moffitt, 1990).

The pseudo-panel approach, first introduced by Deaton (1985), consists of categorizing "similar" individuals in a number of cohorts, which can be constructed over time, and then treating the average values of the variables in the cohort as synthetic observations in a pseudo-panel. In Cuesta et al. (2011), a pseudo-panel approach is used to study the differences in mobility across the Latin American region. They find a high level of unconditional mobility and significant differences across countries. Canelas (2010) also uses this technique to measure poverty, inequality, and mobility in Ecuador and finds a decrease in poverty but persistent inequality between 2000 and 2009.

The rest of the paper is organized as follows: Section 2 presents the data treatment and the construction of the pseudo-panel. Section 3 presents the models of mobility: unconditional and conditional. Section 4 presents the results and the analysis, and Section 5 concludes.

2. Database Treatment and Documentation

The main objective of this paper is to describe the linkages between entrepreneurship and mobility. In this section, we first analyze the intragenerational mobility experienced by Ecuadorians (unconditional mobility) and approach the potential role of entrepreneurship in improving mobility (conditional mobility). Given that individual data panels are nonexistent in Ecuador, the use of pseudo-panels was required. We start by explaining how we constructed the instrument.

2.1 Database Treatment

The data used for the construction and estimation of the pseudo-panel were obtained from the National Employment and Unemployment Survey (ENEMDU for its Spanish acronym) collected by the National Institute of Statistics and Census (INEC). The census is taken at the national level in November every 10 years, and the results are processed and made public the following month. This data collection methodology has been applied since 2003. In some years, national census data are presented in May or June. To avoid any seasonal bias due to variations in the levels of economic activity at different times of the year, only those surveys presented in December were used.

The database used to estimate the pseudo-panel is constructed as a series of independent cross-sections, one for each period analyzed. To determine the period in which the pseudo panel ought to be constructed, it is first necessary to examine the changes made by INEC in the methodology for both the determination of the sample and the estimation of the relevant variables. There are two important changes made in the last decade in the ENEMDU's methodology which are so significant that, without taking them in account, any estimation made for the whole period would suffer from serious bias. The first occurred in December of 2003, before which time only the urban population was analyzed. The definition of what constitutes an urban settlement was also changed in 2003 to include centers with more than 2,000 inhabitants rather than the 5,000 used earlier. The definitions of several labor variables were also modified, and others were included. The second set of changes introduced by INEC in September 2007 consisted of several modifications in labor market definitions and classifications, but there were no significant changes in the variables used for this study (even though income estimation underwent some changes, which will be discussed below).

Because of the loss of information that would result from the construction of a larger panel (in the time dimension), only the 2003-2010 period was studied. To maintain the consistency of the data for the period analyzed, special attention was paid to the changes in the methodology, variable classification, and labels used. The method used for the estimation of individual income was also changed, and new income criteria were introduced after 2007. In addition, even if a survey question was not modified, some of the responses were changed, which in some cases made it impossible to use the variable for the whole period. To account for all of these issues, the income series was constructed using the previous methodology, and all of the other variables included were previously processed to ensure their statistical comparability. Using this methodology, income is calculated as any payment, either monetary or in-kind, received by the individual on a regular basis (daily, weekly, or monthly). Two types of income sources were considered: income generated by work and income derived from capital, investment, contractual, or non-contractual transfers. A monthly income series was then constructed by adding all sources of personal revenue.

The ENEMDUs were processed in order to obtain the pertinent variables at household level, as the information relevant for this study on an individual level is collected by INEC. Data mining techniques were used and, with the use of Structured Query Language (SQL), income and other covariates were aggregated at the desired level.

The first key concept in determining the effect of entrepreneurship on income mobility is the definition of which households are to be considered entrepreneurs. The focus of the study is those households that are entrepreneurs by choice rather than entrepreneurs due to lack of options (a group that is difficult to correctly identify considering the scant information available). In order to reduce the probability of error at the moment of classification, only those households in which at least one member currently employs other workers are considered entrepreneurs.

2.2. Construction of the Pseudo-panel

In order to analyze the dynamic nature of income mobility, household income needs to be observed over time. Given the absence of panel data, a pseudo-panel must be constructed. The pseudo-panel approach consists of categorizing "similar" individuals into a number of cohorts, which can be constructed over time, and then treating the average values of the variables in the cohort as synthetic observations in a pseudo-panel. Even though this approach has many limitations compared to real panel data, it reduces several problems characteristic of real panel data. First, it greatly diminishes the problem of sample attrition, hence allowing the possibility for the construction of larger panels in the time dimension. A second contribution is that, as the observations are obtained by averaging different observations in a cohort, the possibility of measurement error is greatly reduced (provided the cohorts are adequately constructed).

The efficiency and consistency of the estimators depends, among other things, on the criterion used for the construction of the different cohorts and the asymptotic nature of the data assumed. Several of the requirements for the consistency of pseudo-panel estimation are discussed by Verbeek and Nijman (1992), who recommend that the choice of the variables for the discrimination of the cohorts in the sample should follow three criteria:

- The cohorts are chosen such that the unconditional probability of being in a particular cohort is the same for all cohorts.
- The variables chosen should be constant over time for each individual, because individuals cannot move from one cohort to another. This maintains the independence of the different cohort observations.
- These variables should be observed for all individuals in the sample. This could be remedied by the use of unbalanced panel methods, but due to the short time span of the constructed pseudo-panel, this alternative is not considered.

Following these assumptions, cohorts were constructed using gender and date of birth of the household head. To determine the number of cohorts to be constructed, first the distribution of the date of birth variable was tested with conventional goodness of fit methods, but no traditional distribution seemed to adjust the data correctly. To ensure relatively similar probabilities of belonging to a birth cohort, the aggregated data for year of birth for the eight periods were divided into deciles. This avoids the possibility that a cohort in a given period becomes too small to provide an accurate estimation of its true characteristics. After considering the weights of the observations due to sample stratification, the following deciles were obtained:

Table 1. Date of Birth (z_i^1) Cohorts Criterion

Date of Birth (z_i^1) cohorts criterion
$z_i^1 < 1934$
$1934 \leq z_i^1 < 1942$
$1942 \leq z_i^1 < 1949$
$1949 \leq z_i^1 < 1954$
$1954 \leq z_i^1 < 1958$
$1958 \leq z_i^1 < 1962$
$1962 \leq z_i^1 < 1966$
$1966 \leq z_i^1 < 1971$
$1971 \leq z_i^1 < 1977$
$1977 \leq z_i^1$

Source: Authors' calculations.

Another criterion used to determine the number of cohorts is the gender of the household head (z_i^2). The conjunction of the two variables (considering the criterion proposed for the date of birth) results in 20 cohorts per year and 160 synthetic observations in the pseudo-panel. The distribution of the observations and their corresponding expanded population values (by the use of sampling weights) in each of the categories explained are presented in the Appendix. The inclusion of the gender of the household head as a determinant for the conformation of the cohorts makes the probability of belonging to a cohort uneven for most cohorts (as male household heads are more frequently found). As the synthetic observations are calculated with different sample sizes, a systematic heteroskedasticity component is introduced to the error. The methods to correct this problem are discussed in Gurgand et al. (1997). This problem becomes less relevant in cohorts constructed with a large number of observations, since the variance of the mean approaches zero as this number tends to infinity.

2.3 Treatment of Outliers

The household income series each year is irregular, as its standard deviation is between 2 and 4 times the mean. The asymmetries presented by the data may complicate the estimation of any inference model applied. This is also maintained at a cohort level, and important differences in variances between each cohort average are observed. These differences make the heteroskedasticity component, described in the previous section, more important. To account for this problem, data mining techniques are applied to determine and exclude outliers. A median of absolute deviations (MAD) approach is used to determine outliers in each cohort as, due to the nature of the series, the median is a better central tendency measure than the mean. Under this scheme, the following univariant filter was applied to each observation, and observations that satisfy this restriction are considered outliers:

$$\frac{\left|x_{ct}^{i} - median_{ct}^{j} x_{ct}^{j}\right|}{MAD_{ct}} > 10^{2}$$

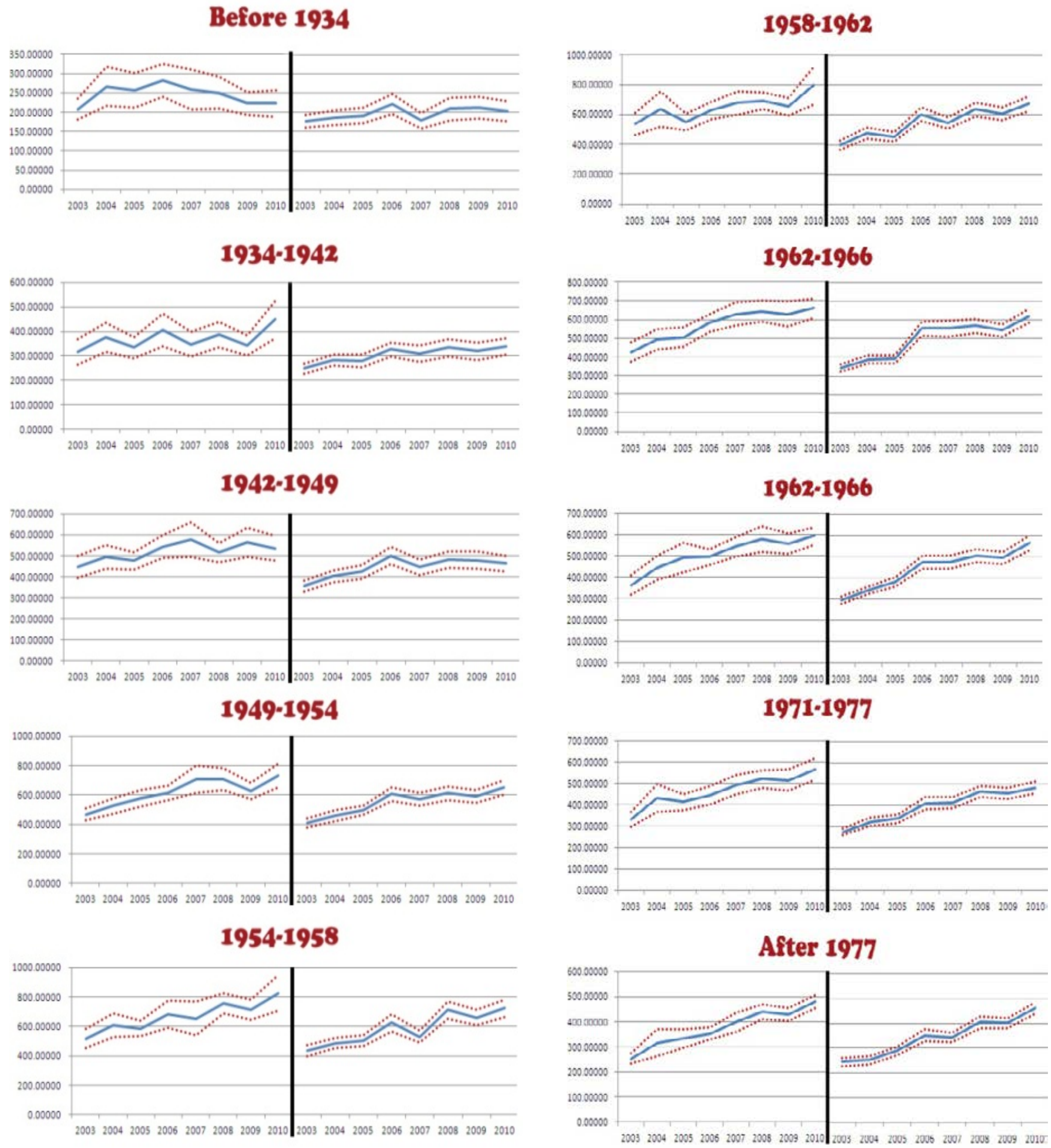
As shown in the formula, the method is applied at a cohort level for each period. Approximately 1.2 percent of the sample was determined to be an outlier. In Figures 1 and 2, the average income estimated for male and female cohorts is presented before and after the MAD treatment is applied. The red dotted bands denote the 95 percent confidence interval for the estimated means (solid blue line) for the period analyzed. The graph to the left of each vertical black line corresponds to the estimated cohort mean of household incomes before the univariant filter is applied, and to the right the results excluding outliers are presented.

In observing the two figures, it is noteworthy that the error bands on male cohorts are smaller than those observed for female cohorts, possibly due to the lower number of observations in the female cohorts. But even after considering those wider confidence intervals, for most of the years studied and most of the birth cohorts, male-headed households experience a significantly higher income than female-headed households (at a 95 percent confidence level). Without the application of the MAD univariant filter, some birth cohorts exhibit very irregular behavior, and in the case of female cohorts some of the error bands explode (raising serious concerns about the validity of those estimations in a pseudo-panel context). However, once

[2] Traditionally the tolerance criterion is set at 4.5, but this resulted in the loss of 12 percent of the sample, including an important percentage of entrepreneurs.

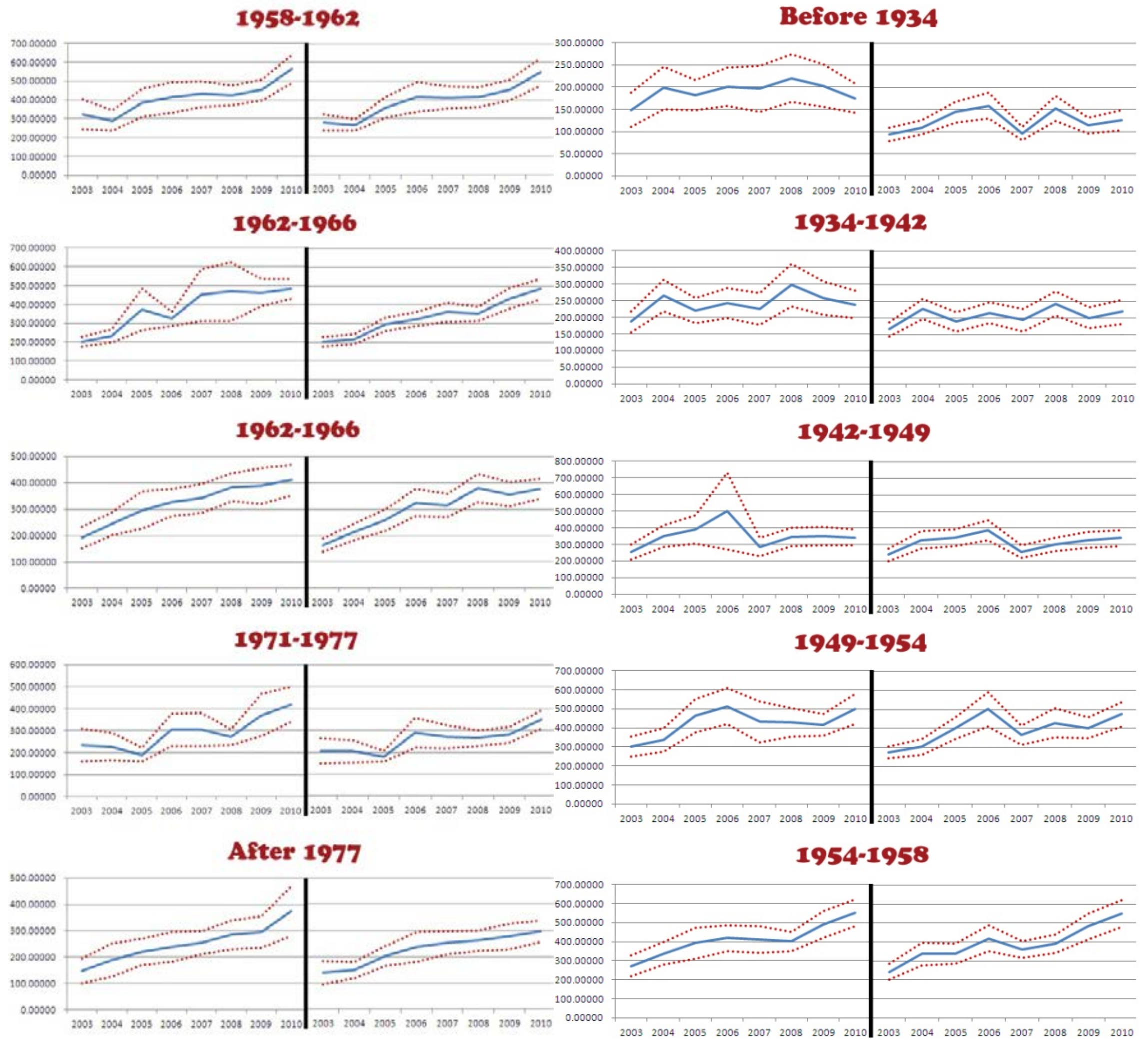
outliers are excluded, the error bands decrease considerably and the behavior of the income mean becomes smoother.

Figure 1. Average Income Before and After MAD Treatment for Male Cohorts



Source: Authors' calculations.

Figure 2. Average Income Before and After MAD Treatment for Female Cohorts



Source: Authors' calculations.

Additional descriptive statistics regarding the cohorts constructed after the outlier treatment are presented next. For most of the male cohorts, the percentage of households located in urban areas is significantly lower than the one presented by female cohorts. This might be due to a more traditional family life in rural areas, which makes single-parent families, or female-headed households, a less common occurrence. It is also important to note that the number of entrepreneur households headed by women is significantly lower than those headed by men. This is accentuated by the fact that almost half of all female entrepreneur households include a member, different from the household head, who owns a business.

Table 2. Urban Ratio

	Male Households								Female Households							
	2003	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
$z_i^1 < 1934$	56%	58%	56%	54%	53%	53%	53%	54%	59%	63%	64%	66%	62%	63%	62%	63%
$1934 \leq z_i^1 < 1942$	59%	61%	55%	56%	61%	60%	60%	59%	68%	67%	62%	63%	65%	67%	66%	71%
$1942 \leq z_i^1 < 1949$	64%	64%	62%	61%	61%	61%	59%	58%	71%	74%	72%	73%	71%	72%	72%	72%
$1949 \leq z_i^1 < 1954$	66%	67%	65%	66%	66%	68%	63%	63%	75%	75%	79%	79%	77%	77%	74%	76%
$1954 \leq z_i^1 < 1958$	70%	70%	67%	68%	67%	68%	67%	66%	81%	81%	80%	77%	80%	75%	74%	73%
$1958 \leq z_i^1 < 1962$	69%	68%	70%	71%	67%	69%	63%	66%	78%	80%	78%	76%	83%	84%	78%	79%
$1962 \leq z_i^1 < 1966$	68%	69%	68%	73%	68%	67%	67%	66%	72%	76%	78%	79%	78%	76%	77%	77%
$1966 \leq z_i^1 < 1971$	66%	67%	68%	66%	67%	66%	66%	66%	77%	76%	78%	82%	79%	78%	81%	80%
$1971 \leq z_i^1 < 1977$	67%	69%	70%	69%	68%	67%	70%	68%	76%	73%	80%	74%	77%	71%	82%	82%
$1977 \leq z_i^1$	71%	69%	71%	75%	70%	70%	73%	73%	84%	84%	86%	77%	84%	82%	86%	83%

Source: Authors' calculations.

Table 3. Entrepreneurship Ratio

	Male Households								Female Households							
	2003	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
$z_i^1 < 1934$	6%	11%	9%	9%	5%	6%	4%	3%	3%	3%	2%	4%	2%	3%	2%	2%
$1934 \leq z_i^1 < 1942$	9%	13%	11%	11%	8%	9%	7%	5%	5%	8%	5%	7%	5%	5%	3%	2%
$1942 \leq z_i^1 < 1949$	9%	13%	13%	13%	11%	13%	7%	7%	4%	10%	5%	6%	5%	5%	3%	4%
$1949 \leq z_i^1 < 1954$	9%	12%	12%	14%	12%	11%	9%	7%	7%	7%	7%	10%	6%	6%	4%	3%
$1954 \leq z_i^1 < 1958$	9%	15%	12%	16%	8%	11%	10%	8%	7%	7%	5%	7%	6%	5%	6%	3%
$1958 \leq z_i^1 < 1962$	8%	13%	12%	12%	12%	10%	10%	7%	5%	4%	5%	5%	3%	5%	5%	4%
$1962 \leq z_i^1 < 1966$	8%	12%	10%	13%	10%	11%	7%	8%	3%	7%	6%	6%	5%	6%	3%	1%
$1966 \leq z_i^1 < 1971$	5%	10%	10%	10%	10%	10%	8%	9%	5%	6%	4%	6%	6%	3%	2%	3%
$1971 \leq z_i^1 < 1977$	6%	9%	8%	9%	9%	8%	6%	5%	3%	5%	4%	7%	3%	4%	1%	3%
$1977 \leq z_i^1$	5%	6%	5%	6%	3%	4%	4%	2%	3%	3%	2%	3%	2%	5%	1%	1%

Source: Authors' calculations.

Table 4. Percentage of Entrepreneurs who are Household Heads

	Male-headed Households								Female-headed Households							
	2003	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
$z_i^1 < 1934$	78%	69%	76%	71%	72%	76%	62%	68%	37%	52%	30%	49%	58%	40%	21%	42%
$1934 \le z_i^1 < 1942$	76%	88%	88%	87%	72%	77%	70%	77%	26%	35%	34%	52%	58%	45%	49%	39%
$1942 \le z_i^1 < 1949$	73%	82%	85%	79%	80%	80%	82%	69%	53%	66%	70%	50%	63%	80%	56%	59%
$1949 \le z_i^1 < 1954$	78%	81%	78%	82%	86%	79%	87%	80%	84%	74%	75%	66%	79%	44%	39%	40%
$1954 \le z_i^1 < 1958$	77%	80%	80%	78%	82%	89%	82%	90%	61%	81%	57%	58%	84%	80%	81%	66%
$1958 \le z_i^1 < 1962$	85%	82%	90%	81%	88%	84%	78%	81%	61%	84%	72%	85%	74%	84%	88%	77%
$1962 \le z_i^1 < 1966$	90%	84%	83%	82%	80%	92%	94%	92%	55%	79%	82%	96%	71%	90%	95%	87%
$1966 \le z_i^1 < 1971$	85%	79%	89%	85%	88%	93%	89%	86%	66%	67%	60%	86%	70%	96%	83%	97%
$1971 \le z_i^1 < 1977$	84%	86%	83%	87%	86%	86%	86%	82%	95%	93%	75%	87%	100%	82%	87%	90%
$1977 \le z_i^1$	79%	91%	79%	98%	90%	77%	82%	90%	86%	60%	100%	70%	79%	62%	74%	87%

Source: Authors' calculations.

3. Income Mobility and Entrepreneurship

3.1 The Unconditional Model

Income mobility presents a measure of the relationship between past and present income. This relationship can be represented by the following equation:

$$y_{i,t} = \beta y_{i,t-1} + \psi_{i,t}$$

where $y_{i,t}$ represents the income of household i in period t, $\psi_{i,t}$ is a composite error term and β is a measure of the unconditional income convergence (β=0 represents a perfect income mobility and β=1 represents an absence of income mobility or perfect convergence).

Since information for the same individual is not available in the different years sampled, a pseudo panel approach was taken in order to estimate the β parameter. As previously mentioned the synthetic observations are constructed with the average values of the household observations in each cohort. The dependent variable used for the estimation of the model is the log of the average household's income for the cohort and the period studied, which makes the β parameter a measure of the elasticity of past and present income. The respective cohort model can be expressed as follows:

$$\ln(\bar{y}_{c,t}) = \beta_1 \ln(\bar{y}_{c,t-1}) + \bar{\psi}_{c,t}$$

For ease of exposition, the logarithm of the income variable for each cohort will still be represented as $\bar{y}_{c,t}$.

Fields and Ok (1999) demonstrate that this measure of income mobility is the only one to have a set of desired properties (scale invariance, symmetry, multiplicability, and additive separability).

3.2 The Conditional-Entrepreneurship Model

As shown in Cuesta et al. (2011), the measure of unconditional income mobility tends to underestimate the true mobility experienced by households in an economy. The effect of household covariates on income mobility can be estimated by an extension of the previous model:

$$\bar{y}_{c,t} = \beta_1\bar{y}_{c,t-1} + \beta_2\bar{\varepsilon}_{c,t-1}\bar{y}_{c,t-1} + \beta_3\bar{\varepsilon}_{c,t} + \beta_4 f_c\bar{y}_{c,t-1} + \beta_5\overline{X^1}_{c(t),t} + \beta_6\overline{X^2}_{c(t)} + \bar{\psi}_{c,t}$$

where:

$\bar{y}_{c,t}$= the Neperian logarithm of the average household income of cohort c in period t. The income variable is previously deflated considering the purchasing power parity (PPP) index reported by the World Bank.

$\bar{\varepsilon}_{c,t}$= the proportion of households which are considered to be entrepreneurs in cohort c and period t. This regressor is believed to be predetermined, a concept that will be clarified below.

f_c= a dichotomous variable which takes the value of 1 if c is a female cohort and 0 otherwise.

$\overline{X^1}_{c(t),t}$= a set of time-variant household covariate averages for cohort c and period t. These regressors are believed to be exogenous.

$\overline{X^2}_{c(t)}$= a set of time-invariant household covariate averages for cohort c. The term $c(t)$ is included to denote that the average is taken on period t and, as the sample mean is an error-driven measurement, differences may be observed over time. These variables need not be truly time-invariant, but the rate at which they vary may be too subtle to be observed in one period (variables that present a staircase behavior and which require more than one period to register a change fall in this category). An obvious example is the gender of the household head, due to the construction of the cohorts. No assumptions about the relationship of these covariates with the error term are made.

β_5 and β_6= vectors of the pseudo-elasticity of said covariates on present incomes.

$\bar{\psi}_{c,t}$= a composite error term determined by the next equation,

$$\bar{\psi}_{c,t} = \bar{\lambda}_c + \bar{u}_{c,t}$$

where $\bar{\lambda}_c$ is a time-invariant intrinsic cohort "c" component which cannot be observed, and $\bar{u}_{c,t}$ is an error term. The different possible assumptions for these terms are considered below.

The total measure of income mobility can be expressed as:

$$\frac{\delta\bar{y}_{c,t}}{\delta\bar{y}_{c,t-1}} = \beta_1 + \beta_2\bar{\varepsilon}_{c,t-1} + \beta_4 f_c$$

where

β_1= the part of income convergence that is only explained by past income.

β_2= the effect that a marginal increase in the entrepreneurship percentage in the cohort has on its income convergence.

β_4= the variation in income convergence experienced by female cohorts.

If the number of observations in each cohort is sufficiently large, $\bar{y}_{c,t}$, $\bar{\varepsilon}_{c,t}$, $\overline{X^1}_{c(t),t}$ and $\overline{X^2}_{c(t)}$ will provide accurate estimators of the true cohort means. An unbiased estimator of the relevant model parameters can be obtained through traditional dynamic panel estimation methods (like the two-step least squares estimation or a more efficient estimator obtained by GMM methods).

3.3 Estimation by Dynamic Panel Methods

The dynamic nature of the model presented results in a series of complications when estimating the autoregressive parameters. Because the estimation centers on following the same cohorts over time, as previously indicated, an unobserved fixed component is introduced into the equation. The following expression is obtained by replacing the composite error term with its determinants:

$$\bar{y}_{c,t} = \beta_1\bar{y}_{c,t-1} + \beta_2\bar{\varepsilon}_{c,t-1}\bar{y}_{c,t-1} + \beta_3\bar{\varepsilon}_{c,t} + \beta_4 f_c\bar{y}_{c,t-1} + \beta_5\overline{X^1}_{c(t),t} + \beta_6\overline{X^2}_{c(t)} + \bar{\lambda}_c + \bar{u}_{c,t} \quad (1)$$

By expressing this equation in the $\mathrm{t}-1$ period, one can show that the unobserved component is correlated with previous income (this is also likely to hold for the other contemporaneous covariates), which introduces a missing variable problem that would render standard estimation methods biased. However, several assumptions can be made about the nature of the error components (depending on the licenses the researcher is willing to take), and from which different consistent estimation methods can be derived.

The following assumptions are made about the cohort covariates and the error components:

- The error term is believed to be serially uncorrelated,[3]

$$E\left(\bar{\mu}_{c,t}\bar{\mu}_{c,t-1}\right) = 0 \qquad (2)$$

- Entrepreneurship is believed to be predetermined,

[3] This hypothesis can be tested and if the error's autocorrelation cannot be rejected at a given confidence level, then other considerations, which are later specified, must be taken into account.

$$E(\bar{\varepsilon}_{c,t-s}\bar{u}_{c,t}) = 0 \qquad \forall \quad s > 0 \qquad (3)$$

This means that the error term is correlated with contemporaneous or future entrepreneurship levels. This assumption is made because entrepreneurship is believed to be endogenous, as it is difficult to establish a causal relationship between this variable and the contemporaneous average household income for each cohort.

- The time-variant cohort covariates are assumed to be exogenous, but not strictly so.

$$\left(\overline{X^1}_{c(t),t-s}\bar{u}_{c,t}\right) = 0 \qquad \forall \;\; s \geq 0 \qquad (4)$$

- No assumptions are made about the nature of the time-invariant covariates.
- Building on the belief that the errors are serially uncorrelated, a natural supposition is made:

$$\mathrm{E}(\bar{\mu}_{c,t}\bar{y}_{c,t-1}) = 0 \qquad (5)$$

The fact that the model is estimated at a cohort level provides an alternative to conventional dynamic panel estimation methods. Moreover, if it is assumed that:

$$\lambda_i \sim N(0, \sigma^2) \quad \forall \; i \in c \;\wedge\; \forall \;\; c \in C \qquad (6)$$

then the intrinsic cohort average effect will be asymptotically equal to 0 and, with large cohort dimensions, the missing variable problem is solved. Next, instrumental variables are needed to account for the endogeneity of entrepreneurship, after which a two-stage OLS estimation method will provide consistent estimators (even though a robust estimation is recommended because of the existence of an important heteroskedastic factor caused by the use of sample averages as observations). But assumption (6) implies that no intrinsic cohort effect exists, which raises serious questions about the validity of the constructed cohorts as units of study. If one is willing to assume that the intrinsic cohort effect is evened out, then how many important effects suffer the same attrition? Cohorts should be constructed on the basis of homogeneity of the individuals within the group, so as to avoid the loss of essential information and to make mean estimators more significant. If it is assumed that this is not so, then even if the estimations are consistent, the conclusions derived from the study will not be very relevant. Hence, the existence of idiosyncratic cohort effects should be taken as an indicator of accurately constructed cohorts. On account of these issues, and the belief that the cohorts defined for this study are correctly specified, assumption (6) is relaxed.

Thus, estimating the proposed model without accounting for the unobserved cohort effect will result in biased estimators and ordinary least squares (OLS) methods will result in a positively biased estimation (Nickell, 1981). The fact that OLS is expected to provide a positive bias in the estimator if the idiosyncratic component exists is a useful check on the validity of the proposed cohorts studied.

To draw out the fixed effect of the error term, a possible solution would be to apply a mean deviation transformation to equation (1). Under this transformation, the equation variables are expressed as a deviation from their period mean, thus eliminating the time-invariant cohort fixed effect and any other fixed variable (within estimation). But in panels with a short time span, the transformed autoregressive term $(\bar{y}_{c,t-1}{}^{*} = \bar{y}_{c,t-1} - \frac{1}{T-1}(\bar{y}_{c,2} + \cdots + \bar{y}_{c,T}))$ is now negatively correlated with the transformed error term $(\bar{\mu}_{c,t}{}^{*} = \bar{\mu}_{c,t} - \frac{1}{T-1}(\bar{\mu}_{c,2} + \cdots + \bar{\mu}_{c,T}))$ as the $\bar{y}_{c,t-1}$ term correlates negatively with the $-\frac{1}{T-1}\bar{\mu}_{c,t-1}$ term in the transformed error. This bias decreases as the time frame T becomes larger; hence, the within estimators are asymptotically consistent in the time dimension. Bond (2002) points out that these different directions in the bias of both OLS and within estimators provide useful bounds on the accuracy of any other theoretical superior estimator proposed.

Another approach that eliminates the unobserved fixed effects is to apply first differences to equation (1) as shown next:

$$\Delta\bar{y}_{c,t} = \beta_1\Delta\bar{y}_{c,t-1} + \beta_2\Delta(\bar{\varepsilon}_{c,t-1}\bar{y}_{c,t-1}) + \beta_3\Delta\bar{\varepsilon}_{c,t} + \beta_4\Delta(f_c\bar{y}_{c,t-1}) + \beta_5\Delta\overline{X^1}_{c(t),t} + \Delta\bar{\mu}_{c,t} \quad (7)$$

where

$$\Delta\bar{y}_{c,t-1} = \bar{y}_{c,t-1} - \bar{y}_{c,t-2}$$

As first differences are applied, any time-invariant regressor is also eliminated and can no longer be estimated. It can easily be shown that the $\bar{y}_{c,t-1}$ component in the transformed autoregressive term is correlated with the $\bar{\mu}_{c,t-1}$ component in the transformed error term by the dynamic nature of the model. Following the method proposed Holtz-Eakin, Newey, and Rosen (1988) and continued in Arellano and Bond (1991), a generalized method of moments (Hansen 1982) approach is taken to account for the endogeneity presented in (7). Building on the first moment conditions given by (3), (4) and (5), a set of instruments is available to account for this problem. As the number of available instruments is quadratic in the time dimension of the panel, many problems can be encountered in finite samples. Roodman (2006) presents various methods available to account for the over identification problem derived from large instrument matrixes.

Another possible estimation method would be the use of system GMM, presented in Blundell and Bond (1998).[4] The advantage of system GMM is that it permits the estimation of time-invariant variable parameters by the inclusion of both level and difference instrument sets, but it also requires the use of large instrument matrices. Due to the instrument proliferation relative to the small sample size that would occur if the previous estimation method were used, the traditional difference GMM approach was taken. But, as previously noted, this eliminates from the estimation any time-invariant regressor.

4. Results

The results of the estimation of the conditional and unconditional income mobility models are presented. As previously explained, the models are also estimated by the use of OLS and within estimation to obtain reasonable bounds for the autoregressive parameter and to evaluate the results obtained by GMM.

4.1 The Unconditional Model

The unconditional model was estimated using difference GMM procedures. As previously indicated, it is expected that a heteroskedastic component exists in the error term; hence, a robust correction in the variance and covariance matrix of the errors is applied. A total of 140 observations were used in the estimation of the model (since one period is lost due to the application of first differences), and a collapsed instrument matrix[5] containing a total of two instruments was used.

[4] This method requires an additional assumption that the deviation of the first observation from the steady state is uncorrelated with the fixed effect.
[5] See Roodman (2006).

Table 5. Unconditional Model Results

	Coefficients		
	Within Estimation	**OLS Estimation**	**GMM Estimation**
$\bar{y}_{c,t-1}$	0.476379*** (0.05998)	0.922871*** (0.02993)	0.864752*** (0.08274)
Constant	3.438388*** (0.38695)	0.559065*** (0.19341)	-

Source: Authors' calculations.
Note: *** Significant at 1%

It can be observed that the GMM estimator for the autoregressive term is inside the bounds given by the OLS and within estimators. The total unconditional convergence is estimated at 0.86, slightly below the 0.9 obtained for the rest of Latin America (Ñopo et al., 2011). As expected, the OLS estimator is larger than the one provided by GMM methods. This indicates that the cohorts are adequately constructed, as the unobserved cohort fixed effect is present (hence the OLS estimation is positively biased). But due to the robust estimation method applied to the GMM estimation, its 95 percent confidence interval is wide and includes the value estimated by OLS; thus, the difference observed is not significant. Additional relevant statistics for GMM estimation are presented in Table 6.

Table 6. Unconditional Model Relevant Statistics

	Statistic	**P-Value**
Arellano-Bond Test	2.95	0.003
Sargan Test*	0.03	0.870
Hansen Test *	0.05	0.824

Source: Authors' calculations
Note: *The statistic has a chi-squared distribution with 1 degree of freedom.

The Arellano-Bond test for autocorrelation is used to determine if the errors are serially uncorrelated (assumption (2)). The null hypothesis is that no second-order autocorrelation is present in the transformed error component (which translates to first order autocorrelation in level equation (1)); and as the hypothesis is not rejected, the assumption that there is no autocorrelation present in the error holds.

The Sargan and Hansen tests are used to determine the quality of the instrument matrix used. The null hypothesis is that the instruments are not exogenous (which means that instrument matrix is not valid). Thus, for the consistency of the GMM estimators, this test must be rejected. The Sargan test is not robust, and the Hansen statistic is a robust measure but its validity is reduced as the number of instruments used increases (a frailty not shared by the Sargan statistic). Therefore high p-values obtained for this test are generally construed as a warning of misspecification. However, taking into account the small number of instruments used and the fact that the Sargan test also presents high p-values, the hypothesis of exogeneity of the instruments is confidently rejected.

4.2 The Conditional-Entrepreneurship Model

Only one cohort covariant was considered for the $\overline{X^1}_{c(t),t}$ vector: the average number of residents whose income represents more than 25 percent of their household income for cohort c in period t. This measure is considered to be much more volatile in time than any other household covariate (including age of the household head, number of residents per household, education level for older cohorts, etc.), and thus would be able to survive the first differences taken. The following results were obtained:

Table 7. The Conditional Model: Results

	Coefficients		
	OLS Estimation	**Within Estimation**	**GMM Estimation**
$\bar{y}_{c,t-1}$	0.944132*** (0.0385)	0.651593*** (0.10039)	0.7739*** (0.12009)
$\bar{\varepsilon}_{c,t-1}\bar{y}_{c,t-1}$	-0.369551*** (0.10023)	-0.16449* (0.08786)	-0.24995* (0.12764)
$f_c\bar{y}_{c,t-1}$	0.002318 (0.00608)	-0.27818** (0.10851)	-0.297292* (0.15998)
$\overline{X^1}_{c(t),t}$	0.054548 (0.09872)	1.423304*** (0.20309)	2.237625*** (0.44658)
$\bar{\varepsilon}_{c,t}$	3.087679*** (0.56992)	0.891305 (0.57498)	1.793539** (0.84013)
Constant	0.292977 (0.25073)	1.030288** (0.46863)	-

Source: Authors' calculations.
Note: * Significant at 10%;** Significant at 5%;*** Significant at 1%.

As occurred in the unconditional model, the first autoregressive term is inside the bounds given by the other estimation methods. The difference between the OLS and GMM estimators is more notorious in the conditional model, but it is still not significant at a 95 percent level. This will be very difficult to accomplish considering the small size of the pseudo-panel and the robust estimations made.

The total income mobility can be expressed as follows:

$$\frac{\delta\bar{y}_{c,t}}{\delta\bar{y}_{c,t-1}} = 0.7739 - 0.25\bar{\varepsilon}_{c,t-1} - 0.2973f_c$$

This means that an increase of 1 percent in the level of entrepreneurship in a cohort in period $t-1$, translates into an increase of 0.0025 in the income mobility of said cohort in period t. An interesting result is that female cohorts experience significantly higher income mobility, as their base convergence level can be expressed as (since $f_c = 1$ for female cohorts):

$$\frac{\delta \bar{y}_{c,t}}{\delta \bar{y}_{c,t-1}} = 0.4766 - 0.25\bar{\varepsilon}_{c,t-1}$$

To complement the reduction of income convergence that occurs with an increase in the percentage of entrepreneurs in a given cohort, this increase also positively affects future income.

Additional relevant statistics for GMM estimation are presented below:

Table 8. Conditional Model: Relevant Statistics

	Statistic	P-Value
Arellano-Bond Test	1.59	0.111
Sargan Test*	33.4	0.001
Hansen Test *	17.15	0.192

Source: Authors' calculations.
Note: *The statistic has a chi-squared distribution with 13 degrees of freedom.

The Arellano-Bond test is rejected at a 10 percent level, so first-order correlation in the error term is to be expected. On account of this issue, assumptions (2), (3), (4) and (5) need to be corrected as follows:

$$E(\bar{\mu}_{c,t}\bar{X}_{c(t),t-s}) = 0 \;\; \forall\, s \geq 1 \quad E\left(\bar{\mu}_{c,t}\bar{y}_{c,t-2}\right) = 0 \;\forall\, t \geq 3$$

$$E\left(\bar{\mu}_{c,t}\bar{\mu}_{c,t-s}\right) = 0 \quad \forall \quad s > 0$$

$$E(\bar{\varepsilon}_{c,t-s}\bar{u}_{c,t}) = 0 \quad \forall \quad s > 1$$

$$\left(\overline{X^1}_{c(t),t-s}\bar{u}_{c,t}\right) = 0 \quad \forall \quad s \geq 1$$

$$\mathrm{E}\left(\bar{\mu}_{c,t}\bar{y}_{c,t-2}\right) = 0 \quad \forall \quad t \geq 3$$

The following instrument matrix was constructed:

For every $t \geq 4$

- $\bar{y}_{c,t-s} \;\forall\, s \geq 3$,
- $\bar{\varepsilon}_{c,t-s}\bar{y}_{c,t-s} \;\forall\, 3 \leq s \geq 4$,
- $f_c\bar{y}_{c,t-1} \;\forall\, 3 \leq s \geq 5$

- $\bar{\varepsilon}_{c,t-2}$
- $\overline{X^1}_{c(t),t-1}$

As was done with the unconditional model, the instrument matrix constructed was collapsed to reduce the number of instruments without loss of information. This resulted in an instrument matrix with a total of 18 instruments and 120 observations available for the estimation of the model.

The Sargan statistic is not robust and, due to the heteroskedasticity of the error term (guaranteed by the pseudo-panel approach taken), the fact that the null is not rejected should not be alarming. The Hansen test is rejected at a 10 percent significance level, but the value is not high enough as to generate doubt (an empirical rule of thumb is to consider p-values approaching 0.3 or higher as suspicious) so no miss specification signs are present.

5. Conclusion

There have been few attempts to measure the determinants of income mobility in Latin America, due mainly to the limited information collected in those countries and the lack of panel data. This is especially true for Ecuador, where few such studies have been undertaken. But building on recently developed methods of estimation without the use of panel data and other efforts to reduce the dependence of consistent autoregressive estimators on a large time frame in a dynamic panel scheme, estimations of income mobility and its determinants were achieved for the Ecuadorian case.

Through the construction of a pseudo-panel and the application of difference GMM estimation methods, unconditional and conditional income mobility models were estimated. As expected, and in accordance with other empirical studies, the unconditional model tends to underestimate true income mobility, as there are other factors not included in the equation that explain future income and are correlated to previous income (there is a missing variable problem and estimations are biased). The inclusion of other cohort covariates in the proposed model reduces the bias and permits the analysis of other determinants of income mobility and future income (these might be affected by public economic policy). One of those factors, and the one of particular interest for this paper, is the percentage of entrepreneurs in the different cohorts.

The results of the GMM estimations revealed that entrepreneurship not only reduces income convergence by 0.0025 but also increases future average cohort income by 1.79 percent per percentage-point increase in the cohort's entrepreneurship rate. This means that entrepreneurship not only positively affects income generation on average, but also makes it easier to generate such an increase. Another interesting result is that households headed by women tend to experience more income mobility. This, paired with the fact that their income is significantly lower than those of male-headed households, is a clear indication of the vulnerability of female-headed households.

The classification of entrepreneurship used in this study suffers from data censorship, as only families currently in charge of a business are considered entrepreneurs. Households that owned a business but went bankrupt would fall in the non-entrepreneur category. This leads not only to an overestimation of the positive effect mentioned in future income but also to a possible further increase in the positive effect on income mobility. However, considering that under normal conditions most businesses that go bankrupt follow a process that takes time, the censorship should not be problematic. This is because in any given period, some entrepreneurs are thriving while others are failing. This effectively reduces the average income they perceive over time and considerably diminishes any possible bias.

The results of this study indicate that public policy should put special emphasis on promoting incentives for the development of entrepreneurship as a strategy for economic development.

References

Acemoglu, D., and F. Zilibotti. 1997. "Was Prometheus Unbound by Chance? Risk, Diversification and Growth." *Journal of Political Economy* 105(4): 709–51.

Acs, Z., and and J.E. Amorós. 2008. "Entrepreneurship and Competitiveness Dynamics in Latin America." *Small Business Economics* (31): 305–22.

Alesina, A., and R. Perotti. 1996. "Income Distribution, Political Instability and Investment." *European Economic Review* 105(4): 709–51.

Amorós, J., and O. Cristi. 2010. "Poverty, Human Development and Entrepreneurship." In: M. Minitti, editor. *The Dynamics of Entrepreneurship: Theory and Evidence*. Oxford, United Kingdom: Oxford University Press.

Amorós, J. E., A. Leguina, and I. Gutiérrez. 2010. "Análisis de la Actividad Emprendedora en Sectores de Comercio en América Latina: Una Aproximación desde el Global Entrepreneurship Monitor." Santiago, Chile: Universidad del Desarrollo - FUNDES

Arteaga, M.E. and V. Lasio. 2009. "Empresas Dinámicas en Ecuador: Factores de Exito y Competencias de sus Fundadores." *Academia, Revista Latinoamericana de Administración* 42: 49-67.

Arteaga, M.E. 2011. "Perfil de los Emprendedores de la Región 5." Guayaquil, Ecuador: Escuela Superior Politécnica del Litoral. Manuscript.

Autio, E. 2007. "Global Report on High-Growth Entrepreneurship." Babson Park, United States and London, United Kingdom: Global Entrepreneurship Monitor.

Canelas, C. 2010. "Poverty, Inequality and Income Mobility: The Case of Ecuador. A Pseudo-Panel Approach. Doctoral Dissertation." Paris, France: Paris School of Economics.

Castellani, F., and G. Parent. "Social Mobility in Latin America." Paris, France: OECD Development Centre. Mimeographed document.

Corporación Latinobarómetro 2011. 2010 Report. Accessed at: www.latinobarómetro.org. June 2, 2011.

Cuesta, J., H. Ñopo and G. Pizzolito. 2011. "Using Pseudo-Panels to Estimate Income Mobility in Latin America." Review of Income and Wealth 57(2): 224-246.

Deaton, A. 1985. "Panel Data from a Series of Cross-Sections." *Journal of Econometrics* 30: 109-126.

Doepke, M., and F. Zilibotti. 2005. "Social Class and the Spirit of Capitalism." *Journal of the European Economic Association* 3: 516–24.

Easterly, W. 2001. "The Middle Class Consensus and Economic Development." *Journal of Economic Growth*, 6(4): 317–35.

Franco, R., M. Hopenhayn and A. León. 2011. "Crece y Cambia la Clase Media en América Latina: Una Puesta al Día." *Revista Cepal* 103: 7–26.

Heckman, J., and R. Robb. 1985. "Alternative Models for Evaluating the Impact of Interventions: An Overview." Journal of Econometrics (30): 239-267.

Hisrich, R.D., and G.C. Brush. 1986. *The Woman Entrepreneur: Starting, Financing, and Managing a Successful New Business*. Lexington, Massachusetts, United States: Lexington Books.

Instituto Nacional de Estadísticas y Censos (INEC). 2005. "Las Condiciones de Vida de los Ecuatorianos, Encuesta de Condiciones de Vida." Quinta Ronda. Accessed June 2, 2011 at www.inec.gob.ec.

IPSA Group. 2010. *Ecuador Overview 2010*. Quito, Ecuador: IPSA Group.

Kantis H., P. Angelelli and V. Moori 2005. *Desarrollo Emprendedor: América Latina y la Experiencia Internacional*. Washington, DC, United States: Inter-American Development Bank and FUNDES Internacional.

Kelley, D. J., N. Bosma, and J. E. Amorós. 2011. *Global Entrepreneurship Monitor: Global Report 2010*. St. Louis, United States: Global Entrepreneurship Research Association.

Kharas, H. 2010. "The Emerging Middle Class in Developing Countries." Working Paper 285. Paris, France: OECD Development Centre.

Lazear, E.P. 2005. "Entrepreneurship." *Journal of Labor Economics* 23(4): 649-680.

Lasio, V., M. E. Arteaga, and G. Caicedo. 2009. "Global Entrepreneurship Monitor Ecuador 2008." Guayaquil, Ecuador: Escuela Superior Politécnica del Litoral.

---. 2010. "Global Entrepreneurship Monitor Ecuador 2009." Guayaquil, Ecuador: Escuela Superior Politécnica del Litoral.

Moffitt, R. 1990. "Estimating Dynamic Models with a Time Series of Repeated Cross Sections." Providence, United States: Brown University. Mimeographed document.

Nicolaou, N., S. Shane, L. Cherkas, J. Hunkin, and T. D. Spector. 2008. "Is the Tendency to Engage in Entrepreneruship Genetic?" *Management Science* 54(1): 167-179.

Organisation for Economic Co-operation and Development (OECD). 2010. “Family Affair: Intergenerational Social Mobility across OECD Countries.” In: *Economic Policy Reforms: Going for Growth*. Paris, France: OECD.

----. 2010. *Latin America Economic Outlook*. Paris, France: OECD Development Centre.

Pressman, S. 2007. “The Decline of the Middle Class: An International Perspective.” *Journal of Economic Issues* 40(1): 181–200.

Quadrini, V. 2000. "Entrepreneurship, Saving and Social Mobility." Review of Economic Dynamics 3: 1-40.

Ravallion, M. 2009. “The Developing World’s Bulging (but Vulnerable) ‘Middle Class.’” Policy Research Working Paper 4816. Washington, DC, United States: World Bank.

Revista Lideres. 2011. “Clase Media Creció su Poder de Compra en Cuatro Años.” Accessed: January 17, 2012 at www.lideres.ec.

Shane, S. 2009. “Why Encouraging More People to Become Entrepreneurs is Bad Public Policy.” *Small Business Economics* 33: 141–49.

Solimano, A. 2008. “The Middle Class and The Development Process: International Evidence.” Santiago, Chile: Economic Commission for Latin America and the Caribbean. Mimeographed document.

Torche, F., and L.F. López-Calva. 2010. “Stability and Vulnerability of the Latin American Middle Class.” New York, United States: New York University. Manuscript.

Unger, J. et al. 2011. “Human Capital and Entrepreneurial Success: A Meta-Analytical Review.” *Journal of Business Venturing* 26: 341–58.

Valdez, R. 2011. Clase Media: La Consentida de los Proveedores Inmobiliarios en Ecuador. America Economia On Line. www.americaeconomia.com.

Verbeek, M. 1996. “Pseudo Panel Data.” In: L. Matyas and P. Servestre, editors. *The Econometrics of Panel Data*. Norwell, United States: Kluwer Academic Publishers.

Verbeek, M., and T. Nijman. 1992. “Can Cohort Data be Treated as Genuine Panel Data?” *Empirical Economics* 17: 9-23.

World Economic Forum. 2010. *Global Competitiveness Report*. Davos, Switzerland: World Economic Forum.

Appendix Table 1. Distribution of the Households Sampled by Gender and Date of Birth Cohorts

Cohorts		Years Analyzed								Total
		2003	2004	2005	2006	2007	2008	2009	2010	
Men	1	1521	1487	1401	1328	1156	1176	1145	1121	10335
	2	1399	1601	1325	1392	1205	1259	1315	1405	10901
	3	1535	1523	1438	1420	1524	1682	1645	1755	12522
	4	1549	1510	1518	1501	1357	1308	1388	1426	11557
	5	1150	1353	1251	1242	1198	1183	1247	1322	9946
	6	1491	1412	1337	1445	1200	1385	1382	1406	11058
	7	1506	1577	1548	1339	1367	1402	1328	1516	11583
	8	1655	1731	1595	1803	1728	1723	1725	1774	13734
	9	1671	1639	1497	1515	1794	1789	1552	1581	13038
	10	961	1038	1220	1347	1862	1758	1787	1933	11906
Women	11	680	657	642	620	526	612	564	589	4890
	12	505	580	530	570	500	536	506	602	4329
	13	486	495	429	482	522	582	625	727	4348
	14	434	435	381	415	417	408	459	531	3480
	15	305	360	310	353	342	342	401	453	2866
	16	356	346	314	317	349	427	417	442	2968
	17	337	356	307	319	361	386	350	435	2851
	18	312	329	309	391	366	456	439	467	3069
	19	211	252	258	267	303	350	370	432	2443
	20	195	209	225	279	332	368	406	473	2487
Total		18259	18890	17835	18345	18409	19132	19051	20390	150311

Source: Authors' calculations based on ENEMDU surveys.

Appendix Table 2. Distribution of the Households Weighted by Gender and Date of Birth Cohorts

	Cohorts	2003	2004	2005	2006	2007	2008	2009	2010	Total
		Years Analyzed								
Men	1	222185	213682	217222	204647	172527	181443	181781	171545	1565031
	2	195668	238828	205626	219123	198276	201554	226031	226984	1712091
	3	244078	241450	237137	227344	250774	276392	277756	284957	2039888
	4	239114	242831	258934	260577	235487	234598	243217	246474	1961232
	5	179716	220244	219791	225065	216272	212070	217332	218750	1709240
	6	236965	224501	234392	253949	219800	247841	238696	241102	1897246
	7	240246	255779	268285	247245	248280	257454	238871	260530	2016690
	8	264702	276059	270947	303896	317966	304104	307826	310869	2356368
	9	267446	277079	279140	263016	328642	327426	284393	277867	2305008
	10	154527	166747	216726	240924	347137	328959	341834	352987	2149840
Women	11	99501	96330	99457	102256	88582	100515	97725	98631	782996
	12	76802	86569	87160	92989	84152	89649	88349	105824	711495
	13	80186	81541	74399	84685	95134	105482	114474	129457	765357
	14	72619	74104	72840	81452	82510	78154	81774	92864	636317
	15	55332	63672	59513	68102	63566	61199	71520	82187	525090
	16	58501	59247	54766	57536	70491	89688	78716	84324	553269
	17	53132	59559	59351	61261	67628	74120	69591	88904	533544
	18	52854	50106	60060	80833	68931	85654	85545	92480	576463
	19	35325	37220	48858	45945	58482	62280	74424	82875	445407
	20	30134	33381	42505	47290	63648	70016	82707	93536	463217
Total		2859031	2998930	3067109	3168135	3278284	3388597	3402560	3543144	25705790

Source: Authors’ calculations based on ENEMDU surveys.