Intergenerational Mobility in the Presence of Informal Labor Markets

Paula Cesana
Queen Mary University of London

November 24, 2025

Click here for the most recent version

This paper studies intergenerational mobility in the context of Chile, an economy with a significant informal labor market. Using longitudinal data, I document two key empirical facts. First, labor informality is associated with higher income uncertainty. Second, the share of time individuals spend in informal employment is correlated with parental background, suggesting a positive association in labor informality across generations. Moreover, I show most of this association can be explained by substantial intergenerational persistence in education and occupations. I then propose and estimate a model of human capital investment and occupation choice under uncertainty to quantitatively assess the role of the income uncertainty channel in intergenerational mobility, namely, persistence in education and occupations, and in labor informality. In the model, higher income uncertainty reduces parental investment in children's human capital, limiting access to higher-skilled occupations, which are also less affected by informality. I show that reducing parental income risk in informal employment increases the share of educated individuals and enhances upward mobility. These findings underscore how labor informality can shape future labor market outcomes and perpetuate barriers to intergenerational mobility.

1. Introduction

Labor informality, broadly defined as employment outside any regulatory framework, is a salient feature of many labor markets, specially in low and middle-income countries. It affects around 51.3% of the workforce in Latin America, 85.1% in Southern Asia and 85.2% in Africa (ILO, 2024). In particular, in Latin America, labor informality coexists with high income inequality and low intergenerational mobility. While previous research has emphasized the relationship between informality and income inequality, one critical question remains underexplored: what is the link between informality and intergenerational mobility? How does labor informality shape the long-term outcomes of future generations?

This paper takes a step toward addressing that question. The key channel I study is whether greater uncertainty associated with informal employment undermines investment in the next generation's human capital, thereby limiting upward mobility. If this mechanism holds, it could create a cycle in which parents exposed to larger uncertainty are unable to adequately invest in the children's human capital, increasing the likelihood that their children engage in informal work. In turn, children engaged in informal work would face greater uncertainty themselves, perpetuating the cycle. Thus, the so-called "informality trap" could extend beyond a single generation, persisting across generations.

I study this question in the context of Chile, using longitudinal survey data from the *Encuesta de Protección Social* (EPS) from 2002 to 2020[1]. This is a relevant setting for two reasons. First, as other countries in the region, Chile has a both a sizable prevalence of informal employment and low intergenerational mobility. Recent estimates show that the informal employment share amounts to 27.5% of the workforce (ILO, 2024), while usual indicators of intergenerational mobility rank among the lowest in the OECD (Munoz and Van der Weide 2025)[2]. Second, Chile is one of the few countries with these characteristics for which extensive longitudinal survey data is available. The survey includes retrospective information on parents' education and occupation, as well as individuals' labor histories dating back to 1980, allowing me to characterize individual labor histories and relate them to family background.

In the first part of the paper, I document key empirical facts about labor informality and intergenerational mobility in Chile. First, I show that labor informality is indeed associated with higher income uncertainty. I define an informal employment spell based on the conditions under which a worker is hired or performs their job, regardless of occupation (i.e., the type of job or tasks performed). Exploiting the longitudinal data, I run fixed

[1] I am currently updating the results to include the recent release of the 2023 wave

[2] For the 1980s cohort, the Global Database on Intergenerational Mobility reports that Chile has an intergenerational income elasticity of 0.58 and a correlation between parents' and children's years of education of 0.55, compared with OECD averages of 0.38 and 0.41, respectively. Higher values indicate lower intergenerational mobility, meaning that children's outcomes are more strongly tied to their parents'.

effects regressions and show that informal employment spells exhibit larger dispersion in residual earnings, which I interpret as a measure of income uncertainty. This pattern also holds at the occupation level, suggesting that, even within narrow employment categories, informal employment spells are subject to greater uncertainty.

Second, I show that an individual's "lifetime informality" is associated with their parental background. I define lifetime informality as the share of time individuals spend in informal employment, while parental background is measured based on educational attainment and main occupation. I find that a higher likelihood of informality in parents is associated with a greater share of time spent in informality by their children. Specifically, a 10 percentage point increase in a parent's likelihood of informality is associated with a 2.58 percentage point increase in lifetime informality for males and a 2.42 percentage point increase for females. I interpret this as suggestive evidence of intergenerational persistence in labor informality.

To explore the channels driving this association, I further document that there is substantial intergenerational persistence in education and occupations, two features that have been well-documented in other economies. First, a son (daughter) is more likely to be high-educated if their father (mother) is high-educated. On average, around 70% of individuals attain the same education level as their parent[3]. Second, individuals are more likely to select into the same occupations as their parents. At the 1-digit level, sons are 84% more likely to work in a given occupation if their father does as well, while this number is 72% for daughters and mothers. The overall share of individuals who coincide with their parent's occupation is 22.8% for males and 21.9% for females.

Accounting for "followers" –individuals who attain the same education level and/or work in the same occupation as their parents– explains approximately 70% of the initial association between informality and parental background for males and around 90% for females. Overall, this evidence underscores that intergenerational persistence in education and occupations are key mechanisms through which labor informality can persist across generations.

Motivated by the empirical facts, in the second part of the paper I propose a model of human capital investment and occupation choice under uncertainty. The goal of the model is to quantitatively assess the role of income uncertainty associated with informality in shaping intergenerational mobility, namely, education and occupation persistence, and on informality.

The model features two generations who act over two periods. In the first period, parents engage in the labor market and decide whether to invest in their children's human capital or not. In the second period, children choose occupations and are hired either formally or informally with some probability that varies with education and occu-

[3]In order to account for secular trends, I let the definition of 'high-skilled' to evolve over time: at least high school for the parent's generation, and at least some college for the child's generation.

pation. There is uncertainty in labor income for both generations, which is captured by a shock to labor income whose variance depends on whether employment is formal or informal. Moreover, children have occupation-specific preferences that are correlated with their fathers' occupations. This feature captures the idea that children derive utility from following similar occupational paths and thus provides a channel that generates intergenerational occupational persistence.

The key mechanism of the model is that higher income uncertainty leads parents to reduce investment in their children's human capital. In turn, children with lower levels of human capital are more likely to work informally and face greater income uncertainty themselves.

I estimate the model using the EPS Chilean data for pairs of sons and fathers from 2002 to 2020. I externally calibrate the parameters governing the income processes and informality probabilities based on the "lifetime informality" measure discussed above. The remaining parameters are estimated through a simulated method of moments, by matching simulated moments from the model to their empirical counterparts. Preferences for occupations are informed from occupation shares in the data, conditional on the father's occupation, while the parameters governing the father's investment choice are identified using data for the shares of high-educated sons conditional on the father's background. The model reproduces the main patterns of education and occupation persistence, informality and overall shares of high-educated sons and employment by occupation.

Finally, I perform a series of counterfactual exercises to evaluate the role of informality-driven uncertainty in shaping intergenerational outcomes. I find that reducing uncertainty in informal employment in the parent's generation to match that of their formal counterparts increases the share of high-educated sons by 3.6 percentage points (pp), from a baseline share of 50.8%. Driven by this, overall intergenerational persistence in education and in occupations decrease by 2.2 pp and 0.6 pp respectively, from their baseline levels of 63.4% and 21.7%. These results appear modest at the aggregate level, but are more pronounced when considering patterns of upward mobility. Among children of low-educated parents, educational persistence declines by 5 pp, while among children of parents in lower-skilled occupations, occupational persistence decreases by 1.7 pp. These results also imply a weaker intergenerational association between the father's and son's informality.

Overall, this paper proposes and quantifies a mechanism through which informality contributes to low intergenerational mobility, emphasizing how uncertainty in labor income can constrain both educational attainment and occupational outcomes across generations. I interpret the results as lower-bound effects, as they capture only uncertainty related to labor income, while other aspects of informality, such as differential unemployment risk or reduced access to non-labor income such as social security benefits, could likely amplify these effects.

Related literature

This paper relates to multiple strands of literature on intergenerational mobility, human capital investment, and labor informality, particularly in the context of developing countries. Next, I review the main empirical patterns on intergenerational mobility and labor informality, and the channels or determinants in which the literature has focused.

Intergenerational Mobility: empirical patterns. Intergenerational mobility in Latin America is notably low compared to other regions, reflecting a broader pattern in which educational and income mobility are lower in developing countries compared to high-income ones (Van der Weide et al. 2024). Research on mobility in the Latin American region has predominantly focused on educational attainment, largely due to data limitations, and has often examined its relationship with inequality. In this context, while educational mobility remains limited, it has shown improvement in recent decades. Neidhöfer, Serrano, and Gasparini (2018) highlights rising trends in relative and absolute educational mobility across 18 countries in the region, driven by increased upward mobility among children from less-educated families. However, broader indicators of well-being –such as job stability and homeownership– reveal minimal progress over time, suggesting limited opportunities for substantial socioeconomic mobility (Neidhöfer, Ciaschi, and Gasparini 2022). Relatedly, Brunori, Ferreira, and Neidhöfer (2024) emphasize the role of family background, such as education and occupation, in reproducing inequality. Finally, studies on income mobility, while scarcer, consistently underscore high persistence. For example, estimates for Chile (Nunez and Miranda 2010) and Brazil (Dunn 2007) reveal strong intergenerational income persistence. Recent research has addressed some data challenges by leveraging administrative records (Britto et al. 2022; Leites et al. 2022). In particular, Britto et al. (2022) provide comprehensive estimates for Brazil, incorporating both formal and imputed informal income data, and find a strong degree of income persistence, highlighting the role of location and assortative mating, among other factors.

I contribute to this literature by documenting key empirical facts associated with intergenerational persistence using longitudinal data for Chile. Moreover, I focus not only on education and occupation outcomes, but explore other important labor market outcomes, specifically informality.

Explaining intergenerational mobility. A significant body of literature has studied the role of human capital investment in intergenerational mobility, drawing on the contributions of Becker and Tomes (1979, 1986). Particularly related to this paper is research examining these dynamics in the context of uninsured income risk (Abbott 2022; Agostinelli et al. 2024). These studies show that parental income risk negatively affects child-related expenditures and the development of children's skills or abilities. Similarly, related literature highlights the timing of parental investments and the importance of balanced income profiles, particularly, during critical stages of the life cycle (Lee and Seshadri 2019; Carneiro et al. 2021). Finally, closely related is the study by Bobba, Flabbi, and Levy (2022), which examines individuals' decisions on their own human capital within a search-and-matching framework with formal and informal jobs, although the mechanism is not related to risk but to returns to human capital. Similarly, López García (2015) studies a similar question using a life-cycle model, in which individuals choose schooling and labor market participation in the case of Chile.

I build on this literature by examining the role of parental risk in shaping both children's human capital investments and their future labor market outcomes, in particular, explicitly addressing intergenerational mobility in occupations.

Intergenerational persistence in occupations has also been documented in developed countries (Long and Ferrie 2013; Constant and Zimmermann 2003). The literature highlights various channels driving this persistence. One line of research emphasizes the transfer of occupation-specific human capital (Laband and Lentz 1983; Hellerstein and Morrill 2011) or, more broadly, the intergenerational transmission of skills and abilities (Lee and Seshadri 2019; Abbott et al. 2019). Another strand highlights the role of preferences in shaping occupational choices, and their intergenerational transmission (Doepke and Zilibotti 2008; Escriche 2007). A further important channel is the influence of intergenerational networks on occupational outcomes (Kramarz and Skans 2014; Corak and Piraino 2011; Magruder 2010). In particular, Lo Bello and Morchio (2022) develop a model of occupational choice with search frictions, examining multiple channels and finding that networks account for the largest share of occupational persistence in the UK.

In this paper, I explicitly incorporates the preferences channel to account for occupational persistence in the theoretical framework and, different from previous studies, I consider intergenerational persistence in occupations in the context of uncertainty.

Labor informality. The extent and prevalence of labor informality in Latin America have been extensively documented (for example, Ulyssea (2020); Gasparini and Tornarolli (2009)). Informality is more prevalent among women, low-skilled workers, and individuals at both ends of the age spectrum. A large portion of informality consists of low-skilled self-employed workers, who are often not registered and do not contribute to social security. There is also a substantial income gap between formal and informal workers, which persists even after controlling for observable characteristics. Furthermore, workers who shift from formal to informal employment typically experience wage losses, which provides further evidence on the informality income penalty (Maurizio and Monsalvo

2021). Recent studies have also highlighted that informal workers face greater income volatility compared to their formal counterparts (Engbom et al. 2022; Gomes, Iachan, and Santos 2020).

This paper focuses on one specific aspect of informality: higher income uncertainty. I document similar empirical patterns as those in recent contributions, but further show that informality is the most important determinant of income uncertainty, even when considering narrow employment groups.

An important body of literature examines why informality persists and its implications for the wage distribution, productivity, and welfare. Much of this work uses search models to explore firms' decisions to hire formally or informally (Bosch and Esteban-Pretel 2012; Meghir, Narita, and Robin 2015). Meghir, Narita, and Robin (2015) find that tighter enforcement raises formal sector wages and welfare. Using similar models, related research further examines worker self-selection into informal jobs Albrecht, Navarro, and Vroman (2009); Haanwinckel and Soares (2021); Narita (2020). In particular, Haanwinckel and Soares (2021) highlight the role of workforce composition as a driver of reductions in informality, while Narita (2020) also considers selection into self-employment. Another important contribution is that of Ulyssea (2018), who develops am equilibrium framework to explain two margins of informality –business registration and off-the-books hiring– and concludes that reducing informality does not always improve welfare or productivity. More recently, Erosa, Fuster, and Martinez (2023) build on this model to include selection into enterpreneurship.

So far, in this paper, I treat informality probabilities as exogenous to workers, which is more aligned with the contributions that highlight the role of the firm in the decision to hire formally or informally. In future work, I plan to expand on this aspect, by incorporating labor demand and/or accounting for self-selection of workers into formal and informal employment.

Outline. The rest of the paper is organized as follows. In Section 2, I document the empirical facts; in Section 3, I present the model; in Section 4, I discuss the model quantification; in Section 5, I perform the counterfactual exercise, to finally conclude in Section 6.

2. Empirical facts

In this section I document two key novel empirical facts that will also motivate the model. First, I show informal employment is associated with greater uncertainty; second, that lifetime informality is associated with parental background.

2.1. Data and definitions

The main data source for the empirical analysis is the *Encuesta de Protección Social* (EPS), a longitudinal survey conducted in Chile across seven rounds spanning from 2002 to 2020. Two key features of the EPS are, first, that it provides retrospective information on family background, including the education and main occupation of their parents. Second, that given its longitudinal nature, I can track individuals' labor spells and their characteristics over time. The survey also asks for retrospective information about labor histories dating back to 1980. I use this information to construct the lifetime indicators as described below.

For the empirical analysis, I restrict the sample to individuals aged 25 to 60. The baseline sample consists of 22,737 individuals and a total of 74,614 spells. On average, individuals feature in the survey 3.9 times, and the average tracking time is 17.5 years. Following previous studies (Bosch and Esteban-Pretel 2012), I define an employment spell as formal if the individual is a wage worker with a signed contract, or a self-employed worker in professional or technical occupations; otherwise, I classify the spell as informal. Employers are excluded from the analysis. Using this definition, the average informality rate during the survey period is around 29%.

In Appendix A, I provide more details on the EPS data, along with descriptive statistics for both the panel and cross-section data.

2.2. Informality and income uncertainty

In this section, I document that informal employment spells are associated with larger uncertainty in labor income. This is evidenced by the fact that informal employment spells exhibit larger dispersion in residual earnings. Moreover, I document this pattern holds even when accounting for occupation and education, which suggests that labor informality is a key margin associated to income uncertainty, even within narrow employment groups.

First, in order to get a measure for labor income uncertainty, I exploit the longitudinal data and run a fixed effects regression for individual i and spell t:

$$\ln y_{it} = \alpha_i + \beta x_{it} + \lambda_t + \epsilon_{it}$$

where y_{it} is a measure of income, x_{it} is a set of controls, α_i and λ_t are individual and time fixed effects, and the residuals ϵ_{it} are idiosyncratic shocks. The baseline measure for labor earnings is real monthly income in the main occupation, as reported at the time of the survey interview, excluding individuals with zero earnings.[4] In the baseline specification,

[4]This measure of income excludes contributions to social security, health, bonuses, etc. Real income is obtained by deflating labor income using monthly CPI data from the Chilean Central Bank. Since the interview dates are not specified, I use the average inflation during the data collection period for each wave (see Appendix A). Observations with zero earnings are excluded to ensure consistency, as they are coded as missing values in the 2015 wave.

the set of controls includes age and age squared.

Next, I compute the distribution of residuals ϵ_{it} separately for formal and informal employment spells, and interpret the moments of these distributions as the properties of earnings uncertainty faced by workers in spell t.

Figure 1 shows the distributions of residual earnings for formal and informal workers, pooled over the full period from 2002 to 2020. I compute and interpret the first four moments of these distributions following the earnings dynamics literature (Guvenen et al. 2021). These are shown in Table 1. The broad patterns and interpretation are similar to what is reported by Engbom et al. (2022) for the case of Brazil.

As shown in Table 1, the mean of the residual log earnings distribution is negative for informal employment spells and slightly positive for formal employment spells. Moreover, the standard deviation of residual log earnings is higher for informal employment spells (0.438) compared to formal employment spells (0.321)[5]. The same pattern is observed for the difference between the 90th and 10th percentiles of the distributions ($P_{90} - P_{10}$), which is robust to outliers. I interpret this greater dispersion of residuals in informal employment as indicative of higher income uncertainty.

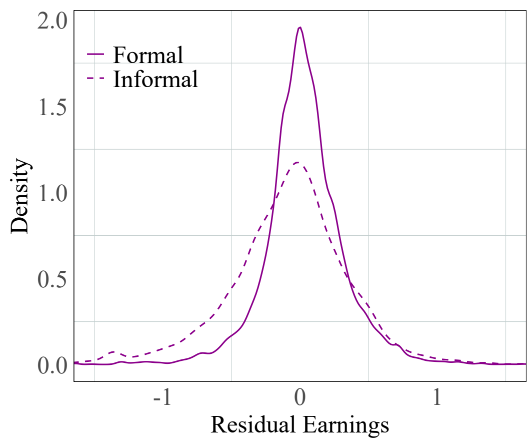
For higher order moments, I report the Kelley skewness and the excess Crow-Siddiqui kurtosis, which are quantile-based measures and therefore robust to outliers[6]. The distribution of residuals for informal workers is left-skewed, while it is right-skewed for formal employment spells. This indicates that informal spells are associated with a higher probability of negative shocks. Lastly, while the distributions of residuals are leptokurtic for both formal and informal spells, formal spells exhibit higher kurtosis, reflecting more extreme deviations from the mean.

These patterns are unchanged when considering hourly earnings, and when including further controls such as the spell's occupation and the individual's education level. I show these in Appendix B. This means that even within occupation and education groups, informality is a key determinant of income risk.

[5]This difference is significant at the 5% level

[6]The Kelley skewness is defined as $S_k = \frac{(P_{90}-P_{50})-(P_{50}-P_{10})}{P_{90}-P_{10}}$, and the Crow-Siddiqui kurtosis as $\kappa_{cs} = \frac{P_{97.5}-P_{2.5}}{P_{75}-P_{25}}$. The excess Crow-Siddiqui kurtosis is computed as $\kappa_{cs} - 2.91$, which corresponds to its value in a Normal distribution.

FIGURE 1. Density of residual log earnings, 2002-2020



Note: The figure shows the Kernel density of the distributions of residuals of a fixed effects regression for all EPS survey waves, separately for formal (solid line) and informal (dashed line) employment spells.

TABLE 1. Moments of residual log earnings, 2002-2020

	Formal	Informal
Mean	0.028	-0.073
Standard deviation	0.321	0.438
p90-p10	0.680	1.022
Kelley skewness	0.059	-0.070
Excess C-S kurtosis	1.265	0.744
Observations	28,244	11,104

Note: The table shows moments of the distributions of residuals of a fixed effects regression for all EPS survey waves, separately for formal and informal employment spells.

Further evidence on this is presented in Table 2, in which I show the results of a regression analysis that relates the uncertainty measure to demographic and occupation characteristics. As shown in the table, informality is always a key margin associated to dispersion in residual earnings. When including education, occupation and their interaction as further controls, the coefficient for informality remains virtually unchanged.

TABLE 2. Regression results: income uncertainty and labor spell characteristics (dependent variable: standard deviation of log residual earnings)

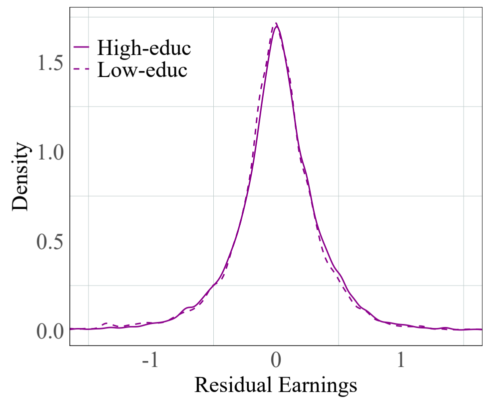
	(1)	(2)	(3)	(4)	(5)
Intercept	0.285***	0.278***	0.292***	0.286***	0.247***
	[0.015]	[0.018]	[0.031]	[0.032]	[0.040]
Informal (dummy)	0.146***	0.146***	0.145***	0.145***	0.143***
	[0.018]	[0.018]	[0.019]	[0.019]	[0.018]
Education		✓		✓	✓
Occupation			✓	✓	✓
Education*Occupation					✓
Observations	71	71	71	71	71
R^2	0.500	0.504	0.522	0.526	0.625

Finally, when considering income uncertainty across education groups only, the data shows almost no variation at all. This is shown in Figure 2, which shows the estimated density and moments of the distribution of residual earnings for low-educated and high-educated individuals separately. In turn, there is variation across occupations, which is shown in Figure 3, displaying a positive correlation between the uncertainty measure and the informality share by occupation[7].

[7] Table A6 in the Appendix shows the measure of labor income uncertainty across other relevant groups

FIGURE 2. Residuals of log earnings, by education level

A. Residuals of residual log earnings

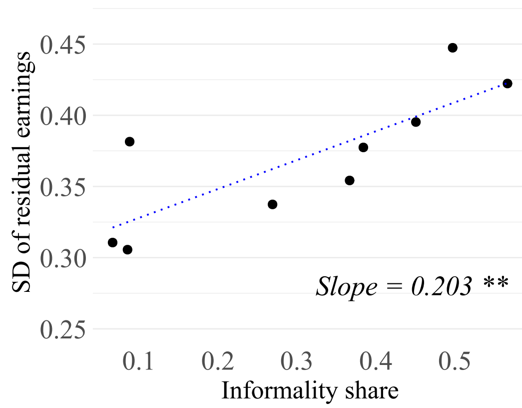


B. Moments of residual log earnings

	Low-Educ	High-Educ
Mean	0.007	-0.002
Standard deviation	0.367	0.368
$P_{90} - P_{10}$	0.808	0.797
Kelley skewness	-0.002	-0.017
Exc. C-S kurtosis	1.326	1.419
Observations	15,216	26,645

Note: Residuals of log earnings are calculated from a fixed effect regression with age controls. Low-educated individuals are those with less than high-school, while high-educated individuals are those with completed high school or more. This pattern is unchanged when defining the education groups based on college attendance (see Appendix B)

FIGURE 3. Standard deviation of residual log earnings, by occupation groups and informality shares



Note: Each dot represents an occupation, which are labeled following the 1-digit level ISCO-08: (1) Managers; (2) Professionals; (3) Technicians and Associated Professionals; (4) Clerical Support Workers; (5) Service and Sales Workers; (6) Agricultural, Forestry and Fishery Workers; (7) Craft and Related Trades Workers; (8) Operators and Assemblers; (9) Elementary Occupations.

2.3. Lifetime informality and parental background

In this section, I first document the association between lifetime informality and parental background. To explore the drivers of this association, I then examine two additional channels of intergenerational persistence, in education and occupations. Finally, I show that accounting for education and occupation 'followers' explains most of the initial association, which suggests it operates through education and occupation outcomes.

First, to measure lifetime informality, I exploit the longitudinal data to compute the proportion of time each individual spends in informal employment. This is defined as the number of months an individual reports working in an informal job divided by the total number of months observed. [8]. Interestingly, this measure reveals high persisistence of formality status at the individual level, with approximately 71% of males and 72.5% of females spending all their observed time in either formal or informal employment.

Second, I define parental background as a combination of the parent's education level and main occupation, which are the two key variables available in the data. Based on the characteristics of the parent's sample, I categorize education into two levels: less than high school, and high school or more. Occupations are classified into 9 categories according to the 1-digit level of the International Standard Classification of Occupations (ISCO). This results in a total of 18 distinct parental backgrounds. More details on the parent's sample are provided in Appendix A.3.

A limitation of the data is the lack of information about informality for the parents' generation. However, since informality is strongly correlated with gender, education, and occupation, I assign an "informality likelihood" for each parental background based on the informality measure discussed above. There is great heterogeneity in this measure, ranging from over 75% for low-educated agricultural workers and managers (self-employed), to less than 10% for high-educated professionals and technicians. The complete table for lifetime informality by occupation and education is shown in Appendix D.

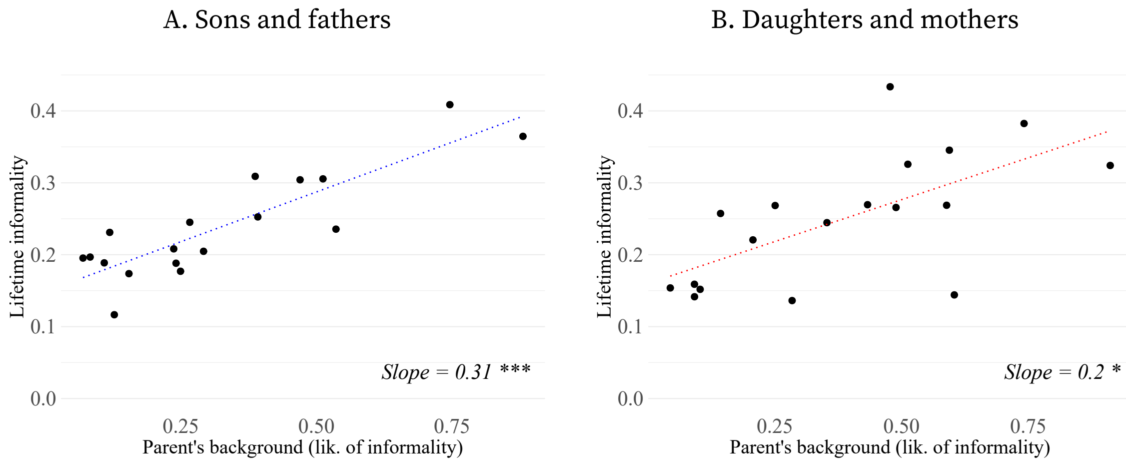
Figure 4 illustrates the association between the share of time individuals spend in informality and their parental background, separately for males (relative to fathers) and females (relative to mothers). The figure reveals a positive association between these variables: individuals with parents from backgrounds associated with higher levels of informality tend to spend a larger share of time in informal jobs[9].

In the next paragraphs, in order to explore the determinants of this association, I document there is substantial intergenerational persistence in education and occupations.

[8] I include only individuals with at least one year of observations in order to ensure certain degree of attachment to the labor market.

[9] In Appendix D, I show that this relation is robust to using individuals born before 1970 to compute the parents' "likelihood of informality"

FIGURE 4. Share of time in informality by family background



Note: Lifetime informality is the share of time in informal employment (vertical axis). A parent's background is defined as the combination of education and occupation. Each parental background is assigned an informality likelihood (horizontal axis).

In this section, I examine the association between individuals' educational attainment and their parents' educational attainment. For the individuals' education level, I use the most frequently reported level of education, while the parent's education level is defined as reported by their child.

Table 3 presents educational attainment levels for males and females, both overall and conditional on their parent's education. For males, the calculations are based on their fathers' education, while for females, they are based on their mothers'. The table highlights significant persistence in educational attainment: while 44.5% of males, on average, achieve more than a high school education, this share increases to 66.4% for those with high school-educated fathers but drops to 30.2% for those with fathers who have less than a high school education. The same pattern is observed for females: on average, 39.8% attain more than a high school education, with this share rising to 71.1% for those whose mothers have higher education and falling to 29.4% for those whose mothers have less than a high school education.

This intergenerational association in educational attainment persists when accounting for standard covariates, as shown in Table 4. For males, having a father with at least a high school education increases the likelihood of attending college by 17%, while for females, it increases the likelihood by 10%. Similarly, having a mother with at least a high school education increases the likelihood of college attendance by approximately 17% for males and 15.7% for females.

TABLE 3. Educational attainment by parent's education, 2002-2020

	Males			Females		
Parent's education	Less than HS	HS	College	Less than HS	HS	College
Less than HS	0.358	0.339	0.302	0.364	0.342	0.294
HS or more	0.082	0.254	0.664	0.062	0.226	0.711
Average	0.293	0.263	0.445	0.299	0.303	0.398

Note: The table shows the individual's educational attainment for males and females (less than high school, high school, some college), both on average and conditional on their parent's educational attainment (less than high school, high school or more). Shares are computed conditional on the father's education for males, and conditional on the mother's education for females.

TABLE 4. Conditional probability of college education, 2002-2020

	Males			Females		
	(1)	(2)	(3)	(1)	(2)	(3)
Mean probability	0.476	0.478	0.486	0.414	0.414	0.424
Intercept	0.301***	0.427***	0.383***	0.294***	0.391***	0.291***
	[0.006]	[0.027]	[0.029]	[0.005]	[0.013]	[0.025]
Father's education	0.453***	0.278***	0.170***			0.100***
(high school or more)	[0.010]	[0.011]	[0.013]			[0.013]
Mother's education			0.170***	0.417***	0.252***	0.157***
(high school or more)			[0.013]	[0.010]	[0.011]	[0.013]
Controls		✓	✓		✓	✓
Observations	8,472	8,379	7,814	10,190	10,046	8,356
R2	0.195	0.294	0.324	0.147	0.243	0.266

Note: Columns 1 show the unconditional association between own and parental education. Controls in columns 2 include cohort, siblings and parent's occupation. Controls in columns 3 also add the other parent's education and occupation.

Occupation persistence

Next, I present evidence on intergenerational persistence in occupations, that is, the probability that an individual and their parent work in the same occupation. For the parent's occupation, I use the main occupation reported by the surveyed individual. For the individual's occupation, I define the main occupation as the one they hold most frequently, considering all labor spells and their durations. This measure for main occupation demonstrates strong persistence at the individual level: 51.5% of individuals spend all their time in the same occupation, and over 94% of individuals spend more than half of the time in the same occupation[10].

[10] I include only individuals with at least one year of observations. Repeating the analysis using all occupation spells yields similar results.

In order to show intergenerational persistence in occupations, I first construct likelihood ratios, defined as the ratio between the probability of working in a particular occupation conditional on the parent working in that occupation, and the unconditional probability of working in that occupation (Lo Bello and Morchio 2022). Table 5 illustrates the likelihood ratios for males (relative to their fathers) and females (relative to their mothers). All likelihood ratios exceed 1, indicating that the conditional probability of working in an occupation is consistently higher than the unconditional probability: an individual is more likely to work in a given occupation if their parent worked in it. On average, the likelihood ratio is 1.84 for males and 1.72 for females, meaning they are 84% and 72% more likely, respectively, to work in a given occupation if their parent worked in it.

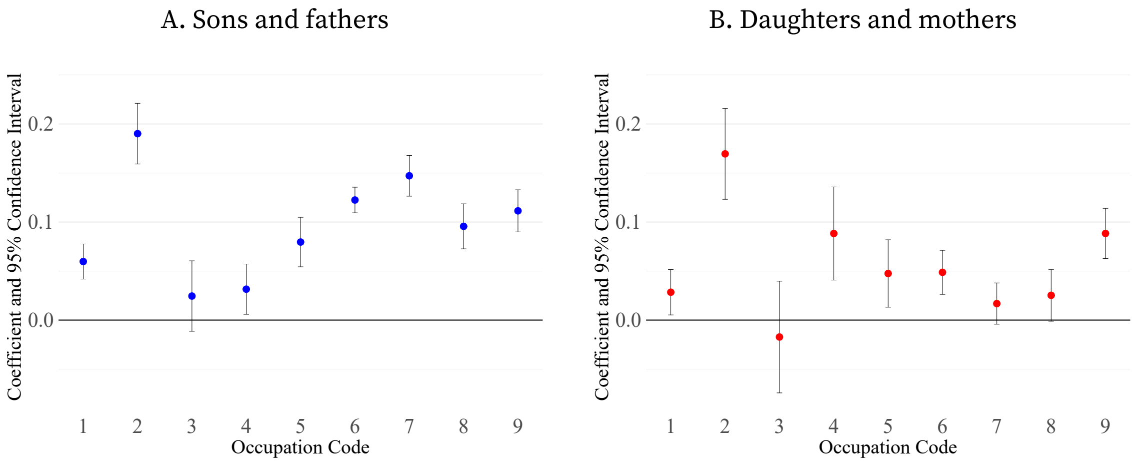
This pattern persists even after controlling for usual covariates. Figure 5 plots the coefficients of occupation persistence by occupation, conditional on education and cohort. For males (Panel A), the coefficients are significant for all but one occupation: having a father in a given occupation increases the probability of working in that occupation by between 2.9% and 18.1%. For females (Panel B), the persistence is lower on average, with significant effects ranging from 2.5% to 17%. In Appendix C, I report the full regression results and repeat the estimations conditional on additional covariates. I also show that occupation persistence is higher for the lower-educated individuals.

TABLE 5. Occupation persistence (likelihood ratios)

Occ. Code	Occupation Group	Males	Females
1	Managers	2.30	1.43
2	Professionals	3.57	2.56
3	Technicians and Assoc. Prof.	1.15	1.48
4	Clerical Support Workers	0.93	1.93
5	Service and Sales Workers	1.34	1.26
6	Agric., Forestry and Fishery Workers	4.16	5.86
7	Craft and Related Trades Workers	1.60	1.87
8	Operators and Assemblers	1.44	1.33
9	Elementary Occupations	1.86	1.53
	Average	1.84	1.72

Note: Likelihood ratios are calculated as the ratio between conditional and unconditional occupation shares. For males, the ratios are computed conditional on their father's occupation; for females, they are computed conditional on their mother's occupation.

FIGURE 5. Conditional occupation intergenerational persistence



Note: The figure plots the persistence coefficients by occupation (1-digit level), after controlling for cohort and education. Occupational codes are as follows: (1) Managers; (2) Professionals; (3) Technicians and Associated Professionals; (4) Clerical Support Workers; (5) Service and Sales Workers; (6) Agricultural, Forestry and Fishery Workers; (7) Craft and Related Trades Workers; (8) Operators and Assemblers; (9) Elementary Occupations.

Reassessing the intergenerational association in informality

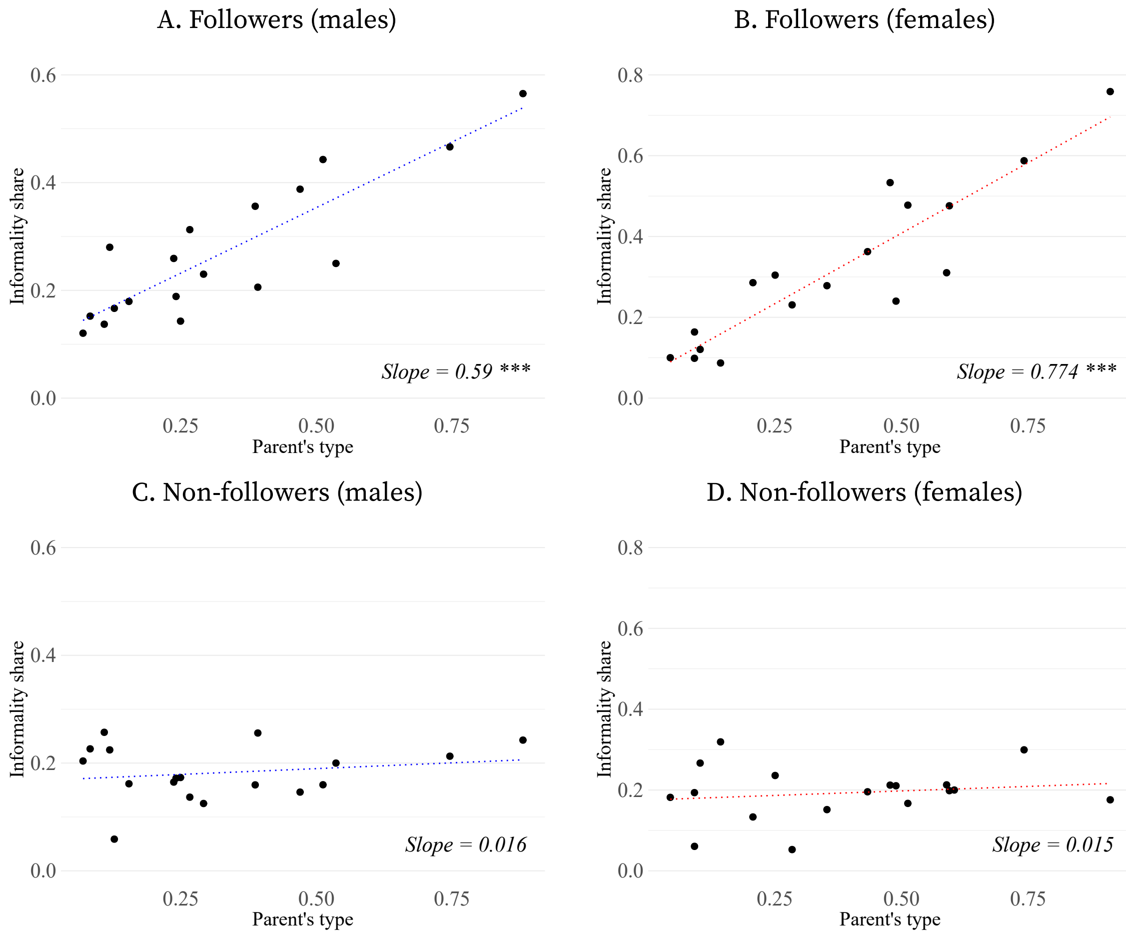
Following the previous analysis of intergenerational persistence in education and occupations, this section explores how the relationship between labor informality and parental background differs for individuals who follow their fathers' educational and occupational choices. Specifically, I investigate whether part of the observed association in labor informality can be attributed to children aligning with their parents in terms of education and occupation.

In order to show this, consider first Figure 6, in which I plot the association between lifetime informality and parental background, separately for followers and non-followers. A 'follower' is an individual who coincides with their parent in either education or occupation. The figure reveals a striking pattern: for followers, the association is very strong, while for non-followers the association becomes almost flat. This pattern holds both for males and females.

Consistent with these aggregate patterns, Table 6 provides evidence from regressions of lifetime informality on parental background at the individual level. The table includes both the unconditional association and models that incorporate interaction terms to differentiate between followers and non-followers. The full regression results are reported in Appendix D.

As shown in Table 6, for males, the unconditional association (column 1) shows that a 10 percentage point (pp) increase in the likelihood of a father being informal is associated

FIGURE 6. Share of time in informality by family background



Note: Lifetime informality is the share of time in informal employment (vertical axis). A parent's background is defined as the combination of education and occupation. Each parental background is assigned an informality likelihood (horizontal axis). A follower is defined as an individual who coincides with their parent in occupation and/or education level. For males, association is computed based on the father's background; for females, it is computed based on the mother's background.

with a 2.58 pp increase in a son's lifetime informality. Accounting for education followers reduces this association by nearly half (column 2), while accounting for occupation followers explains around half of the association (column 3). Finally, when including interaction terms for both education and occupation followers, the original association is reduced by 70%. Interestingly, even after accounting for these follower interactions, the remaining association between parental background and informality is still positive and significant: a 10 pp increase in the likelihood of a father being informal is associated with a 0.8 pp increase in lifetime informality. This remaining association suggests that additional, unexplored channels contribute to the persistence of informality, which will

TABLE 6. Regression results: lifetime informality and parental background (dependent variable: share of time in informality)

a. Males

	(1)	(2)	(3)	(4)
Intercept	0.117***	0.114***	0.137***	0.118***
	[0.013]	[0.016]	[0.014]	[0.017]
Parental background	0.258***	0.138***	0.145***	0.079**
	[0.023]	[0.032]	[0.025]	[0.034]
Education followers		✓		✓
Occupation followers			✓	✓
Cohort controls	✓	✓	✓	✓
Observations	6,659	6,652	6,591	6,584
R^2	0.061	0.099	0.100	0.122

b. Females

	(1)	(2)	(3)	(4)
Intercept	0.133***	0.152***	0.157***	0.163***
	[0.020]	[0.024]	[0.022]	[0.027]
Parental background	0.242***	0.109**	0.129***	0.026
	[0.035]	[0.043]	[0.039]	[0.047]
Education followers		✓		✓
Occupation followers			✓	✓
Cohort controls	✓	✓	✓	✓
Observations	3,399	3,392	3,363	3,356
R^2	0.046	0.096	0.084	0.126

Note:: A follower is an individual who coincides with their parent in either occupation or education. For males, followers are defined with respect to fathers; for females, with respect to mothers.

be addressed in future research[11].

For females, the unconditional association (column 1) shows that a 10 pp increase in the likelihood of a mother being informal is associated with a 2.42 pp increase in the daughter's lifetime informality. Accounting for education and occupation followers individually explains, as in the case of males, around half of this association in each case. When interactions for both education and occupation followers are included together, they account for nearly 90% of the initial association. The remaining association is positive but not statistically significant.

Crucially, this evidence suggests two important takeaways. First, parental background is strongly associated with informality, as parents with higher likelihoods of working

[11] Regressions of lifetime informality against parental background, controlling not only for followers but for all education and occupation outcomes, point in the same direction (see Appendix D).

informally tend to have children who spend more time in informality. Second, this association is primarily driven by intergenerational persistence in education and occupation, suggesting that the relationship operates through these choices.

3. Model

I propose a model of human capital investment with uncertainty and occupation choice. In the model, parents invest in the children's human capital, and children select into occupations, which might be formal or informal. Throughout the model, I define an occupation as a job, while informality refers to the nature of the worker's hiring arrangement.

The core mechanism in the model is that parents' labor market outcomes affect their investment in the child's human capital, which in turn impacts the child's future occupation choice. In particular, the model shows that parents facing greater income uncertainty invest less in human capital. In turn, children with lower levels of human capital are more likely to work informally and experience greater income uncertainty themselves. Ultimately, the main goal of the model is to quantify the extent to which a child's labor outcomes (education, occupation and informality) are related to parental income uncertainty, and derive implications for intergenerational mobility in education and occupations.

3.1. Setup

Consider two generations: parents and children. A household consists of one parent and one child and lives over two periods. In the first period, the parent invests in the child's human capital. In the second period, the child selects into an occupation and gets hired either formally or informally.

Parents are heterogeneous in terms of occupations j and human capital z. A parent's 'type' is given by the pair $\{j, z\}$. Occupations j belong to a finite set J, while human capital z can take one of two levels: high or low, denoted as $z \in \{z_l, z_h\}$. The sets of occupations and human capital levels are common for parents and children. Moreover, children have preferences for occupations, which are tied to their parent's occupation, constituting a channel for occupation persistence. Finally, in both periods there is uncertainty in labor income.

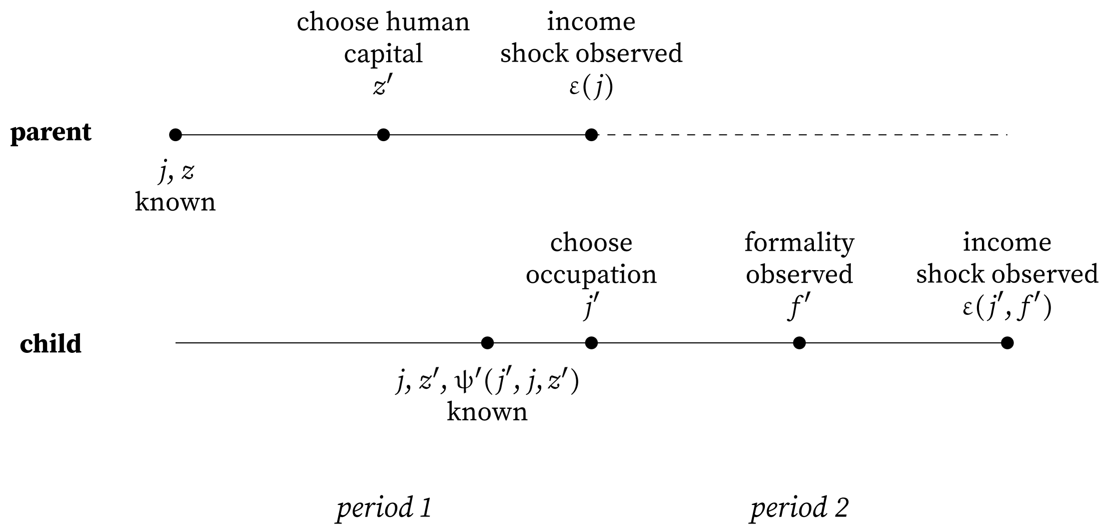
In the next section, I provide an overview of the model by describing its timing, followed by a more detailed explanation of the two periods.

Timing

The timing of the model is displayed in Figure 7. In the first period, a parent of type $\{j, z\}$ chooses the child's human capital z'. After this decision, the parent observes an occupation-specific shock $\varepsilon(j)$ to their labor income, which depends on occupation j.

In the second period, children draw a vector of preferences for occupations $\psi'(j', j, z')$. This vector characterizes the preference for all occupations $j' \in J$, conditional on their parent's occupation j, and their own human capital z'. Given preferences and the state variables j and z', children choose an occupation j'. Following this choice, they observe

FIGURE 7. Model timing



their formality status –either formal or informal, such that $f' \in (0;1)$–, and finally, they observe a shock $\varepsilon(j', f')$ to their labor income. This shock depends both on their occupation and formality status.

The key timing assumption is that the income shocks are observed only after the choices are made. Crucially, the human capital investment choice is made under uncertainty: parents are uncertain about their own income stream, their child's occupation, and the child's future income stream.

First period: Parents

In the first period, parents of type $\{j, z\}$ choose their child's human capital z'. By default, human capital is set at the lower level, z_l. For the child to acquire the higher human capital level, z_h, parents need to pay a cost. I assume this cost can be split into a pecuniary component $\bar{x}$, and an individual component x_i, which summarizes all non-pecuniary costs and benefits associated with the human capital investment choice. I further assume x_i is random and drawn from a Logistic distribution with location parameter μ_x and scale parameter β_x.

Finally, there is uncertainty in the income process: income $y(j, z)$ is determined by type-specific wages and a mean-preserving income shock $\varepsilon(j)$, which is occupation-specific, and is distributed with variance σ_j^2.

$$y(j, z) = w(j, z)\varepsilon(j) \tag{1}$$

Parents derive utility from both pecuniary and non-pecuniary sources. A parent's expected period utility, conditional on type $\{j, z\}$, is given by:

$$U_i(j,z) = \mathbb{E}\Big\{u\big[y(j,z) - \bar{x}\big]\Big\} + x_i \tag{2}$$

where the utility function exhibits prudence such that $u''' > 0$.

Second period: children

In the second period, children choose an occupation j', conditional on their parent's occupation and their own level of human capital determined in the previous period. When choosing occupation j', children are also assigned a formality status f', which refers to the way in which they are hired. I assume this assignment is random, conditional on occupation j' and human capital z': with probability $p(j', z')$, an individual is assigned a formal job ($f' = 1$); with probability $1 - p(j', z')$, an individual is assigned an informal job ($f' = 0$). These probabilities are known and exogenous[12].

Labor income is also uncertain and is spent solely on consumption. Income is defined as follows:

$$y(j', z', f') = w(j', z', f')\, \varepsilon(j', f'), \tag{3}$$

where wages $w(j', z', f')$ depend on occupation, human capital, and formality. The mean-preserving income shock $\varepsilon(j', f')$ depends on both occupation and formality status and is distributed with variance $\sigma^2_{j'f'}$.

Finally, children have preferences for occupations, denoted by $\psi(j', j, z')$, which can be interpreted as non-pecuniary costs and benefits associated with employment in occupation j'. I assume these preferences are derived upon entry to the labor market and consist of two components. The first component is determined by the parent's occupation, denoted by $\phi(j', j, z')$. This parameter captures the intergenerational transmission of occupational preferences, where a parent's occupation j directly influences the child's preference for occupation j'. In other words, having a parent in occupation j shapes the child's taste for that occupation. Additionally, I let this parameter vary with the child's level of human capital z' to reflect different patterns of selection across human capital level. The second component is an idiosyncratic preference for occupation j', denoted $\eta_i(j')$. I assume the idyosincratic preference is drawn from a type-1 extreme value distribution with scale parameter β_j.

[12] Implicitly, this assumption implies informality is determined by the demand-side, i.e. firms (for example, as in Ulyssea 2018), which I am not modeling here. In future work I plan to expand on this by endogeneizing informality probabilities. Finally, the fact that informality probabilities depend on occupation and human capital aims at capturing the well-known stylized facts.

Overall, preferences for occupation j', conditional on j and z' are written as:

$$\psi_i(j', j, z') = \phi(j', j, z') + \eta_i(j') \tag{4}$$

The expected utility of a child choosing occupation j' is then given by:

$$U_i'(j', j, z') = p(j', z')\,\mathbb{E}\left\{u[y(j', z', 1)]\right\} + \left(1 - p(j', z')\right)\mathbb{E}\left\{u[y(j', z', 0)]\right\} + \psi_i(j', z', j)$$

This expression highlights that, with probability $p(j', z')$, the child derives utility from the formal employment ($f' = 1$), while with probability $1 - p(j'z')$, the child derives utility from the informal employment. For convenience, I rewrite the expression as:

$$U_i'(j', j, z') = V'(j', j, z') + \phi(j', j, z') + \eta_i(j') \tag{5}$$

where $V'(j', j, z')$ represents the pecuniary component of the expected utility.

3.2. Household's Problem

In this section, I derive and characterize the model's solution. The two key endogenous objects are the share of parents that invest in their child's human capital, by parental type, and the share of children that select into different occupations, conditional on their background. I present the model's solution backwards, beginning with the child's problem.

Child's Problem: Occupation Choice

An individual with human capital z', preferences $\psi_i(j', j, z')$, and a parent in occupation j chooses occupation j' to solve the following problem:

$$\max_{j'} U_i'(j', j, z')$$

where $U_i'(j', j, z')$ satisfies Equation 1.5.

From the assumption that the idiosyncratic preferences $\eta_i(j')$ are drawn from a type-1 extreme value distribution with shape parameter β_j, the share of individuals choosing occupation j' is given by:

$$\mu(j', j, z') = \frac{\exp\left(\frac{V'(j', z') + \phi(j', j, z')}{\beta_j}\right)}{\sum_{j'} \exp\left(\frac{V'(j', z') + \phi(j', j, z')}{\beta_j}\right)} \tag{6}$$

From this expression, consider the determinants of occupational choice in the model. Clearly, the probability of selecting into occupation j' increases with the expected utility in that occupation, which is determined by the expected pecuniary utility $V(j', z')$ and the preference parameter $\phi(j', j, z')$.

In turn, $V(j', z')$ increases with $p(j', z')$ (the probability of formality) if there are utility gains from being employed in a formal job. These gains are determined by the income gap between formal and informal employment, and by the difference in the variance of the income shock between formal and informal employment. Specifically, the utility of having a formal job is higher when the income is greater and the variance of the income shock is smaller.

Similarly, $V(j', z')$ increases with human capital z' if there are expected gains from human capital in that occupation. These gains depend on the relationship between human capital and income, the relationship between human capital and the probability of formality, and the interaction between human capital, informality, income, and the variance of income shocks. Ultimately, the share of individuals choosing occupation j' increases with human capital if the expected returns in occupation j' are larger than in other occupations. Further discussion of these results can be found in Appendix E.

Parent's Problem: Human Capital Choice

A parent maximizes both their own utility and the child's future utility and decides whether to invest in human capital or not. If the parent invests, they pay the pecuniary ($\bar{x}$) and non-pecuniary costs (x_i), and the child acquires the higher level of human capital z_h. If the parent does not invest, no costs are incurred, and the child's human capital is set at the lower level z_l. The parent will then choose z' by comparing the values of the two possible outcomes. Note that parents do not know their children's preferences for occupations, so their occupation choice is uncertain. Therefore, parents compute the expected utility of their child participating in the labor market with low versus high human capital.

The value of investing in human capital, v_i, conditional on the parent's type and x_i is given by:

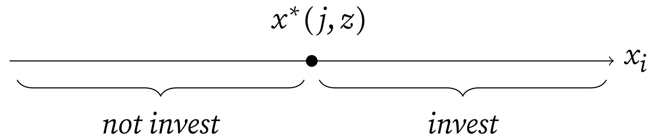
$$v^i(j, z, x_i) = \mathbb{E}\left\{u\left[y(j, z) - \bar{x}\right]\right\} + x_i + \alpha\beta_j \log\left(\sum_{j'} e^{\frac{V'(j', z_h) + \phi(j', j, z_h)}{\beta_j}}\right) + \tilde{\gamma}\beta_j \tag{7}$$

where α denotes the altruism parameter, and $\tilde{\gamma}$ is the Euler-Mascheroni constant.

The value of not investing is given by:

$$v^{ni}(j, z) = \mathbb{E}\left\{u\left[y(j, z)\right]\right\} + \alpha\beta_j \log\left(\sum_{j'} e^{\frac{V'(j', z_l) + \phi(j', j, z_l)}{\beta_j}}\right) + \tilde{\gamma}\beta_j \tag{8}$$

FIGURE 8. Parent's investment choice, conditional on type



The parent's choice is the solution to:

$$\max\left\{v^i(j,z,x_i),\ v^{ni}(j,z)\right\}$$

A parent of type $\{j,z\}$ optimally chooses to invest when $v^i(j,z,x_i) \geq v^{ni}(j,z)$. Since v^i is increasing in x_i and v^{ni} does not depend on x_i, there exists a threshold $x^*(j,z)$ for which parents of type $\{j,z\}$ are indifferent between investing and not investing. In other words, for $x_i = x^*$, the gains from increased human capital in the next generation perfectly compensate the associated cost:

$$x^*(j,z) : v^i(j,z,x_i) = v^{ni}(j,z)$$

Substituting Equations 7 and 8, the definition of the threshold by parental type is:

$$\begin{aligned} x^*(j,z) = {} & \mathbb{E}\left\{u\left[y(j,z)\right]\right\} - \mathbb{E}\left\{u\left[y(j,z) - \bar{x}\right]\right\} \\ & - \alpha\beta_j\left[\log\left(\sum_{j'} e^{\frac{V'(j',z_h)+\phi(j',j,z_h)}{\beta_j}}\right) - \log\left(\sum_{j'} e^{\frac{V'(j',z_l)+\phi(j',j,z_l)}{\beta_j}}\right)\right] \end{aligned} \tag{9}$$

A parent invests whenever $x_i \geq x^*(j,z)$, that is, whenever the net non-pecuniary benefit of investing is "sufficiently high" (Figure 8). Note the level of the threshold is dependent on the parent's type $\{j,z\}$ since both the parent's utility and the child's future utility vary with the parent's type. It is straightforward to see that the higher the threshold, the lower the share of parents that will invest.

Given the assumptions on the distribution of the non-pecuniary cost x_i, the share of parents of type $\{j,z\}$ who invest is defined as follows:

$$s(j,z) = 1 - F_x\left[x^*(j,z)\right] = 1 - \frac{1}{1 + e^{-\frac{x^*(j,z)}{\beta x}}} \tag{10}$$

where F_x is the CDF of a Logistic distribution.

From Equations 9 and 10, consider the determinants of the share of parents who invest conditional on type, $s(j,z)$. First, note that the share $s(j,z)$ is increasing in the altruism parameter α and in the expected returns to human capital in the next generation (which, as discussed earlier, are determined by the joint effect of human capital on income, on the probability of formality, and on the interaction between formality probability, income, and the variance of the income shock). Moreover, $s(j,z)$ decreases with the costs of education, increases with income $w(j,z)$, and –under the prudence assumption on the utility function– decreases with the variance of the parental income shock. In other words, higher income uncertainty in occupation j decreases the share of parents of type $\{j,z\}$ who invest. Further discussion and the derivation of this result are presented in Appendix E.

3.3. Labor demand and market clearing

To close the model, I assume that, in both periods, labor demand H is such that the labor market clears. In the first period:

$$H = \sum_j \sum_z N(j,z) \tag{11}$$

where $N(j,z)$ is the supply of parents of type $\{j,z\}$, which is exogenous in the model.

In the second period:

$$H' = \sum_j \sum_z N'(j,z) \sum_{j'} \left[\mu(j',j,z_h)s(j,z) + \mu(j',j,z_l)\,(1 - s(j,z)) \right] \tag{12}$$

where $N'(j,z)$ is the supply of children with parents of type $\{j,z\}$.

3.4. Discussion: Aggregate Implications

From the key endogenous variables, that is, the share of parents who invest, by type, and the children's occupational choices, I derive aggregate implications for intergenerational persistence in education and occupations and for informality.

Education Persistence. Define education persistence, e, as the share of individuals that attain the same human capital level as their parent. This can be written as a function of the endogenous variables as follows:

$$e = s_i(j,z_h)n(j,z_h) + (1 - s_i(j,z_l))\,n(j,z_l) \tag{13}$$

- $n(j,z_h)$ and $n(j,z_l)$ are shares of parents with high and low human capital

Occupation Persistence. Define occupation persistence, o, as the share of children that work in the same occupation as their parent. This can be written as:

$$o = \sum_j \sum_z n'(j,z) \left[\mu(j', j, z_h) s_i(j,z) + \mu(j', j, z_l) \left(1 - s_i(j,z)\right) \right] \quad \text{for } j' = j \tag{14}$$

where $n'(j,z)$ is the share of children with parents of type $\{j,z\}$.

Informality. The informality rate, i, is the weighted sum of the share of informal workers by occupation and human capital level. It can be written as a function of the endogenous variables and parameters $n'(j,z)$ and $p(j,z)$ as follows:

$$i = \sum_j \sum_z n'(j,z) \sum_{j'} \left[\left(1 - p(j', z_h)\right) \mu(j', j, z_h) s^i(j,z) + \left(1 - p(j', z_l)\right) \mu(j', j, z_l) \left(1 - s^i(j,z)\right) \right] \tag{15}$$

Finally, in a similar way, we can write the overall share of high skilled individuals and the aggregate employment shares by occupations as a function of the endogenous variables.

4. Quantitative Exercise

In this section, I present the steps to conduct the quantitative exercise. The estimation strategy involves fixing certain parameters and calibrating others to align the model with the observed data. The following sections detail the steps involved in the quantitative analysis and evaluate the model fit.

4.1. Data and definitions

The primary data source for the quantitative analysis is the EPS-Chile, covering the years 2002 to 2020. I restrict the sample to individuals aged 25 to 60 and to pairs of fathers and sons.

Occupations and Human Capital. The number of occupations, J, is set to 9, corresponding to the 1-digit level of occupational classifications according to ISCO-08. Human capital levels are defined based on education levels. For fathers, these levels are: (i) less than high school, and (ii) completed high school or higher. For sons, the levels are: (i) up to high school, and (ii) more than completed high school.

Given these definitions, I classify individuals into groups based on their education and occupation. Fathers are classified using retrospective information reported by their sons. Sons are assigned a level of education and a main occupation based on the most frequently observed values in the panel data. This approach is representative of labor histories, as there is strong persistence in occupations at the individual level, as noted in the empirical analysis. For males, around 50% of individuals spend their entire observed working life in their main occupation, and over 93% spend more than half of that time in it.

Informality. As before, I define informality as working as a wage worker without a contract or being self-employed in non-professional or non-technical occupations. Sons are classified as formal or informal workers based on their most frequently observed formality status. This classification is representative of labor histories, as 71% of males spend their entire observed working life in either formal or informal employment spells. Formal job probabilities $p(j,z)$, are computed separately for parents and children, based on the "lifetime" informality measure when dividing the sample in two (born before and after 1970).

4.2. Further assumptions for identification

Utility Function. For the utility of consumption, I assume a Constant Relative Risk Aversion (CRRA) utility function with parameter γ:

$$U(c) = \frac{c^{1-\gamma}}{1-\gamma}$$

Labor income and lifetime utilities. In the model, labor income is defined as lifetime income by type, for parents:

$$y(j,z) = w(j,z)\varepsilon(j) \tag{16}$$

and for children:

$$y(j',z',f') = w(j',z',f')\varepsilon(j',f') \tag{17}$$

In order to bring this to the data, I assume both generations participate in the labor market during T periods, and each period they get income y_t, defined as:

$$\log y_t = a + bt + \varepsilon_t \tag{18}$$

and

$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t \quad \nu_t \sim N(0, \sigma_\nu^2) \tag{19}$$

Note this implies:

$$\log y_t \sim N(a + bt, \frac{\sigma_v^2}{1-\rho^2}) \tag{20}$$

The parameters a, b, ρ and σ_v^2 are group and generation-specific. Crucially, ρ and σ_v^2 vary with occupation and formality status.

In turn, for each generation, total utility is now the sum of the discounted flow of utilities during T periods:

$$U = \sum_{t=1}^{T} \beta^{t-1}\mathbb{E}\{u(c_t)\} \tag{21}$$

The nature of the model and its solution remain unchanged, as I still assume that both parents and children have only one discrete choice. For children, this implies a single occupation choice; for parents, a choice of whether to invest in human capital or not. In particular, under this formulation, I assume that if parents decide to invest, they commit to some cost x each period.

Ultimately, the stream of lifetime utilities is rewritten as follows. For children, I rewrite $V(j', z')$ of Equation 6 of the model as:

$$V(j', z') = p(j,z)\sum_{t=1}^{T}\beta^{t-1}\mathbb{E}\left\{u\left[y_t(j', z', 1)\right]\right\} + \left[1 - p(j,z)\right]\sum_{t=1}^{T}\beta^{t-1}\mathbb{E}\left\{u\left[y_t(j', z', 0)\right]\right\} \tag{22}$$

where $\log y_t \sim N\left(a(j', z', f') + b(j', z', f'), \frac{\sigma_v^2(j', f')}{1-\rho^2(j', f')}\right)$

For parents, the expected utility terms in Equation 9 are rewritten as:

$$\mathbb{E}\{u\left[y(j,z)\right]\} = \sum_{t=1}^{T}\beta^{t-1}\mathbb{E}\{u\left[y_t(j,z)\right]\} \tag{23}$$

$$\mathbb{E}\{u\left[y(j,z) - \bar{x}\right]\} = \sum_{t=1}^{T}\beta^{t-1}\mathbb{E}\{u\left[y_t(j,z) - x_t\right]\} \tag{24}$$

where $\log y_t \sim N\left(a(j,z) + b(j,z), \frac{\sigma_v^2(j)}{1-\rho^2(j)}\right)$

The expectations are computed in closed-form as detailed in Appendix E. Finally, I set $T = 12$, spanning from ages 25 to 60. Each period t broadly corresponds to 3 years, in order to maintain consistency with the average spacing of the EPS survey waves (see Appendix A).

4.3. Calibration strategy

In this section, I detail the steps for the model calibration. A summary of the parameters, their sources, and the estimation procedure is presented in Table 7.

TABLE 7. Calibration summary

Parameter	Description	Value	Source/Target
γ	CRRA parameter	2	-
β	Discount factor	0.885	-
α	Altruism parameter	0.32	Lee and Seshadri (2019)
p_f	Probability formal		Shares of formal workers by group
a, b_t, ρ, σ_v^2	Wage process parameters		Income and income growth, autocorrelation of error terms, variance of residuals
ϕ β_j	Preferences for occupations and scale parameter	 0.597	Occupation shares and dispersion
$\log x$ μ_x β_x	Cost of education and logistic distribution parameters	10.77 3.961 1.052	Shares of high-skilled sons by father's type

Note: The values for the formality probabilities and the wage process parameters are detailed in Appendix F.

First, I fix some parameters exogenously. The CRRA parameter is set to the usual value of $\gamma = 2$ and the annual discount factor is set to 0.96. Since in the model one period is around three calendar years, I set $\beta = 0.96^3 = 0.885$. The altruism parameter is set to $\alpha = 0.32$ (Lee and Seshadri 2019)

Next, I estimate a set of parameters outside the model. These are: the probabilities of formal workers by occupation and human capital, which is calculated as detailed before, and the parameters of the wage process. These are obtained from a reduced form approach. First, I run a regression of the form:

$$\ln y_{it} = \alpha + \beta x_{it} + \lambda_t + \epsilon_{it}$$

where x_{it} is a set of controls for age group, education level, occupation and informality. The values for a are obtained from the intercept of the regression, by group, while those for b, the time trend, are obtained from the coefficient for age group. Next, I regress the residuals ϵ_{it} on its lagged values separately for each combination of occupation and informality to pin down the autorregressive parameter ρ and the the variance of residuals ν_t, σ_ν^2, by group:

$$\epsilon_{it} = \rho\epsilon_{it-1} + \nu_t$$

The values of the parameters estimated outside the model are detailed in Appendix F. In particular, the estimates for σ_v are consistent with the empirical analysis: within all occupations, the variance of the shock is larger for informal workers.

The rest of the parameters is estimated internally by the simulated method of moments (SMM). I proceed in two steps. First, I recover the preferences for occupations and the extreme value parameter β_j, which are informed by the occupation shares and their dispersion. Second, I estimate the parameters that drive the parental human capital investment choice, which are informed by the shares of high-educated children by parental type.

Estimation of occupation choice parameters

From equation 6 in the model, the relative occupation share between two occupations j' and l', conditional on parent's occupation j and own human capital z' is:

$$\ln\left[\frac{\mu(j',j,z')}{\mu(l',j,z')}\right] = \frac{1}{\beta_j}\left[V'(j',z) - V'(l',z) + \phi(j,j',z') - \phi(j,l',z')\right]$$

where the expected utility terms $V(j',z')$ and $V(l',z')$ are computed as defined in Equation 22.

I set the relative preferences $\phi(j',j,z') - \phi(l',j,z')$ for every pair of occupations j' and l' and β to match the observed occupation shares and their dispersion in the data.

The scale parameter is estimated to $\beta = 0.5968$. Figure 9 displays the average relative preference for the father's occupation, conditional on education. Specifically, this is given by:

$$\frac{1}{9}\sum_{l'=1}^{9}\phi(j',j,z') - \phi(l',j,z')$$

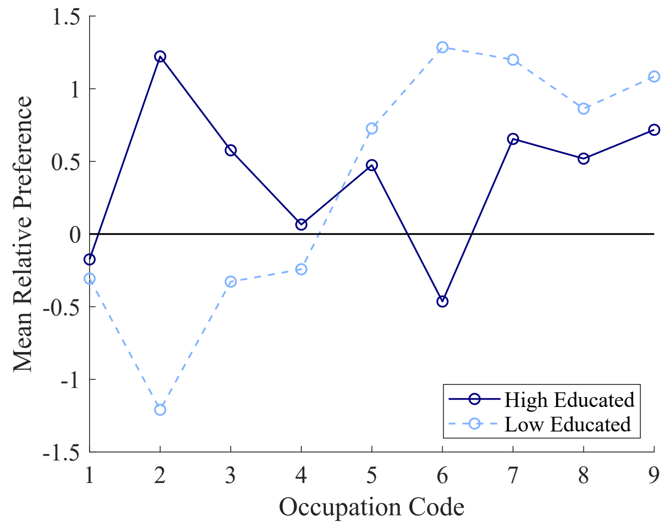
for $j' = j$.

In other words, I compute the average relative preference for the father's occupation relative to all other occupations. Positive numbers indicate that the preference for the father's occupation is larger than for other occupations. Figure 9 shows that this is indeed the case for most occupations and both education levels. Overall, the patterns broadly coincide with what was documented in the empirical section: the recovered preferences for white-collar occupations are larger for high-educated individuals; in the other occupations, they are larger for low-skilled individuals. Appendix F gives more detail on the estimated relative preferences.

Estimation of human capital investment choice parameters

The costs of education x and parameters of the logistic distribution μ_x and β_x are informed by the shares of high-educated children by parental type. Once the preferences parameters are recovered as detailed in the previous section, I use equations 9 and 10

FIGURE 9. Average relative preferences for father's occupation, by occupation and education level



Note: The figure shows the average relative preference for the father's occupation. Positive numbers indicate that the average preference for the father's occupation is larger than for other occupations. Occupational codes (1-digit level) are as follows: (1) Managers; (2) Professionals; (3) Technicians and Associated Professionals; (4) Clerical Support Workers; (5) Service and Sales Workers; (6) Agricultural, Forestry and Fishery Workers; (7) Craft and Related Trades Workers; (8) Operators and Assemblers; (9) Elementary Occupations.

of the model, and iterate over x, μ_x, β_x to minimize the distance between predicted investment shares and those observed in the data.

The logistic distribution parameters are estimated to $\mu_x = 3.9$ and $\beta_x = 0.81$. The education cost is approximated to $\log x_t = 10.77$, which corresponds to around 15% of the mean income.

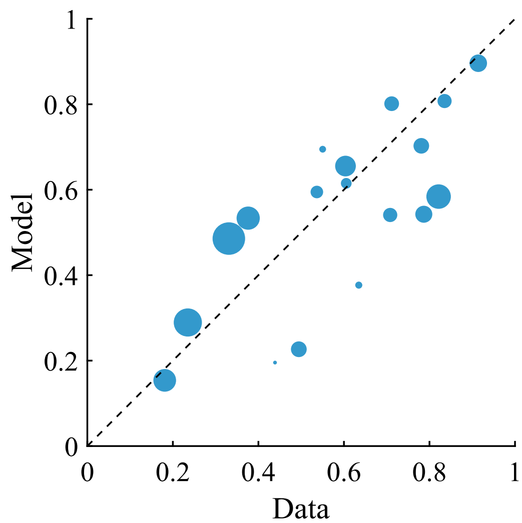
4.4. Model fit

In this section I comment on the overall model fit, as illustrated in Figures 10 and 11 and in Table 8.

Figure 10 compares the predicted human capital investment shares by parental type with those observed in the data. The model captures the heterogeneity in investment shares across fathers' types, although some discrepancies remain between the observed and predicted shares. Overall, as shown in Table 8, the model underestimates the proportion of high-educated individuals, which is 45.8% in the model compared to 47% observed in the data. Nevertheless, since occupation shares by education level are exactly matched, the model reproduces the overall employment shares by occupation well (Figure 11).

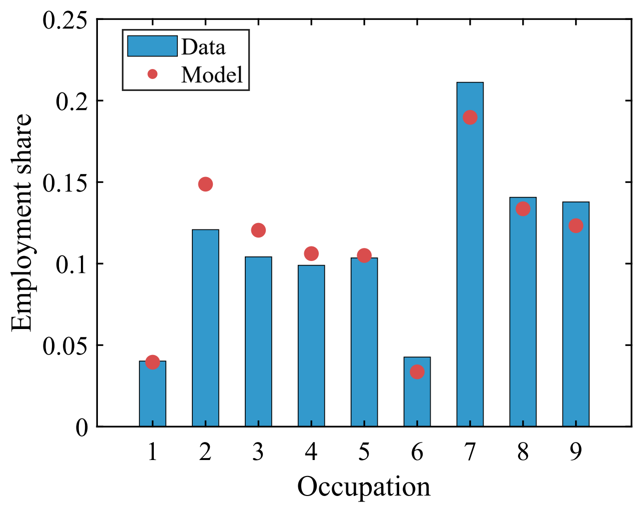
Regarding intergenerational persistence in education and occupations, Table 8 shows that the model underestimates both measures. However, it accurately reflects the pattern that persistence is higher among individuals with lower levels of education. Finally, as a result of the patterns in Figures 10 and 11, the model closely matches overall informality rates as well as informality rates by education level.

FIGURE 10. Simulated vs observed human capital investment shares



Note: The figure shows the predicted human capital investment shares by parental type against those observed in the data. Each dot represents a parent type, defined by the combination of occupation and education, and the size of the dot represents the share of a parental type in the data.

FIGURE 11. Simulated vs observed employment shares by occupation



Note: Occupational codes (1-digit level) are as follows: (1) Managers; (2) Professionals; (3) Technicians and Associated Professionals; (4) Clerical Support Workers; (5) Service and Sales Workers; (6) Agricultural, Forestry and Fishery Workers; (7) Craft and Related Trades Workers; (8) Operators and Assemblers; (9) Elementary Occupations.

TABLE 8. Model fit and results of counterfactual exercises

	Data	Model	Counterfactual A	Counterfactual B
Share high-skilled	0.512	0.508	0.623	0.545
Education persistence				
Overall	0.720	0.634	0.556	0.612
High-skilled	0.754	0.669	0.713	0.686
Low-skilled	0.699	0.608	0.442	0.558
Occupation persistence				
Overall	0.228	0.217	0.198	0.210
High-skilled occ. (1-3)	0.584	0.599		0.604
Low-skilled occ. (4-9)	0.763	0.745		0.738
Informality				
Overall	0.249	0.249	0.229	0.243
Intergen. assoc. (correlation)	0.698	0.706		0.686

5. Counterfactual exercise

To evaluate the role of informality in shaping intergenerational persistence and children's future labor market outcomes, I conduct two counterfactual exercises using the structural model.

In the first couterfactual "A", I measure the overall effect of the presence of informality on intergenerational outcomes, by assume no informality in the parent's generation. Specifically, I set the probability of working in a formal job $p(j, z)$ equal to 1 for all j and z. This eliminates both the wage differences and the higher uncertainty associated with informality. The results are displayed in Table 8 and Figure 12, and show a substantial impact on human capital investment and intergenerational outcomes.

In the second counterfactual "B", I isolate the role of income uncertainty by equalizing the variance of income shocks across formal and informal employment in the parent's generation. That is, I set $\sigma^2_v(j, f = 0) = \sigma^2_v(j, f = 1)$ for all occupations j, thereby removing the additional income volatility associated with informality, while keeping the other variables unchanged. The results are displayed in Table 8 and Figure 13. Overall, the share of high-educated sons increases by 3.6 pp, from a baseline share of 50.8%. Driven by this, overall intergenerational persistence in education and in occupations decrease by 2.2 pp and 0.6 pp respectively, from their baseline levels of 63.4% and 21.7%.

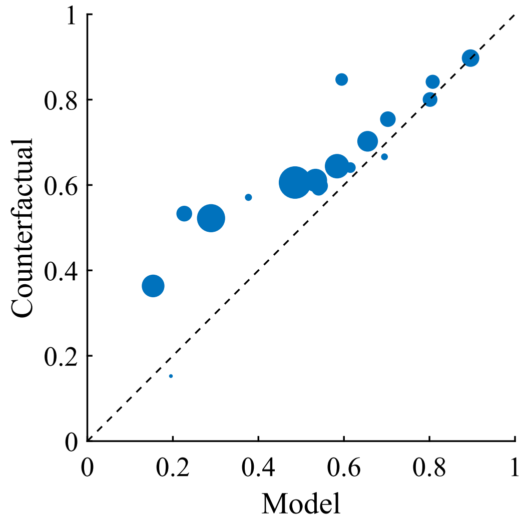
These results appear modest at the aggregate level, but are more pronounced when considering patterns of upward mobility. Among children of low-educated parents, educational persistence declines by 5 pp, while among children of parents in lower-skilled occupations, occupational persistence decreases by 1.7 pp. These results imply a weaker intergenerational association between the father's likelihood of informality and the son's lifetime informality. Overall, the exercise shows that the uncertainty channel alone can explain between a quarter and a third of the total changes observed in the first counterfactual.

I interpret the results as lower-bound effects, as they capture only uncertainty related to labor income, while other aspects of informality, such as differential unemployment risk or reduced access to non-labor income such as social security benefits, could likely amplify these effects.

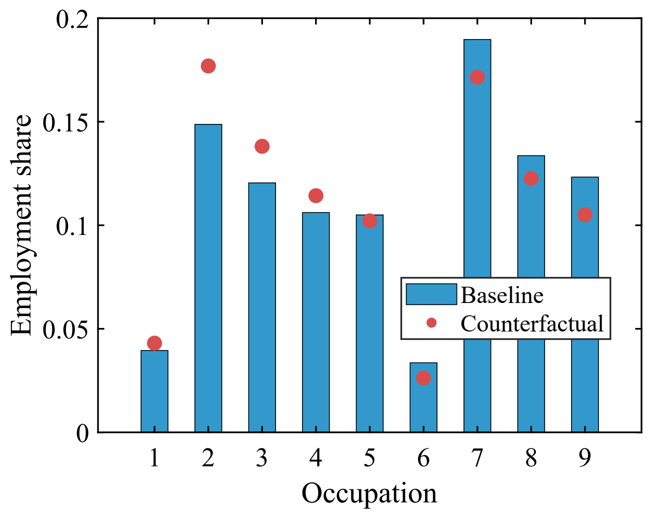
While the counterfactual exercises are theoretical in nature, they offer relevant insights that will be explored in future work. In particular, Counterfactual B suggests that reducing income risk, even without eliminating informality, can still yield substantial intergenerational benefits. In practice, this could be achieved through income-smoothing measures such as non-contributory transfers (by providing a stable income floor for informal workers) or basic social protection coverage (which would reduce the overall risk of total income).

FIGURE 12. Baseline vs Counterfactual A

A. Investment shares by parent type



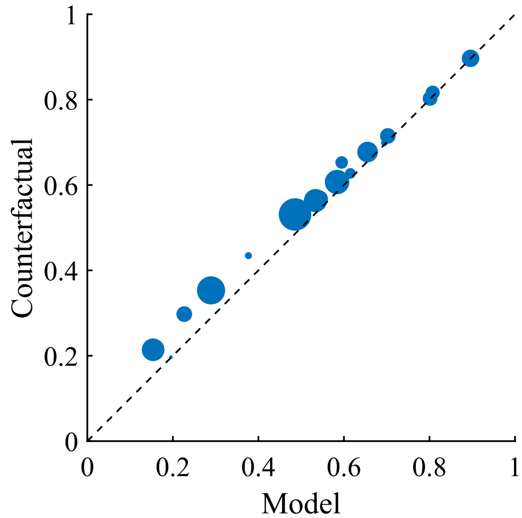
B. Employment shares by occupation



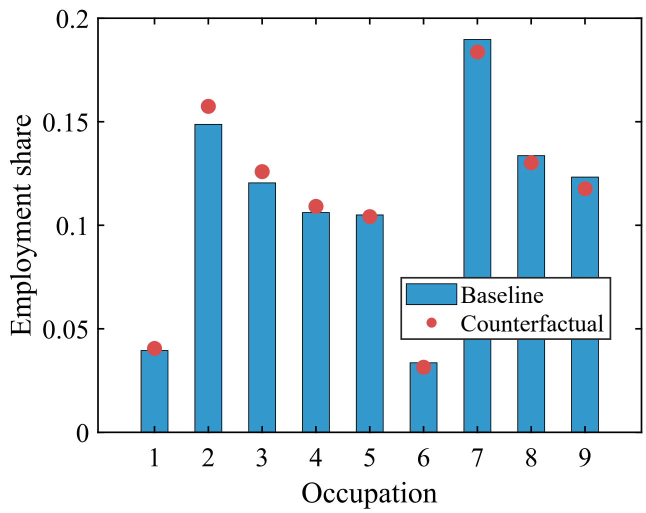
Note: Occupational codes (1-digit level) are as follows: (1) Managers; (2) Professionals; (3) Technicians and Associated Professionals; (4) Clerical Support Workers; (5) Service and Sales Workers; (6) Agricultural, Forestry and Fishery Workers; (7) Craft and Related Trades Workers; (8) Operators and Assemblers; (9) Elementary Occupations.

FIGURE 13. Baseline vs Counterfactual B

A. Investment shares by parent type



B. Employment shares by occupation



Note: Occupational codes (1-digit level) are as follows: (1) Managers; (2) Professionals; (3) Technicians and Associated Professionals; (4) Clerical Support Workers; (5) Service and Sales Workers; (6) Agricultural, Forestry and Fishery Workers; (7) Craft and Related Trades Workers; (8) Operators and Assemblers; (9) Elementary Occupations.

6. Conclusion

In this paper, I take a step towards studying the link between labor informality and intergenerational mobility. I focus specifically on one characteristic of informal employment: income uncertainty. I show that greater uncertainty is a potential mechanism through which intergenerational mobility in both education and labor market outcomes is constrained. This effect is primarily driven by reduced parental investment in children's human capital, which favors occupational following and increases the likelihood of informality, potentially creating a cycle of persistent exposure to uncertainty across generations. Indirectly, this provides suggestive evidence of intergenerational persistence in labor informality, an issue I do not address in the current paper due to data limitations, but which I plan to further explore in future research.

In terms of the current work, other potential extensions include endogenizing the probabilities of obtaining formal employment to better understand the determinants of selection into informal and formal employment. Additionally, the current measure of income uncertainty could be expanded to account for the presence of other workers in the household and for non-labor income, such as social security benefits, which represent another relevant discrepancy between formal and informal work.

References

Abbott, Brant. 2022. "Incomplete markets and parental investments in children." *Review of Economic Dynamics* 44: 104–124.

Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni L Violante. 2019. "Education policy and intergenerational transfers in equilibrium." *Journal of Political Economy* 127 (6): 2569–2624.

Agostinelli, Francesco, Domenico Ferraro, Xincheng Qiu, and Giuseppe Sorrenti. 2024. "Intra-Household Insurance and the Intergenerational Transmission of Income Risk.", National Bureau of Economic Research.

Albrecht, James, Lucas Navarro, and Susan Vroman. 2009. "The effects of labour market policies in an economy with an informal sector." *The Economic Journal* 119 (539): 1105–1129.

Becker, Gary S, and Nigel Tomes. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." *Journal of Political Economy* 87 (6): 1153–1189.

Becker, Gary S, and Nigel Tomes. 1986. "Human capital and the rise and fall of families." *Journal of Labor Economics* 4 (3, Part 2): S1–S39.

Bobba, Matteo, Luca Flabbi, and Santiago Levy. 2022. "Labor market search, informality, and schooling investments." *International Economic Review* 63 (1): 211–259.

Bosch, Mariano, and Julen Esteban-Pretel. 2012. "Job creation and job destruction in the presence of informal markets." *Journal of Development Economics* 98 (2): 270–286.

Britto, Diogo, Alexandre de Andrade Fonseca, Paolo Pinotti, Breno Sampaio, and Lucas Warwar. 2022. "Intergenerational mobility in the land of inequality."

Brunori, Paolo, Francisco HG Ferreira, and Guido Neidhöfer. 2024. "Inequality of opportunity and intergenerational persistence in Latin America.", IZA Discussion Papers.

Carneiro, Pedro, Italo Lopez Garcia, Kjell G Salvanes, and Emma Tominey. 2021. "Intergenerational mobility and the timing of parental income." *Journal of Political Economy* 129 (3): 757–788.

Constant, Amelie F, and Klaus F Zimmermann. 2003. "Occupational choice across generations." *Available at SSRN 487416*.

Corak, Miles, and Patrizio Piraino. 2011. "The intergenerational transmission of employers." *Journal of Labor Economics* 29 (1): 37–68.

Dix-Carneiro, Rafael, Pinelopi K Goldberg, Costas Meghir, and Gabriel Ulyssea. 2024. "Trade and Domestic Distortions: The Case of Informality.", National Bureau of Economic Research.

Doepke, Matthias, and Fabrizio Zilibotti. 2008. "Occupational choice and the spirit of capitalism." *The Quarterly Journal of Economics* 123 (2): 747–793.

Dunn, Christopher E. 2007. "The intergenerational transmission of lifetime earnings: Evidence from Brazil." *The BE Journal of Economic Analysis & Policy* 7 (2).

Engbom, Niklas, Gustavo Gonzaga, Christian Moser, and Roberta Olivieri. 2022. "Earnings inequality and dynamics in the presence of informality: The case of Brazil." *Quantitative Economics* 13 (4): 1405–1446.

Erosa, Andrés, Luisa Fuster, and Tomás R Martinez. 2023. "Public financing with financial frictions and underground economy." *Journal of Monetary Economics* 135: 20–36.

Escriche, Luisa. 2007. "Persistence of occupational segregation: The role of the intergenerational transmission of preferences." *The Economic Journal* 117 (520): 837–857.

Gasparini, Leonardo, and Leopoldo Tornarolli. 2009. "Labor informality in Latin America and the Caribbean: Patterns and trends from household survey microdata." *Desarrollo y sociedad* (63):

13–80.

Gomes, Diego BP, Felipe S Iachan, and Cezar Santos. 2020. "Labor earnings dynamics in a developing economy with a large informal sector." *Journal of Economic Dynamics and Control* 113: 103854.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2021. "What do data on millions of US workers reveal about lifecycle earnings dynamics?" *Econometrica* 89 (5): 2303–2339.

Haanwinckel, Daniel, and Rodrigo R Soares. 2021. "Workforce composition, productivity, and labour regulations in a compensating differentials theory of informality." *The Review of Economic Studies* 88 (6): 2970–3010.

Hellerstein, Judith K, and Melinda Sandler Morrill. 2011. "Dads and daughters: The changing impact of fathers on women's occupational choices." *Journal of Human Resources* 46 (2): 333–372.

Kramarz, Francis, and Oskar Nordström Skans. 2014. "When strong ties are strong: Networks and youth labour market entry." *Review of Economic Studies* 81 (3): 1164–1200.

Laband, David N, and Bernard F Lentz. 1983. "Like father, like son: Toward an economic theory of occupational following." *Southern Economic Journal*: 474–493.

Lee, Sang Yoon, and Ananth Seshadri. 2019. "On the intergenerational transmission of economic status." *Journal of Political Economy* 127 (2): 855–921.

Leites, Martín, Xavier Ramos, Cecilia Rodríguez, and Vilá Joan. 2022. "Intergenerational mobility along the income distribution: estimates using administrative data for a developing country."

Lo Bello, Salvatore, and Iacopo Morchio. 2022. "Like father, like son: Occupational choice, intergenerational persistence and misallocation." *Quantitative Economics* 13 (2): 629–679.

Long, Jason, and Joseph Ferrie. 2013. "Intergenerational occupational mobility in Great Britain and the United States since 1850." *American Economic Review* 103 (4): 1109–1137.

López García, Italo. 2015. "Human capital and labor informality in Chile: a life-cycle approach."

Magruder, Jeremy R. 2010. "Intergenerational networks, unemployment, and persistent inequality in South Africa." *American Economic Journal: Applied Economics* 2 (1): 62–85.

Maurizio, Roxana, and Ana Paula Monsalvo. 2021. "Informality, labour transitions, and the livelihoods of workers in Latin America.", WIDER Working Paper.

Meghir, Costas, Renata Narita, and Jean-Marc Robin. 2015. "Wages and informality in developing countries." *American Economic Review* 105 (4): 1509–1546.

Munoz, Ercio, and Roy Van der Weide. 2025. "Intergenerational income mobility around the world: A new database.", The World Bank.

Narita, Renata. 2020. "Self-employment in developing countries: A search-equilibrium approach." *Review of Economic Dynamics* 35: 1–34.

Neidhöfer, Guido, Matías Ciaschi, and Leonardo Gasparini. 2022. "Intergenerational mobility of economic well-being in Latin America.", Documento de Trabajo.

Neidhöfer, Guido, Joaquín Serrano, and Leonardo Gasparini. 2018. "Educational inequality and intergenerational mobility in Latin America: A new database." *Journal of development economics* 134: 329–349.

Nunez, Javier I, and Leslie Miranda. 2010. "Intergenerational income mobility in a less-developed, high-inequality context: The case of Chile." *The BE Journal of Economic Analysis & Policy* 10 (1).

Ulyssea, Gabriel. 2018. "Firms, informality, and development: Theory and evidence from Brazil." *American Economic Review* 108 (8): 2015–2047.

Ulyssea, Gabriel. 2020. "Informality: Causes and consequences for development." *Annual Review of Economics* 12 (1): 525–546.

Van der Weide, Roy, Christoph Lakner, Daniel Gerszon Mahler, Ambar Narayan, and Rakesh Gupta. 2024. "Intergenerational mobility around the world: A new database." *Journal of Development Economics* 166: 103167.

Appendix A. The EPS survey

A.1. The EPS survey

The *Encuesta de Protección Social* (EPS) is a longitudinal survey conducted in Chile. It has so far 7 rounds: 2002, 2004, 2006, 2009, 2012, 2015, and 2020; the 8th round release is expected for late 2024. The survey is representative at the national level for individuals aged 18 and over since 2004 and can be used for both cross-sectional and longitudinal analysis. The original 2002 sample was expanded in 2004 to include individuals who were not affiliated with the pensions system. Further updates to the sample were conducted in 2012 and 2015 to address attrition and ensure the representativeness of the target population.

I put together and homogenized the data from the seven waves of the EPS survey. After the usual consistency checks —gender, age, education— I also put together a panel dataset. The full dataset contains information on 34,392 distinct individuals and a total of 261,089 labor market spells. Among the individuals in the panel, the average number of appearances is 4.4, which means that, on average, they can be tracked for around 11.5 years. Table A1 lists the data collection period and sample size for each wave of the EPS.

TABLE A1. Data collection period and sample size by wave, EPS

Round	Data collection	Sample size	
		Individuals	Spells
2002	May 2002 – Jan 2003	16,962	76,344
2004	Nov 2004 – May 2005	16,727	29,896
2006	Nov 2006 – July 2007	16,443	27,514
2009	Apr 2008 – Apr 2009	14,463	23,865
2012	Sep 2012 – Aug 2013	15,998	27,020
2015	Apr 2016 – Jul 2016	16,906	46,657
	Dec 2019 – Mar 2020 (in person)	7,800	11,145
2020	Sep 2020 – Dec 2020 (on the phone)	5,031	15,795
	Aug 2020 – Oct 2020 (re-interviews)	2,082	2,853

Note: The 2020 wave was interrupted by Covid-19, leading to the implementation of three separate subwaves: in person, on the phone, and re-interviews. The re-interviews were conducted to individuals who had already been interviewed in 2020 in order to assess the short-term effects of the pandemic.

Source: EPS.

A.2. Descriptive statistics

In the following tables, I provide descriptive statistics of the baseline sample, which includes individuals aged 25 to 60, separately for males and females. Table A2 presents demographic and labor market statistics of the panel dataset. The baseline sample consists of 22,737 individuals and a total of 74,614 spells.

Table A3 shows summary statistics for the cross-section data for selected waves. As shown in the table, there was a noticeable educational expansion throughout the survey waves, alongside a shift toward professional and technical occupations. The informality rate also shows a decreasing trend. The table also shows the informality rate using different definitions, and I show that the levels and trends are similar. In addition to the main measure, the other two criteria are: (i) a wage worker is considered formal if they have a signed contract, while all self-employed workers are considered informal (as in Dix-Carneiro et al. 2024 or Meghir, Narita, and Robin 2015); (ii) a worker is considered formal if they contribute to social security and informal otherwise, a definition commonly used in official statistics.

Finally, in Table A4, I present descriptive statistics on informality for the pooled cross-sectional data. The table displays overall informality, informality among wage workers, and the share of self-employment by demographic and occupational groups, separately for males and females. Informality rates are higher among older workers, those with lower education levels, self-employed individuals, and workers in low-skilled occupations. Overall, informality rates are similar for men and women in the baseline sample; however, women exhibit higher informality rates among wage workers and a lower share of self-employment compared to men.

TABLE A2. Summary statistics. Panel dataset

	Males	Females
Overall share	0.503	0.497
Education (shares)		
Less than high-school	0.293	0.299
High-school	0.263	0.303
More than high-school	0.445	0.398
Cohort (shares)		
1940-1954	0.098	0.106
1955-1969	0.227	0.249
1970-1984	0.347	0.323
1985-1996	0.328	0.322
Average number of spells		
All spells	2.998	2.991
Employment	2.186	1.645
Unemployment	0.593	0.679
Non-participation	0.219	0.667
Obs. (sample)	11,177	11,560
Spells	32,414	34,014

Note: Statistics are computed for individuals aged 25 to 60 and for labor spells occurring during the survey period (2002–2020). To determine education levels, I use the most frequently reported level of education. The birth years used to construct the cohorts are aligned with the reported ages. An employment spell is defined based on the reported occupation at the 1-digit ISCO level.

Source: EPS.

TABLE A3. Summary statistics. Selected survey waves.

	2004		2009		2015	
	Males	Females	Males	Females	Males	Females
Overall share	0.469	0.531	0.493	0.507	0.491	0.509
Age						
Mean	41.67	41.97	41.10	42.08	41.56	42.15
Education shares						
Less than high school	0.483	0.480	0.384	0.406	0.290	0.287
High school	0.258	0.270	0.293	0.311	0.318	0.360
More than high school	0.259	0.250	0.322	0.283	0.393	0.353
Employment shares						
Wage workers	0.699	0.817	0.720	0.795	0.686	0.751
Self-employed	0.252	0.154	0.218	0.156	0.219	0.179
Occupation shares						
Managers	0.057	0.045	0.045	0.036	0.037	0.039
Professionals	0.058	0.122	0.064	0.118	0.102	0.125
Technicians	0.080	0.103	0.098	0.097	0.107	0.139
Clerical workers	0.071	0.149	0.076	0.182	0.072	0.164
Service and sales	0.090	0.224	0.097	0.235	0.084	0.230
Agric., forestry, fishing	0.080	0.025	0.074	0.012	0.047	0.010
Craft workers	0.238	0.052	0.226	0.044	0.240	0.027
Operators, assemblers	0.142	0.023	0.160	0.035	0.151	0.031
Elementary occupations	0.183	0.257	0.159	0.240	0.160	0.235
Informality rate						
Criteria 1 - Baseline	0.358	0.333	0.331	0.323	0.289	0.299
Criteria 2	0.377	0.347	0.350	0.338	0.313	0.318
Criteria 3	0.308	0.301	0.291	0.298	0.250	0.277
Hourly earnings						
Mean	93.59	81.56	107.00	87.84	138.24	109.42
Std dev.	122.06	84.13	166.92	112.84	166.44	109.07
Obs (sample)	5,930	6,071	5,253	5,502	4,170	4,802
Obs (weighted)	3,664,323	4,150,937	4,583,231	4,716,619	4,295,062	4,453,186

Note: Statistics are computed for individuals aged 25 to 60. For informality rates, Criteria 1 defines a formal worker as a wage worker with a signed contract or a self-employed worker in professional or technical occupations. Criteria 2 defines a formal worker as a wage worker with a signed contract, while all self-employed workers are considered informal. Criteria 3 defines a worker as formal if they contribute to social security. Hourly earnings are reported in 2018 Chilean Pesos (CLP).

Source: EPS.

TABLE A4. Informality and occupation categories (2002-2020)

	Total informality		Informality (w.w.)		Self-emp shares	
	Men	Women	Men	Women	Men	Women
Overall	0.286	0.289	0.112	0.165	0.221	0.164
Age group						
25-34	0.223	0.229	0.132	0.139	0.139	0.129
35-49	0.280	0.293	0.099	0.173	0.229	0.161
50 or more	0.401	0.373	0.106	0.195	0.330	0.225
Education group						
Less than high school	0.442	0.525	0.177	0.345	0.315	0.267
High school	0.277	0.312	0.089	0.161	0.205	0.185
More than high school	0.173	0.156	0.088	0.096	0.163	0.098
Occupation category						
Wage workers	0.112	0.165				
Self-employed	0.868	0.903				
Occupations						
Managers	0.493	0.477	0.046	0.105	0.325	0.344
Professionals	0.082	0.093	0.088	0.087	0.153	0.061
Technicians	0.077	0.069	0.086	0.072	0.124	0.060
Clerical workers	0.075	0.094	0.049	0.080	0.029	0.015
Service and sales	0.336	0.452	0.098	0.197	0.231	0.290
Agric., forestry, fishing	0.624	0.456	0.210	0.163	0.491	0.317
Craft workers	0.436	0.648	0.127	0.252	0.333	0.500
Operators, assemblers	0.271	0.241	0.110	0.115	0.173	0.131
Elementary occupations	0.329	0.422	0.170	0.333	0.182	0.125
Sample	28,372	20,300	21,995	17,147	30,503	21,468

Note: Statistics are computed for individuals aged 25 to 60, after pooling together the current spells during the survey period (2002-2020). The columns show total informality, informality in wage workers and shares of self-employment by demographic and occupation groups, separately for males and females.

Source: EPS.

A.3. Parents' sample

Table A5 displays the distribution of education and occupation for the parents of individuals in the baseline sample, based on retrospective information. The education variable refers to the highest level of education completed, while the occupation variable refers to the parent's main occupation when the individual was aged 18 or younger.

As shown in Table A5, a very small share of parents attain an education level higher than high school, which is why I group parents into two education categories: less than high school, and high school or more. There is some degree of non-response, which is more pronounced for fathers. However, since fathers are more likely to be employed, the number of valid observations for occupation data is higher for fathers than for mothers.

TABLE A5. Distribution of parental education and occupation

	Fathers	Mothers
Education		
Less than high-school	0.51	0.61
High-school	0.19	0.20
More than high-school	0.09	0.07
No data	0.22	0.11
Employment		
Had job	0.84	0.49
No data	0.12	0.03
Occupation shares		
Managers	0.06	0.06
Professionals	0.04	0.07
Technicians	0.04	0.04
Clerks	0.05	0.08
Service and sales	0.08	0.20
Agricultural, forestry, fishing	0.12	0.02
Craft workers	0.24	0.07
Operators and assemblers	0.14	0.03
Elementary occupations	0.16	0.38
Other/Not known	0.07	0.05
Valid observations		
Education	17,362	19,529
Occupation	18,140	9,952

Note: Statistics computed for the parents of individuals in the baseline sample (individuals aged 25 to 60.)

Source: EPS.

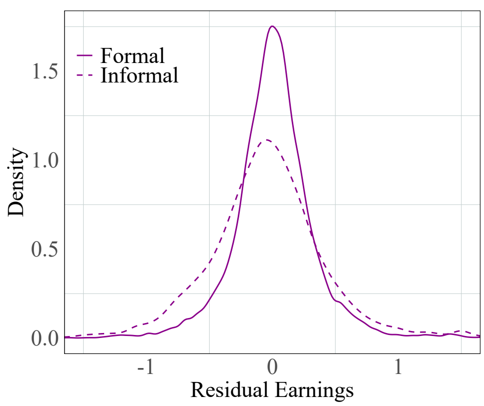
Appendix B. Residual earnings: further results

B.1. Residual earnings distribution by formality status: robustness checks

In this section I compute the densities and moments of log residual earnings for formal and employment spells using alternative measures. In Figure A1, I present the densities and moments calculated from the residuals of log hourly earnings, controlling for age and age squared. In Figure A2, I present the results based on the residuals of log monthly earnings, including additional controls for occupation, education, and informality. The findings confirm that the patterns identified in the main analysis are robust to these alternative specifications.

FIGURE A1. Residuals of log hourly earnings, by formality status

A. Densities of residual log earnings



B. Moments of residual log earnings

	Formal	Informal
Mean	0.012	-0.041
Standard deviation	0.340	0.469
$P_{90} - P_{10}$	0.699	1.074
Kelley skewness	0.008	-0.020
Exc. C-S kurtosis	1.327	1.036
Observations	27,876	10,634

Note: Residuals of log hourly earnings are calculated from a fixed effect regression with age controls.

B.2. Residual earnings by education level

In Figure A3, I present the densities and moments calculated from the residuals of log earnings, separately for low educated and high educated individuals, where the groups are defined based on college attendance.

B.3. Residual earnings by wave, age, occupation and education

Table A6 shows how the measure for income uncertainty varies across other important variables, including time (survey wave), age group, occupation and education (when considering three levels of education).

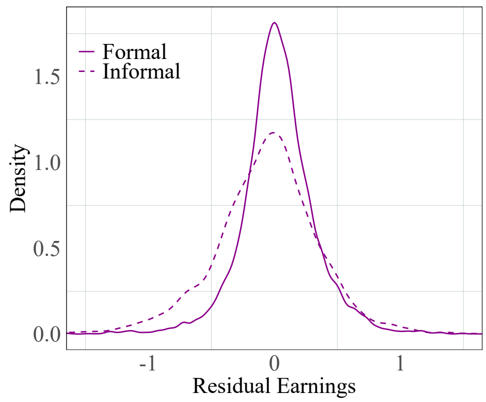
TABLE A6. Standard deviation of residual earnings by groups

Category	Formal	Informal
By Wave		
2002	0.315	0.476
2004	0.281	0.400
2006	0.302	0.413
2009	0.311	0.427
2012	0.331	0.436
2015	0.337	0.432
2020a	0.332	0.481
2020b	0.328	0.464
2020c	0.282	0.559
By Age Group		
25–34	0.300	0.482
35–44	0.309	0.440
45–60	0.328	0.446
By Occupation		
1	0.394	0.431
2	0.363	0.482
3	0.303	0.358
4	0.282	0.391
5	0.294	0.468
6	0.295	0.485
7	0.290	0.457
8	0.305	0.386
9	0.274	0.428
By Education		
< High School	0.282	0.429
High School	0.291	0.422
Some College	0.331	0.490

Source: EPS.

FIGURE A2. Residuals of log earnings, by formality status: additional controls

A. Residuals of residual log earnings



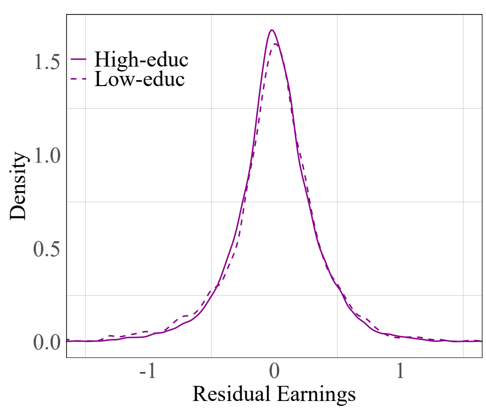
B. Moments of residual log earnings

	Formal	Informal
Mean	0.029	-0.067
Standard deviation	0.317	0.433
$P_{90} - P_{10}$	0.670	1.017
Kelley skewness	0.056	-0.069
Exc. C-S kurtosis	1.234	0.728
Observations	28,233	11,103

Note: Residuals of log earnings are calculated from a fixed effect regression with age, education, occupation and informality controls.

FIGURE A3. Residuals of log earnings, by education level

A. Residuals of residual log earnings



B. Moments of residual log earnings

	Low-Educ	High-Educ
Mean	0.003	-0.003
Standard deviation	0.355	0.389
$P_{90} - P_{10}$	0.776	0.856
Kelley skewness	0.011	-0.059
Exc. C-S kurtosis	1.310	1.488
Observations	28,750	13,111

Note: Residuals of log earnings are calculated from a fixed effect regression with age controls. Low-educated individuals are those with completed high-school or less, while high-educated individuals are those with more than completed high school

Appendix C. Occupational persistence: further results

In this section, I present additional details and robustness checks related to occupation persistence. Tables A7 and A8 display the regression coefficients that are plotted and discussed in Figure 5 of the main text. Next, I assess the impact of including parental education as an additional control and find that this does not significantly alter the results. The coefficients from these regressions are plotted in Figure A4. In turn, Figure A5 shows occupation persistence separately for the high and low-educated. On average, except for white-collar occupations, occupation persistence is somewhat higher for lower-educated individuals.

Finally, I examine the role of both parents' occupations in shaping the occupational outcomes of their children. The coefficients from these regressions are shown in Figure A6. When controlling for both parents' occupations, the results reveal that the father's occupation plays a more significant role in determining the son's occupational choice, while the mother's occupation has a stronger influence on the daughter's occupation.

TABLE A7. Regressions of occupational choice. Males.

	Occ. 1	Occ. 2	Occ. 3	Occ. 4	Occ. 5	Occ. 6	Occ. 7	Occ. 8	Occ. 9
Intercept	0.002	-0.005	0.031***	0.019***	0.080***	0.058***	0.268***	0.164***	0.265***
	[0.005]	[0.007]	[0.008]	[0.006]	[0.008]	[0.006]	[0.012]	[0.010]	[0.010]
Occ. 1	0.060***								
	[0.009]								
Occ. 2		0.190***							
		[0.016]							
Occ. 3			0.025						
			[0.018]						
Occ. 4				0.032**					
				[0.013]					
Occ. 5					0.080***				
					[0.013]				
Occ. 6						0.122***			
						[0.007]			
Occ. 7							0.147***		
							[0.011]		
Occ. 8								0.096***	
								[0.012]	
Occ. 9									0.111***
									[0.011]
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	7,863	7,863	7,863	7,863	7,863	7,863	7,863	7,863	7,863
R2	0.032	0.199	0.078	0.024	0.010	0.077	0.057	0.028	0.094

Note: Each column regresses a dummy variable for the main occupation against a series of controls and dummy variables for the parent's occupation. Controls include the individual's cohort and education level. Occupational codes (1-digit level) are as follows: (1) Managers; (2) Professionals; (3) Technicians and Associated Professionals; (4) Clerical Support Workers; (5) Service and Sales Workers; (6) Agricultural, Forestry and Fishery Workers; (7) Craft and Related Trades Workers; (8) Operators and Assemblers; (9) Elementary Occupations.

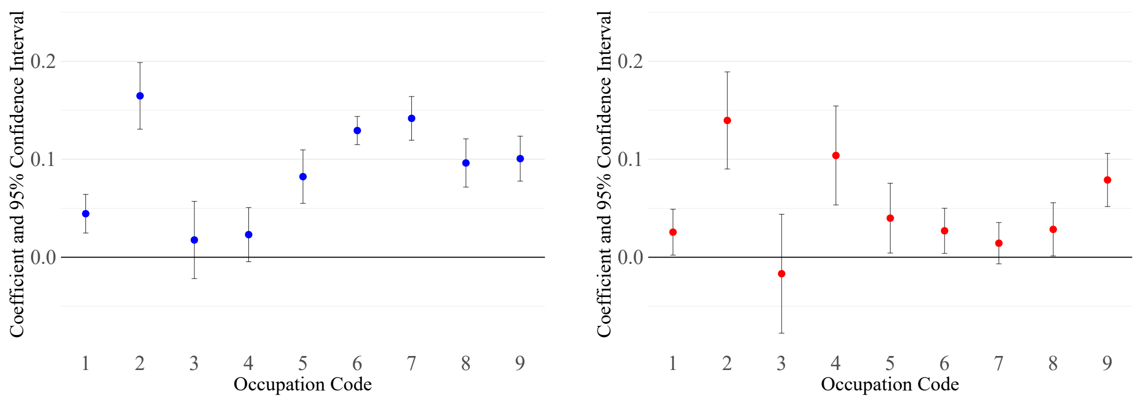
TABLE A8. Regressions of occupational choice. Females.

	Occ. 1	Occ. 2	Occ. 3	Occ. 4	Occ. 5	Occ. 6	Occ. 7	Occ. 8	Occ. 9
Intercept	0.017**	-0.015	0.014	0.073***	0.277***	0.031***	0.065***	0.043***	0.426***
	[0.007]	[0.013]	[0.013]	[0.014]	[0.018]	[0.005]	[0.008]	[0.007]	[0.018]
Occ. 1	0.028**								
	[0.012]								
Occ. 2		0.170***							
		[0.024]							
Occ. 3			-0.017						
			[0.029]						
Occ. 4				0.088***					
				[0.024]					
Occ. 5					0.048***				
					[0.018]				
Occ. 6						0.049***			
						[0.011]			
Occ. 7							0.017		
							[0.011]		
Occ. 8								0.025*	
								[0.013]	
Occ. 9									0.088***
									[0.013]
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	3,677	3,677	3,677	3,677	3,677	3,677	3,677	3,677	3,677
R2	0.009	0.233	0.069	0.055	0.042	0.017	0.012	0.014	0.208

Note: Each column regresses a dummy variable for the main occupation against a series of controls and dummy variables for the parent's occupation. Controls include the individual's cohort and education level. Occupational codes (1-digit level) are as follows: (1) Managers; (2) Professionals; (3) Technicians and Associated Professionals; (4) Clerical Support Workers; (5) Service and Sales Workers; (6) Agricultural, Forestry and Fishery Workers; (7) Craft and Related Trades Workers; (8) Operators and Assemblers; (9) Elementary Occupations.

FIGURE A4. Conditional occupation persistence

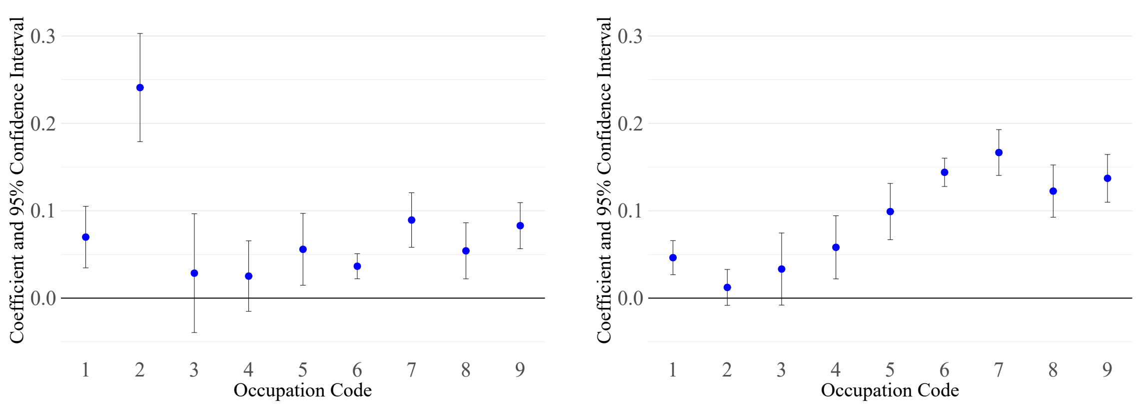
A. Sons and fathers B. Daughters and mothers



Note: The figure plots the persistence coefficients by occupation, after controlling for cohort, own education, and parental education. The coefficients for males are computed with respect to their father's characteristics, while for females, they are computed with respect to their mother's characteristics. Occupational codes (1-digit level) are as follows: (1) Managers, (2) Professionals, (3) Technicians and Associate Professionals, (4) Clerical Support Workers, (5) Service and Sales Workers, (6) Agricultural, Forestry, and Fishery Workers, (7) Craft and Related Trades Workers, (8) Operators and Assemblers, and (9) Elementary Occupations.

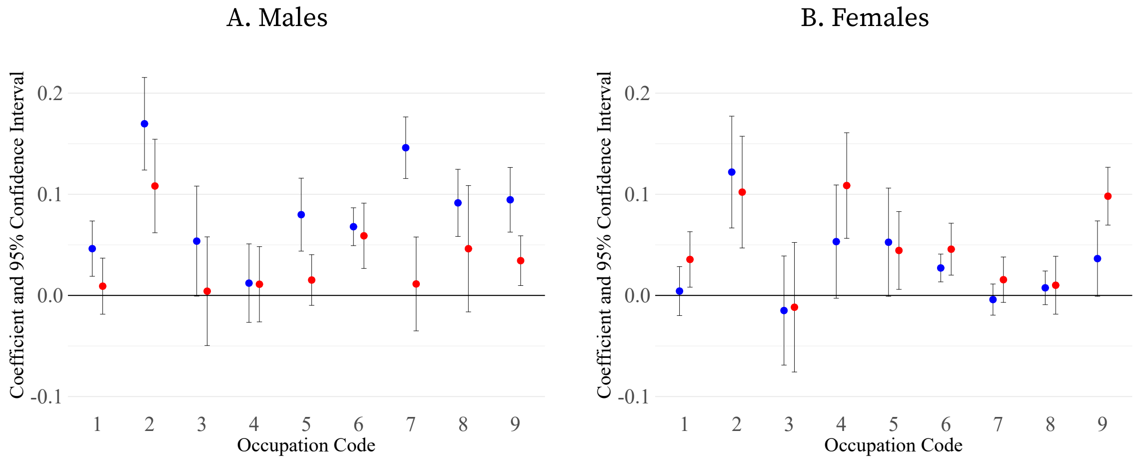
FIGURE A5. Conditional occupation persistence - Sons and fathers

A. High-educated B. Low-educated



Note: The figure plots the persistence coefficients by occupation (1-digit level), after controlling for cohort and education. Occupational codes are as follows: (1) Managers; (2) Professionals; (3) Technicians and Associated Professionals; (4) Clerical Support Workers; (5) Service and Sales Workers; (6) Agricultural, Forestry and Fishery Workers; (7) Craft and Related Trades Workers; (8) Operators and Assemblers; (9) Elementary Occupations.

FIGURE A6. Conditional occupation persistence



Note: The figure plots the persistence coefficients by occupation, after controlling for cohort, own education, and both parents' characteristics. Blue dots represent occupational persistence with respect to the individual's father's occupation, while red dots represent persistence with respect to the mother's occupation. Occupational codes (1-digit level) are as follows: (1) Managers, (2) Professionals, (3) Technicians and Associate Professionals, (4) Clerical Support Workers, (5) Service and Sales Workers, (6) Agricultural, Forestry, and Fishery Workers, (7) Craft and Related Trades Workers, (8) Operators and Assemblers, and (9) Elementary Occupations.

Appendix D. Informality and parental background: further results

D.1. Parental background and informality likelihood

Table A9 shows the average share of time in informal employment by gender, education and occupation. I use this information to input an 'informality likelihood' to each parental background in Section 2.

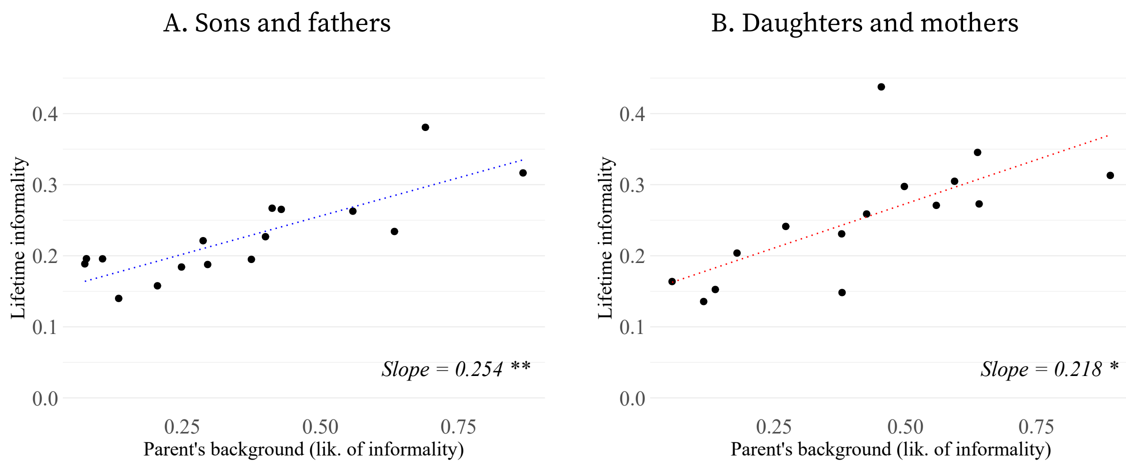
Figure A7 shows the relation between lifetime informality and parental background, when using individuals born before 1970 to compute the parents' "likelihood of informality"

TABLE A9. Average share of time in informal employment by gender, education and occupation

	Males		Females	
	Low Educ	High Educ	Low Educ	High Educ
Managers and self-employed	0.881	0.392	0.912	0.489
Professionals	0.127	0.109	0.044	0.103
Technicians	0.119	0.083	0.207	0.091
Clerical Workers	0.154	0.069	0.143	0.091
Service and Sales	0.512	0.237	0.594	0.353
Agricultural, Forestry, Fishing	0.746	0.536	0.477	0.604
Craft Workers	0.470	0.292	0.742	0.589
Operators and Assemblers	0.266	0.249	0.250	0.284
Elementary Occupations	0.387	0.241	0.512	0.433

Note: Low education corresponds to less than high school, while high education corresponds to high school and more.

FIGURE A7. Share of time in informality by family background



Note: Lifetime informality is the share of time in informal employment (vertical axis). A parent's background is defined as the combination of education and occupation. Each parental background is assigned an informality likelihood (horizontal axis).

D.2. Informality and parental background: further results

In this section, I present the full regression results examining the relationship between lifetime informality and parental background when accounting for whether individuals follow their parents' occupation and education. Table A10 reports the results for males, while Table A11 provides the corresponding results for females. The main coefficients of this Tables are reported and discussed in Table 6 of the main text.

Finally, in Table A12, I report the results of regressions examining lifetime informality against parental background, including education and occupation controls. The results show similar patterns to those when controlling only for "followers"; that is, accounting for educational and occupational choices explains most of the association between informality and parental background. However, a positive and significant association still remains even after accounting for both education and occupation, suggesting that there may be broader mechanisms contributing to the intergenerational persistence of informality.

TABLE A10. Regression results: lifetime informality and parental background. Males (dep. var.: share of time in informal employment)

	(1)	(2)	(3)	(4)
Intercept	0.117***	0.114***	0.137***	0.118***
	[0.013]	[0.016]	[0.014]	[0.017]
Parent's background	0.258***	0.138***	0.145***	0.079**
	[0.023]	[0.032]	[0.025]	[0.034]
Education follower		0.043**		0.082***
		[0.021]		[0.024]
Occupation follower			-0.156***	-0.107***
			[0.026]	[0.041]
Parent's background * Education follower		0.209***		0.104**
		[0.045]		[0.050]
Parent's background * Occupation follower			0.680***	0.555***
			[0.058]	[0.096]
Education follower * Occupation follower				-0.101*
				[0.053]
Parent's background * Educ. follower * Occ. follower				0.170
				[0.120]
Cohort controls	✓	✓	✓	✓
Observations	6,659	6,652	6,591	6,584
R^2	0.061	0.099	0.100	0.122

Note: A follower is an individual who coincides with their father in either education level or main occupation.

TABLE A11. Regression results: lifetime informality and parental background. Female (dep. var.: share of time in informal employment)

	(1)	(2)	(3)	(4)
Intercept	0.133***	0.152***	0.157***	0.163***
	[0.020]	[0.024]	[0.022]	[0.027]
Parent's background	0.242***	0.109**	0.129***	0.026
	[0.035]	[0.043]	[0.039]	[0.047]
Education follower		-0.076**		-0.041
		[0.035]		[0.041]
Occupation follower			-0.188***	-0.165***
			[0.041]	[0.058]
Parent's background * Education follower		0.523***		0.496***
		[0.071]		[0.081]
Parent's background * Occupation follower			0.705***	0.747***
			[0.085]	[0.118]
Education follower * Occupation follower				-0.005
				[0.082]
Parent's background * Educ. follower * Occ. follower				-0.349**
				[0.169]
Cohort controls	✓	✓	✓	✓
Observations	3,399	3,392	3,363	3,356
R^2	0.046	0.096	0.084	0.126

Note: A follower is an individual who coincides with their father in either education level or main occupation.

TABLE A12. Regression results: lifetime informality and parental background (dependent variable: share of time in informal employment)

a. Males

	(1)	(2)	(3)	(4)
Intercept	0.117***	0.292***	0.377***	0.490***
	[0.013]	[0.016]	[0.027]	[0.030]
Parent's background	0.258***	0.144***	0.114***	0.079***
	[0.023]	[0.024]	[0.023]	[0.023]
Education controls		✓		✓
Occupation controls			✓	✓
Cohort controls	✓	✓	✓	✓
Observations	6,659	6,652	6,591	6,584
R^2	0.061	0.104	0.163	0.177

b. Females

	(1)	(2)	(3)	(4)
Intercept	0.133***	0.404***	0.601***	0.717***
	[0.020]	[0.024]	[0.043]	[0.045]
Parent's background	0.242***	0.098***	0.086***	0.065**
	[0.035]	[0.035]	[0.033]	[0.033]
Education controls		✓		✓
Occupation controls			✓	✓
Cohort controls	✓	✓	✓	✓
Observations	3,399	3,392	3,363	3,356
R^2	0.046	0.130	0.228	0.240

Appendix E. Model - Comparative statics

E.1. Determinants of occupation choice

From Equation 6 in the main text, the relative probability of selecting into occupation j' compared to another occupation l' can be written as:

$$\frac{\mu(j', j, z')}{\mu(l', j, z')} = \frac{\exp\left(\frac{V'(j',z')+\phi(j,j',z')}{\beta_j}\right)}{\exp\left(\frac{V'(l',z')+\phi(j,l',z')}{\beta_j}\right)}$$

Taking the logarithm of this ratio yields:

$$\ln\left(\frac{\mu(j',j,z')}{\mu(l',j,z')}\right)=\frac{V'(j',z')+\phi(j,j',z')-V'(l',z')-\phi(j,l',z')}{\beta_j} \tag{A1}$$

where the expected utility $V(j',z')$ is given by:

$$V(j',z')=p(j',z')\mathbb{E}\left\{u\left[w(j',z',1)\varepsilon(j',1)\right]\right\}+(1-p(j',z'))\mathbb{E}\left\{u\left[w(j',z',0)\varepsilon(j',0)\right]\right\}$$

I next perform two comparative statics to show how the probability of selecting into an occupation varies with the probability of formality and with human capital.

Probability of formality

From Equation A1 the share of individuals choosing occupation j' relative to an alternative occupation l' is increasing with the probability of formality in occupation j', $p(j',z')$, if:

$$\frac{\partial\ln\left(\frac{\mu(j',j,z')}{\mu(l',j,z')}\right)}{\partial p(j',z')}=\frac{1}{\beta_j}\left[\frac{\partial V(j',z')}{\partial p(j',z')}\right]>0$$

meaning there are positive returns to formality in occupation j'.

Taking the derivative of $V(j',z')$ with respect to $p(j',z')$, we get the following expression:

$$\frac{\partial V(j',z')}{\partial p(j',z')}=\mathbb{E}\left\{u\left[w(j',z',1)\varepsilon(j',1)\right]\right\}-\mathbb{E}\left\{u\left[w(j',z',0)\varepsilon(j',0)\right]\right\}$$

which is positive if the expected utility of a formal job is larger than the expected utility of an informal job. Under the assumption of a utility function such that $u'''>0$, it can be shown that the expected utility of labor income, $\mathbb{E}\left\{u\left[w(j',z',f')\varepsilon(j',f')\right]\right\}$ is increasing in $w(j',z',f')$ and decreasing in the variance of $\varepsilon(j',f')$.

To see this, assume a CRRA utility function with parameter γ. Using the fact that $\varepsilon(j',f')$ is log-normally distributed with $\mu=-\frac{\sigma^2}{2}$[13], the derivative of $V(j',z')$ with respect to $p(j',z')$ becomes:

$$\frac{\partial V(j',z')}{\partial p(j',z')}=\frac{w(j',z',1)^{1-\gamma}}{1-\gamma}\exp\left(-\frac{(1-\gamma)\gamma}{2}\sigma^2_{j'1}\right)-\frac{w(j',z',0)^{1-\gamma}}{1-\gamma}\exp\left(-\frac{(1-\gamma)\gamma}{2}\sigma^2_{j'0}\right)$$

[13]This assumption is needed for the shock to be mean-preserving. Note this implies:

$$\mathbb{E}\left[\varepsilon(j',f')^{1-\gamma}\right]=\exp\left(-\frac{\gamma(1-\gamma)\sigma^2_{j'f'}}{2}\right)$$

For $\frac{\partial V(j',z')}{\partial p(j',z')} > 0$, the condition becomes:

$$w(j',z',1)^{1-\gamma} \exp\left(-\frac{(1-\gamma)\gamma}{2}\sigma^2_{j'1}\right) > w(j',z',0)^{1-\gamma} \exp\left(-\frac{(1-\gamma)\gamma}{2}\sigma^2_{j'0}\right)$$

Applying logarithm and rearranging:

$$\log\left(\frac{w(j',z',1)}{w(j',z',0)}\right) > \frac{(1-\gamma)\gamma}{2}(\sigma^2_{j'0} - \sigma^2_{j'1}) \tag{A2}$$

Finally, in the case in which $\gamma = 2$ (assumption I use in the quantitative exercise):

$$\log\left(\frac{w(j',z',1)}{w(j',z',0)}\right) + \sigma^2_{j'0} - \sigma^2_{j'1} > 0 \tag{A3}$$

This condition shows that the expected value of the formal job increases with a larger wage gap between formal and informal jobs within occupation j'. Additionally, the returns of a formal job are increasing in the variance of the informal job shock and decreasing in the variance of the formal job income shock. Holding everything else constant, the share of individuals selecting into occupation j' will increase with $p(j',z')$ if condition A3 is satisfied.

Human capital

From Equation A1, the share of individuals choosing occupation j' relative to an alternative occupation l' is increasing in human capital z' if:

$$\frac{\partial \ln\left(\frac{\mu(j',j,z')}{\mu(l',j,z')}\right)}{\partial z'} = \frac{1}{\beta_j}\left[\frac{\partial V(j',z')}{\partial z'} + \frac{\partial \phi(j',j,z')}{\partial z'} - \frac{\partial V(l',z')}{\partial z'} - \frac{\partial \phi(l',j',z')}{\partial z'}\right] > 0$$

This means the expected gains of human capital, which comprise both the pecuniary and non-pecuniary components of utility, are larger in occupation j' compared to the alternative l'.

The effect of the non-pecuniary component on the relative occupation share is straightforward, and depends on how the preferences for occupations j' and l' vary across human capital levels. On the other hand, in order to understand the channels that affect the pecuniary component of utility, I compute the derivative of $V(j',z')$ with respect to z' as:

$$\frac{\partial V(j',z')}{\partial z'} = \frac{\partial p(j',z')}{\partial z'}\left[\mathbb{E}\left\{u\left[w(j',z',0)\varepsilon(j',0)\right]\right\} - \mathbb{E}\left\{u\left[w(j',z',1)\varepsilon(j',1)\right]\right\}\right]$$

$$+ p(j', z')\mathbb{E}\left\{u'\left[w(j', z', 0)\varepsilon(j', 0)\right]\frac{\partial w(j', z', 0)}{\partial z'}\varepsilon(j', 0)\right\}$$

$$+ (1 - p(j', z'))\mathbb{E}\left\{u'\left[w(j', z', 1)\varepsilon(j', 1)\right]\frac{\partial w(j', z', 1)}{\partial z'}\varepsilon(j', 1)\right\}$$

The expected pecuniary utility $V(j', z')$ increases with human capital z' if there are expected gains associated with higher human capital. These gains result from a combination of the effect of human capital on income, its impact on the probability of formality, and the interaction between human capital, the probability of formality, income, and the variance of the income shocks.

E.2. Determinants of parental investment share

In this section, I show that the share of parents who invest, $s(j, z)$ decreases with the variance of the occupation-specific income shock, σ_j^2.

In order to see this, recall Equation 9, which explicitly defines the investment threshold $x^*(j, z)$ for each parental type:

$$\begin{aligned} x^*(j,z) &= \mathbb{E}\left[u\left(y(j,z)\right)\right] - \mathbb{E}\left[u\left(y(j,z) - \bar{x}\right)\right] \\ &\quad - \alpha\beta_j\left(\log\left(\sum_{j'} e^{\frac{V'(j',z_h)+\phi(j',j,z_h)}{\beta_j}}\right) - \log\left(\sum_{j'} e^{\frac{V'(j',z_l)+\phi(j',j,z_l)}{\beta_j}}\right)\right) \end{aligned} \tag{A4}$$

The first part of the expression shows that the threshold increases with the gap between the parent's utility of not investing and investing: the larger the difference between these two utilities, the higher the threshold, and the smaller the share of parents who invest. The second part of the expression shows that the threshold decreases with the human capital gains in the next generation: the higher the expected gains, the lower the threshold, and the larger the share of parents who invest.

The determinants of the human capital gains for the next generation were discussed in the previous section of this Appendix. Ultimately, they depend on the effects of human capital on income, on the probability of formality, and on the interaction between the probability of informality, income, and the variance of the income shock. Therefore, in the rest of this section, I focus on commenting on the determinants of the first part of Equation A4 related to parental utility, and in particular, on the effect of the variance of the income shock.

To that end, I compute the expectation of the parental period utility when investing and not investing. Consider a CRRA utility function with parameter γ and a mean-preserving log-normal income shock $\varepsilon(j)$ with variance σ_j^2. I avoid indexing by parental type $\{j, z\}$

and by occupation j to ease the notation. The expected utility of investing is defined as:

$$\mathbb{E}\left\{u(y-\bar{x})\right\} = \mathbb{E}\left\{\frac{(w\varepsilon-\bar{x})^{1-\gamma}}{1-\gamma}\right\}$$

To evaluate this expression, I perform a Taylor expansion around $\varepsilon = \mathbb{E}(\varepsilon) = 1$:

$$\mathbb{E}\left\{u(y-\bar{x})\right\} \approx$$
$$\frac{1}{1-\gamma}\mathbb{E}\left[(w-\bar{x})^{1-\gamma} + (1-\gamma)(w-\bar{x})^{-\gamma}w(\varepsilon-1) + \frac{(1-\gamma)(-\gamma)w^2(w-\bar{x})^{-\gamma-1}}{2}(\varepsilon-1)^2\right]$$

Next, distributing the expectation and using the fact that for a log-normal distribution $\mathbb{E}(\varepsilon^k) = e^{k\mu+\frac{k^2\sigma^2}{2}}$:

$$\text{(A5)} \qquad \mathbb{E}\left\{u(y-\bar{x})\right\} \approx \frac{(w-\bar{x})^{1-\gamma}}{1-\gamma} - \frac{\gamma w^2(w-\bar{x})^{-\gamma-1}}{2}\left(e^{\sigma_j^2}-1\right)$$

In the case of no investment, $\bar{x} = 0$, and the expression becomes:[14]

$$\text{(A6)} \qquad \mathbb{E}\left\{u(y)\right\} \approx \frac{w^{1-\gamma}}{1-\gamma} - \frac{\gamma w^2 w^{-\gamma-1}}{2}\left(e^{\sigma_j^2}-1\right)$$

Finally, substitute expressions A5 and A6 into the threshold definition A4 and compute the derivative with respect to the variance of the shock σ_j^2:

$$\frac{\partial x^*}{\partial \sigma_j^2} = -\frac{\gamma w^2 w^{-\gamma-1}}{2}e^{\sigma_j^2} + \frac{\gamma w^2(w-\bar{x})^{-\gamma-1}}{2}e^{\sigma_j^2} = \frac{\gamma w^2}{2}\left[(w-\bar{x})^{-\gamma-1} - w^{-\gamma-1}\right]e^{\sigma_j^2}$$

Since $\gamma \geq 0$, this expresion is positive for all $\bar{x} > 0$. This means that the variance of the income shock increases the investment threshold, which in turn, implies a lower share of parents who invest.

[14]In the case where $\gamma = 2$ (assumption used in the quantitative exercise), Expression A5 is

$$\mathbb{E}\left\{u(y-\bar{x})\right\} = \frac{-(w-\bar{x})^2 - w^2\left(\exp(\sigma_j^2)-1\right)}{(w-\bar{x})^3},$$

and Expression A6 is simply

$$\mathbb{E}\left\{u(y)\right\} = \frac{-\sigma_j^2}{w}.$$

Appendix F. Quantitative Exercise - Estimated parameters

F.1. Parameters estimated outside the model

Table A13 reports the values of the parameters estimated outside the model. These correspond to the probabilities of informality, by education level and occupation, and to some parameters of the wage process (autorregressive coefficient and variance of residuals) by occupation and formality status.

TABLE A13. Parameters estimated outside the model

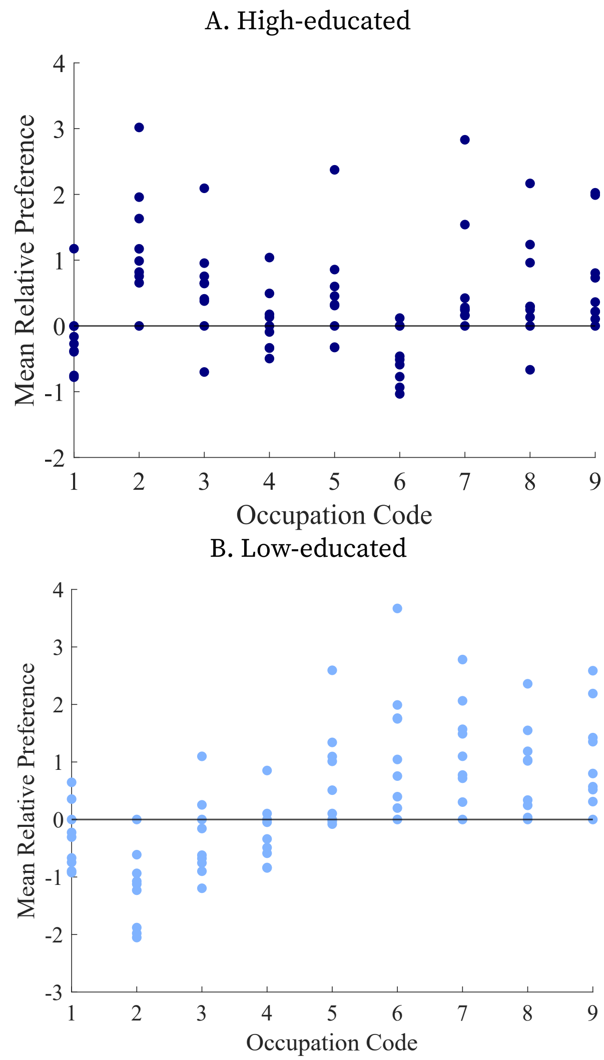
Probability of formality	Autoregressive coefficient		Variance of residuals	
	$\rho(j)$	$\rho(j', f')$	$\sigma_v^2(j)$	$\sigma_v^2(j', f')$
$p(1,1) = 0.213$	$\rho(1) = 0.500$	$\rho(1,1) = 0.668$	$\sigma_v^2(1) = 0.350$	$\sigma_v^2(1,1) = 0.297$
$p(2,1) = 0.899$	$\rho(2) = 0.573$	$\rho(2,1) = 0.564$	$\sigma_v^2(2) = 0.323$	$\sigma_v^2(2,1) = 0.304$
$p(3,1) = 0.914$	$\rho(3) = 0.557$	$\rho(3,1) = 0.552$	$\sigma_v^2(3) = 0.248$	$\sigma_v^2(3,1) = 0.235$
$p(4,1) = 0.916$	$\rho(4) = 0.509$	$\rho(4,1) = 0.545$	$\sigma_v^2(4) = 0.205$	$\sigma_v^2(4,1) = 0.193$
$p(5,1) = 0.637$	$\rho(5) = 0.418$	$\rho(5,1) = 0.435$	$\sigma_v^2(5) = 0.220$	$\sigma_v^2(5,1) = 0.166$
$p(6,1) = 0.299$	$\rho(6) = 0.462$	$\rho(6,1) = 0.217$	$\sigma_v^2(6) = 0.330$	$\sigma_v^2(6,1) = 0.146$
$p(7,1) = 0.610$	$\rho(7) = 0.459$	$\rho(7,1) = 0.456$	$\sigma_v^2(7) = 0.223$	$\sigma_v^2(7,1) = 0.176$
$p(8,1) = 0.756$	$\rho(8) = 0.440$	$\rho(8,1) = 0.468$	$\sigma_v^2(8) = 0.198$	$\sigma_v^2(8,1) = 0.177$
$p(9,1) = 0.663$	$\rho(9) = 0.353$	$\rho(9,1) = 0.276$	$\sigma_v^2(9) = 0.188$	$\sigma_v^2(9,1) = 0.113$
$p(1,0) = 0.754$		$\rho(1,0) = 0.378$		$\sigma_v^2(1,0) = 0.383$
$p(2,0) = 0.891$		$\rho(2,0) = 0.564$		$\sigma_v^2(2,0) = 0.304$
$p(3,0) = 0.916$		$\rho(3,0) = 0.490$		$\sigma_v^2(3,0) = 0.333$
$p(4,0) = 0.929$		$\rho(4,0) = 0.280$		$\sigma_v^2(4,0) = 0.287$
$p(5,0) = 0.757$		$\rho(5,0) = 0.402$		$\sigma_v^2(5,0) = 0.348$
$p(6,0) = 0.546$		$\rho(6,0) = 0.481$		$\sigma_v^2(6,0) = 0.496$
$p(7,0) = 0.724$		$\rho(7,0) = 0.408$		$\sigma_v^2(7,0) = 0.282$
$p(8,0) = 0.710$		$\rho(8,0) = 0.474$		$\sigma_v^2(8,0) = 0.268$
$p(9,0) = 0.726$		$\rho(9,0) = 0.322$		$\sigma_v^2(9,0) = 0.364$

Note: The table shows the estimated parameters for the different combinations of occupation and education (for p), and occupation and formality status (for ρ and σ_v^2). Occupations j' are indexed from 1 to 9, corresponding to the 1-digit level ISCO. Human capital levels z' are two: low-educated ($z' = 0$) and high-educated ($z' = 1$). Formality status f' are also two: formal ($f' = 1$) and informal ($f' = 0$).

F.2. Preference parameters

Figure A8 shows all the values for the estimated occupation preferences relative to the father's occupation. The average relative preferences by occupation and education level are shown in Figure 9 of the main text.

FIGURE A8. Relative preferences for occupations



Note: The figure plots the relative preferences for the father's occupation. Occupational codes (1-digit level) are as follows: (1) Managers, (2) Professionals, (3) Technicians and Associate Professionals, (4) Clerical Support Workers, (5) Service and Sales Workers, (6) Agricultural, Forestry, and Fishery Workers, (7) Craft and Related Trades Workers, (8) Operators and Assemblers, and (9) Elementary Occupations.