Human Capital and Labor Informality in developing countries: A structural dynamic approach

(Draft Paper)

Italo López García (UCL)

April 16, 2013

Abstract

This paper studies the extent to which low human capital development is a determinant of labor informality and self-employment using longitudinal data from Chile. To do so, I estimate a structural dynamic life-cycle model explicitly linking schooling attendance, wages and labor choices. In the model, informality and self-employment are choices driven by individual's comparative advantage, which is consistent with multiple evidence from developing countries that workers self-select into informality, and evidence of a strong positive correlation between schooling and formal labor participation. Nevertheless, the existent literature on labor informality and human capital does not consider potential self-selection into schooling based on expected gains to formality, and does not distinguish schooling from other unobserved skills, like entrepreneurship ability. I study how schooling is a determinant of labor informality by incorporating these two aspects in the context of a human capital investment model in the fashion of Keane and Wolpin (1997) and Attanasio, Meghir and Santiago (2010). The model disentangles human capital into schooling, unobserved skills and sector experience, and allows individuals to self-select into schooling based on expected monetary and non-monetary sector gains. Finally, I use structural estimates to assess the potential benefits of educational policies like college subsidies on the reduction of informality, which in many developing countries account for the biggest share of labor. Preliminary calibration exercises shows that a human capital investment model can accurately fit the data. Estimation results are work in progress.

1 Introduction

The phenomenon of informal labor markets in developing economies has been one of the main concerns for economists and policy-makers for the past decade. According to Gasparini, L. and L. Tornarolli (2007), nearly 40% of the labor force is informal in the case of Latin America, going from lower bounds of 25% in the case of Chile or Uruguay to upper bounds of 60% in the cases of Peru or Colombia. According to Levy (2008) and Meghir, Narita and Robin (2012), the informal sector comprises firms evading taxes and costly labor regulations who hire workers without a contract, are not covered by the social security, have precarious jobs and do not have access to labor benefits. The informal sector often comprises small-scale, self-financed and unskilled labor intensive economic activities, and informal workers tend to be younger, less skilled, and earn less than their counterparts in the formal sector (Thomas (1992); Maloney (1999)). Nonetheless, the study of informality in developing countries is rare. In particular, little has been said about the role of human capital development driving informal labor participation. In the words of the ILO(2004)[1], "informality is itself a trap for unskilled workers, perpetuating a vicious circle of limited human capital and low pay, resulting in low productivity and high income inequality". An interesting question is to study the extent to which low human capital development is an impediment to reduce labor informality and whether educational policies might have an impact on informal participation.

In this paper, I attempt to answer this question by employing a human capital investment model in a dynamic context as a potential vehicle for explaining observed patterns of schooling, labor informality and self-employment, using longitudinal data from Chile. Up to the extent the model is able to replicate life-cycle patterns of schooling, informality and wages, it provides a good understanding of how human capital development determines the informality choice and allows the assessment of the effect of schooling policies like college subsidies on the size of the informal sector. In the model, I distinguish informality from self-employment because there is not full match between them. Furthermore, there is a wide rage of heterogeneity in skills within the formal and the informal. The formal self-employed show significantly higher education and wage profiles than the formal employees, which in turn are more skilled and earn more than the informal. These differences might be due to the interaction between schooling and entrepreneurship abilities. Arbex, Galvao and Gomes (2010) find in Brazilian data that there is a wage premium in the informal sector for high skilled workers, Thus, I open the labor market into four sectors: formal employees, formal self-employed, informal employees and informal self-employed, allowing human capital factors to be linked to them in different ways.

Why do we care so much about informality? Some papers study the relevance of informality for the economy and welfare. For example some authors see informality as an engine of economic growth, allowing firms to operate in less regulated labor markets with lower wages and regulatory costs. Others consider that informality

[1]ILO, 2004. "Chile: Trabajo decente y Calidad de Vida Familiar, 1990-2000"

has negative effects on productivity. Levy (2008) argues that in order to avoid costly labor regulations and contributions, firms operate in suboptimal combinations of capital and labor given the available technology, losing the possibility to achieve economies of scale, access to credits, or risk insurance. Since labor and capital are miss-allocated, similar workers are less productive in the informal sector and are paid lower wages as firms pay a marginal productivity of labor equated to lower expected labor costs. They also lose training opportunities and access to credit, limiting their optimal consumption/saving decision. Informality might also have negative effects in welfare. Maloney et al. (2007) note that salaried and non-salaried informal workers and the unregistered self-employed are not socially protected against health or disability risks, employment shocks, or the risk of having a small pension after retirement.

In this paper, I do not discuss whether informality is good or bad for the economy or welfare. Instead, I attempt to undercover the social factors driving informality, in particular low Human Capital development. To the extent that higher schooling attendance reduce informal participation, there is an important scope for the role of schooling policies shaping more efficient labor markets. Furthermore, I place my work on the competitive markets view to informality (Heckman and Pagés (2004), Magnac (1991), Maloney (1999),Levy (2008)). Labor markets are competitive and the worker chooses to be formal or informal according to their comparative advantage. Given the dynamic nature of choices, forward looking workers self-select into schooling and into informality and self-employment based on expected rewards, which are determined by formal schooling and experience, but also by sector-specific unobserved skills. This is consistent with wage premium in the informal sector in Brazil (Meghir, Narita and Robin (2012);Arbex, Galvao and Gomes (2010)). The relevant outcomes of the estimation process are the sector-specific returns to human capital and preference parameters regarding schooling and sector choices, which along with data on policy variables like tuition fees, can be used for policy simulations.

I contribute to the literature on human capital and labor informality in several aspects. First, I estimate a dynamic extended Roy Model (Roy, 1951) treating the schooling decision as endogenous (Rosen and Willis (1979)) in a dynamic context. Previous studies have estimated the impact of schooling on informality in Latin America. Arbex, Galvao and Gomes (2010) estimate heterogeneous returns to education in the informal sector predicting that schooling is endogenous and its effect on earnings is heterogeneous, reflecting unobservable ability components. Nonetheless, their approach does not consider dynamic effects and they do not analyze separately informality from self-employment. Contreras, de Mello and Puentes (2008) use cross-section analysis in biprobit models to study self-selection into formality, but they take schooling as exogenous. Second, I explicitly estimate returns to ability varying across the four sectors modeling unobserved types in the fashion of Heckman and Singer (1984). If unobserved ability is not accounted for, the causal impact of schooling on informality and will be overstated. Obtaining unbiased sector-returns to schooling is determinant to assess the effect of schooling policies on the size of the informal sector. And finally, differentiating informality from self-employment can explain an u-shaped life-cycle profile for informality rates despite the fact that informality is decreasing for both the self-employed and salaried employees. The reason is a compositional effect: self-employment rates are sharply increasing over the life-cycle.

2 A theoretical approach to informality

The traditional view to labor informality was proposed by Harris and Todaro (1970) and the ILO (1972)[2]. In their view, informality is a consequence of barriers to entry to the formal sector caused by binding minimum wages, which segment the market and create the so called "good jobs". Informal jobs belong to a "residual" sector, composed by workers who are "queuing" for a formal job. This implies that identical workers have a larger utility in the formal sector, where they are paid larger-than-equilibrium wages, and that there is no mobility from the formal to the informal sector. Most of the evidence has rejected this view. Heckman and Pagés (2004) find that in most Latin American countries with high levels of informality, minimum wages are not binding, Chile included. Magnac (1991) finds no evidence of market segmentation using data from Colombia. Levy (2008) argues that if the labor market were distorted in favor of a formal sector, simple policies based on transfers to the informal sector would be enough to offset them. But while transfers to the poor have increased in the last decade through social protection programs, informality rates have not decreased.

A second view, proposed by Maloney (1999), Maloney et al. (2007), and Heckman and Pagés (2004) see labor markets fully or weakly competitive with no barriers to mobility. Informality arises as a result of small and middle size firms and the self-employed escaping from rigid labor regulations or burdensome tax system. Firms and workers are profit and utility maximizing and as a consequence, the formal and informal sector are equally desirable for a worker and a firm at the margin. This view is in line with the Comparative Advantage approach (Heckman and Sedlacek (1985)) and multiple evidence of heterogeneous returns to schooling (Carneiro and Heckman (2003)). The formal and the informal sector are symmetric and competitive, with different production functions, and workers value advantages and disadvantages of each sector and select the one with the highest utility given their tastes and skills. Identical workers are equally productive in both sectors. All these papers have shown strong mobility across sectors in both directions, especially among the poor.

In a refinement of this view, Levy (2008) argues that the informal sector is less productive than the formal sector because there is a misallocation of capital and labor across sectors produced by badly designed social policies, such as social protection programs for the poor, which induce a higher than optimal rate of firms and workers operating in the informal sector. Firms optimize given the constraints imposed by labor regulations.

[2]International Labor Office. 1972. Incomes, Employment and Equality in Kenya. Geneva.

Thus, in order to operate formally, firms must pay higher labor costs, be more productive and have better technology. As a consequence, workers are not equally productive in both sectors and net wages are larger in the formal one. Complementarity of skills and technology implies that in equilibrium firms find it difficult to hire unskilled workers formally. Therefore, utility-maximizing workers choose in which sector to work based on pecuniary and non-pecuniary sector benefits, constrained by their skills. As a consequence, informality is privately but not socially efficient, and low Human Capital development prevents reduction of informality rates. Meghir, Narita and Robin (2012) provide evidence to this view from Brazil. They show that the average formal wage is larger than the informal one because formal firms tend to operate at higher levels of productivity than informal firms, but informal workers are paid more than formal workers for firms operating at the same level of productivity.

3 The data

3.1 Institutional framework

I consider the informal sector as the one composed by firms not registered with the authority, not paying taxes, and not paying neither social security contributions nor labor laws; and by all full-time (more than 20 hours a week) salaried and self-employed workers reporting not to be covered by social contributions.

There is multiple evidence of little labor market segmentation in Chile. Contreras, de Mello and Puentes (2008) argue that the Chile´s tax system is not particularly burdensome, and with regard to labor regulations, the Chilean dictatorship during the 80´s strongly deregulated labor markets decreasing severance pay, dismissal costs and minimum wages, and prohibiting unions activity. A reform in 1980 intended to link contributions with benefits transformed the pay-as-you-go social security system into a full capitalization system, including the pension retirements and health insurance, making Chile the least labor-market regulated Latin American country. Heckman and Pagés (2004) argue that social protection programs for the poor are still very small to incentive informality in Chile, compared to Mexico, Brazil or Argentina.

Regarding human capital factors to informality, Contreras, de Mello and Puentes (2008)show that the probability of working in the formal sector increase with education, job tenure and experience, and that labor informality appears predominantly from self-selection, as workers are estimated to have a high probability of obtaining a formal job if they seek one.

3.2 Descriptives

The "Encuesta de Protección Social" (Social Protection Survey) is a longitudinal survey containing four waves: 2002, 2004, 2006 and 2009. It covers a nationally representative sample of 14.045 individuals who are followed across the three waves with very low attrition rates. In the first wave, individuals are also requested to report family background, all their educational history, and all their labor activities from 1980 onwards, including the type of job performed, the hours of work, the presence of a labor contract, social security coverage, and the labor status (active, inactive, full or part-time, if worked in a firm or as self-employed). Information on wages is available only from 2002 onwards. Since female labor participation is still low (44%), the model will be estimated for males to avoid modeling fertility. In total, the panel for males consists of 6,932 individuals, with 222.403 individual-year observations.

As a policy variable, I use tuition data fees both at the secondary education and college, from the CASEN survey, a nationally representative survey reporting how much the household spent on tuition fees each year. This data is not panel but detrended costs by municipality and year are obtained to simulate schooling choices.

Table 1 shows labor informality rates by age group and gender for the period 1980-2006. Overall labor informality is quite stable over years, but is more prevalent among the youth and the elderly. I take out individuals above 65 years old as the legal age retirement in Chile is 65 for men and 60 for women. Informality rates are higher for males than for females for all age groups. Over time, it has slightly decreased among the youth and slightly increased among the elderly.

	Males					Females				
Age	1980-84	1985-89	1990-94	1995-99	2000-06	1980-84	1985-89	1990-94	1995-99	2000-06
<20	45.3	44.2	37.3	35.0	28.6	41.4	41.7	33.0	31.6	30.4
20-24	33.6	31.8	25.0	22.9	20.4	28.7	26.9	22.2	19.4	16.8
25-29	25.4	27.8	24.1	22.0	20.3	25.9	21.5	20.3	17.6	17.8
30-34	24.6	25.8	26.0	25.9	22.9	23.6	23.8	21.8	22.1	18.5
35-39	23.9	26.6	26.6	27.8	24.9	24.7	24.0	25.7	24.6	25.2
40-44	25.9	27.2	26.7	28.4	27.0	28.8	25.1	24.0	25.8	30.0
45-49	23.1	28.6	28.2	29.9	29.0	21.0	28.6	24.4	24.7	28.6
50-54	24.6	26.0	29.5	32.1	31.7	25.5	22.4	28.1	24.4	27.6
55-59	25.5	28.6	28.1	33.7	36.2	26.1	26.9	23.9	31.3	27.9
60-64	33.8	30.5	33.5	34.7	39.1	36.4	36.4	34.9	29.4	46.0
Total	27.9	28.7	27.1	28.3	28.2	27.0	25.5	24.2	24.0	25.9

Source: Author's estimates from the Social Protection Survey, Chile.

Table 1: Informality Rates by gender, age and year.

Figure 1 shows Informality rates for males by age group over the whole period of study. Informality rates are higher among the youth and the elderly, even after controlling for cohort fixed effects (in Appendix A).

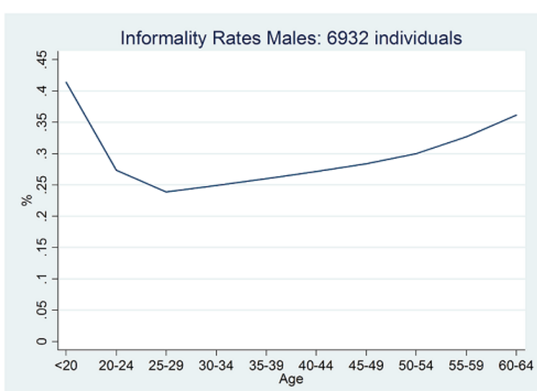

Figure 1: Male Informality Rates over the life-cycle

Figures 2.a shows that informality rates for the self-employed and employees are both decreasing over the life cycle, but informality in the self-employed is consistently larger. Informality rates which are larger for the elderly (Figure 1) are then a composition effect as a result of a larger proportion of self-employed. Maloney et al. (2007) argue that some workers with entrepreneurial abilities start their working life as salaried employees where they accumulate capital and experience to run their own businesses later in life. This is confirmed by Figure 2b. In summary, informality rates are decreasing over the life-cycle, they are much higher for self-employed workers than salaried employees, and that slightly increasing informality rates for the elderly are the effect of more workers becoming self-employed.

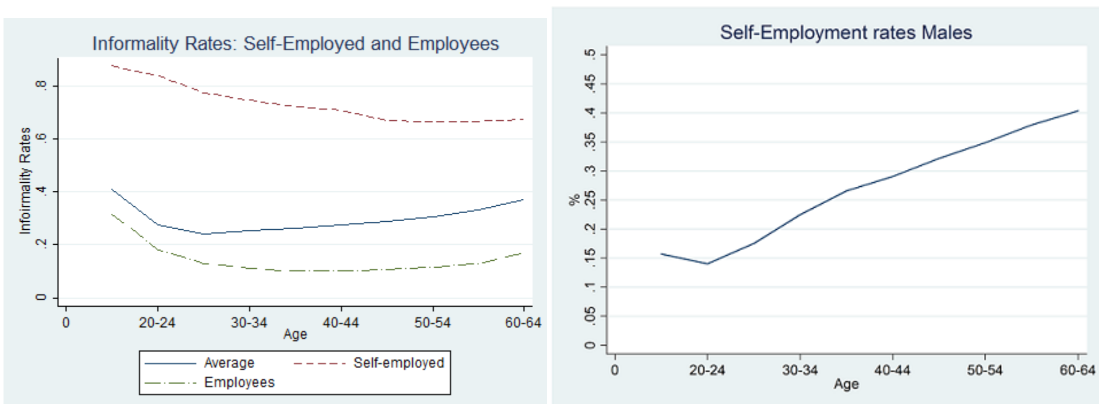

Figure 2: a) Informality Rates by Sub-sector; b) Self-employment rates over the life-cycle

Figure 3 confirms the evidence that formal workers earn in average more than their informal counterparts.

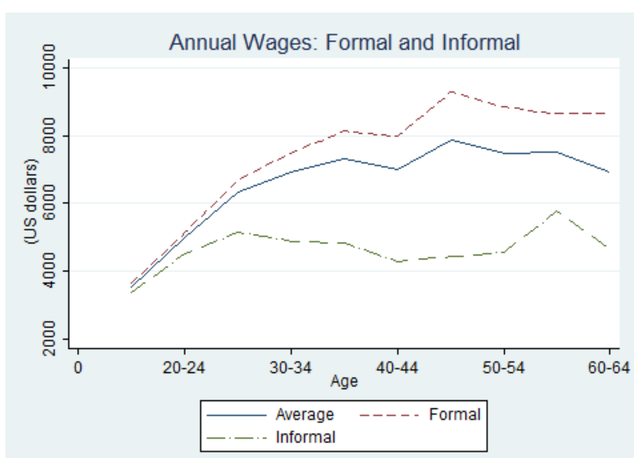

Figure 3: Net Wages in the formal and the informal sector

The data also shows a strong negative correlation between schooling and informality. Figure 4 confirms the results over the life-cycle. Remarkably, there the stronger difference in informality arises between high school degree and high school dropouts. Schooling differentials remain when informality rates are open in self-employed and employees groups (Appendix). Figure 6.

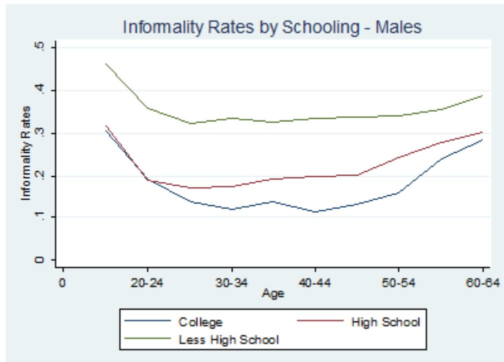

Figure 4: Informality Rates Males and Females by Educational level

Figure 6a shows the annual net wages over the life cycle by schooling levels. Returns to College are significantly higher than to the other two levels, while wages in the formal sector are higher in average than in the informal sector at every schooling level, as shown in Figures 7b, 7c and 7d.

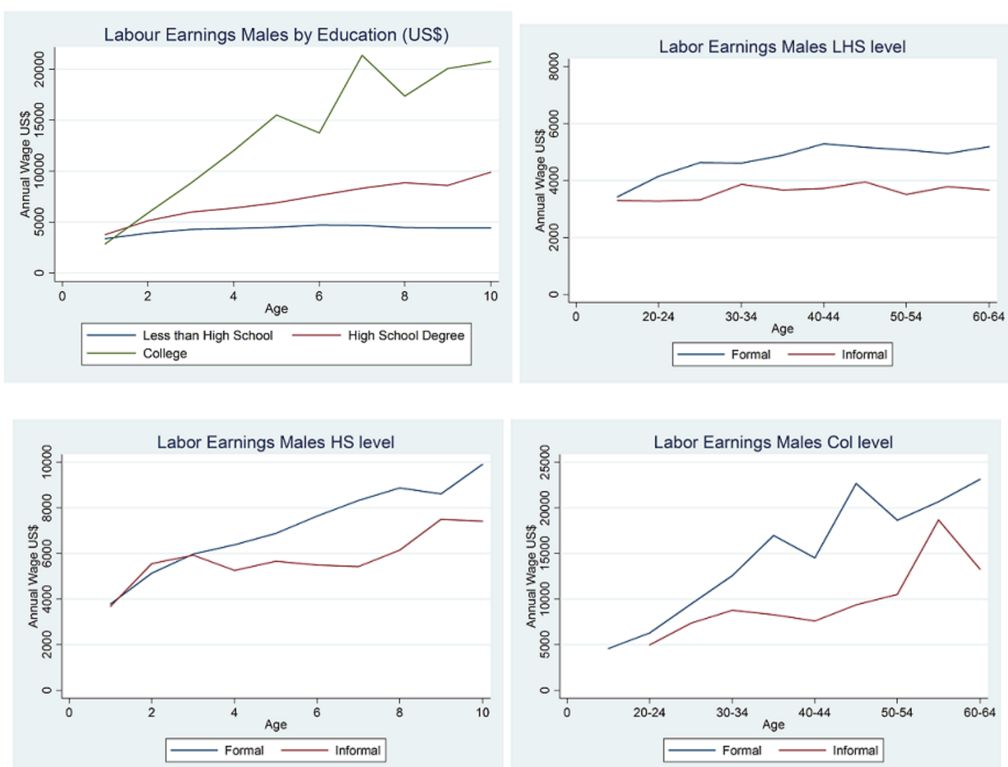

Figure 6: Wages Males a) Overall; b) Less than High School ; c) High School Degree; d) College

Finally, the data also shows strong mobility in/out formality decreasing in schooling. This is consistent with the hypothesis of competitive labor markets as mobility happens in both ways. Figure 8a shows that in fact only 30% of the sample of workers have always worked in the formal sector, and less than 10% have always worked in the informal one. Consistently with Levy (2008)and Maloney et al. (2007), mobility increase for the low-skilled (Figure 8b). Figure 9 shows that about 45% of workers have always worked as employees and 9% always as self-employed and lower mobility is observed at higher levels of schooling.

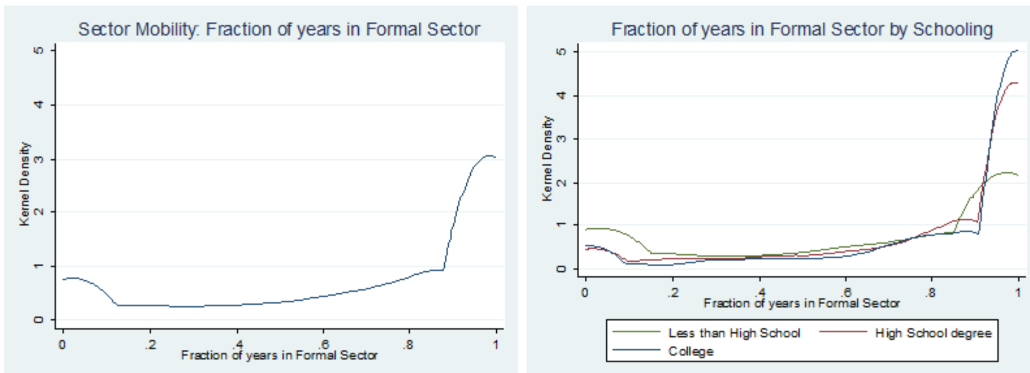

Figure 8: Mobility between Formal and Informal Sectors, a) Overall; b) By education Level

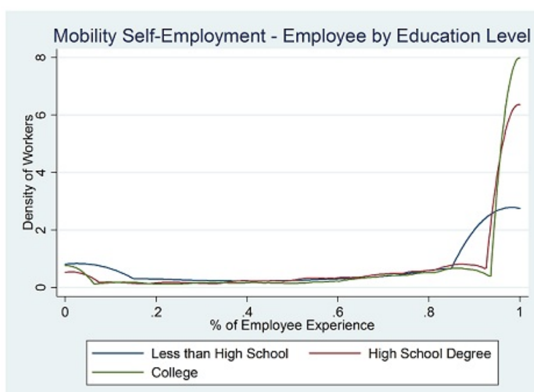

Figure 9: Mobility in/out Self-Employment by education Level

4 The Model

I build on the dynamic discrete choice model with unobserved heterogeneity proposed byKeane and Wolpin (1997), in which labor markets are competitive and utility maximizing workers decide education and labor based on their tastes, skills and expected gains. Wages are a function of the accumulated human capital embodied in a subject. I model individuals as risk neutral as private savings in the Chilean context are quite low. Nonetheless, an interesting extension of the model will be to evaluate choices of risk averse individuals in a context where pension savings only accumulate with labor experience in the formal sector.

In the model, individuals make their first choice at age 14: they can achieve three education levels: "Less than High School", "High School Degree" or "College". Everyone start with the first level at $t = 0$ as primary schooling is compulsory in Chile from 1962 and in the sample 96% of students finish primary schooling. At every period people decide whether to continue to the next schooling level, start working as formal/informal or self-employed/employee, or be voluntarily unemployed. Denote $m = \{FE, FS, IE, IS\}$ one of the four working sector choices, where FE =Formal Employee, FS =Formal Self-employed, IE =Informal Employee, and IS =Informal Self-employed. The choice of remaining unemployed is denoted by U.

The dynamics of the model comes from two sources. First, there is self-selection into schooling and working choices based on state-dependent current rewards on accumulated schooling and sector experience by that period. Second, there is also self-selection into schooling and working choices on the basis of expected sector-specific returns, which make current decisions depend on expected wages.

4.1 The State Space

Denote the state space Ω as the set of all variables which define the state-dependency of individual utilities over time. I estimating a life-cycle model so the time dimension t is the age of the individual. I detrend the data on wages and tuition fees to control for macroeconomic trends, and I take out cohort effects to compute data moments. Ed_{it} is the schooling level of individual i at age t. Then $Ed_{it} = \{LHS, HS, Col\}$ or Less than High School, High School Degree level, and College level.

Regarding labor experience, opening to a fourth-dimensional experience increases the state space exponentially. I assume that sector experience accumulates by self-employment or salaried employee, denoted by $X_{it}^{S}$and X_{it}^{E}, respectively.

Finally, denote the unobserved heterogeneity by μ_i, which entails abilities. More able people tend to self-select more into schooling, and at the same time, people have different set of skills that make them more productive in one sector more than in another, driving the choice. For example, entrepreneurship ability might drive self-selection into informal jobs or into self-employment, while the ability to work in very structured work environments might drive self-selection into formal jobs. Both of them are known by the individual and fixed from age 14, or $t = 0$. On the other hand, they are unobserved to the econometrician and need to be estimated along the rest of parameters. I incorporate permanent unobserved heterogeneity by modeling a discrete number of unobserved types (Heckman and Singer (1984)), where μ_k is an indicator variable that equals 1 if the individual is of type k.

4.2 Flow Utilities

At every period, individuals derive an instantaneous utility from attending schooling, staying at home or working in some economic sector. The costs of that decision are, in the case of schooling, foregone expected earnings or rewards from leisure/home production. When choosing the working sector, I also allow non-monetary rewards coming from sector-specific amenities or fixed costs, which vary along unobserved heterogeneity.

Denote the vector of available choices at t by $\{Ed, m, U\}$. Since everyone at $t = 0$ starts with $Ed_{it} = LHS$, people can make further schooling choices only for High school degree or College, then$Ed = \{HS, Col\}$. Denote U_{it}^{Ed} the instantaneous utility of attending schooling at level Ed. Then,

$$U_{t,k,R}^{Ed} = \sum_{k=1}^{K} \gamma_{1,k}^{Ed}\mu_k - \gamma_2^{Ed} TC_R^{Ed} + \eta_t^{Ed}$$

where TC_{iR}^{Ed} are the tuition costs paid by the household varying by schooling level Ed and municipality R, constructed with detrended costs varying over time and across municipalities. The factor load $\gamma_{1,k}^{Ed}$ represents psychic rewards to schooling or net costs of effort varying with the level of unobserved abilities and with the schooling level. This parameter capture psychic costs or the consumption value of the schooling decision. It also captures heterogeneous family background which translates into financial constraints to attend College. The term η_t^{Ed} is a random shock to indirect costs to schooling.

The utility of working in sector m is

$$U_{t,k}^m = \gamma_2^m W_{t,k}^m + \sum_{k=1}^{K} \gamma_{1,k}^m \mu_k + \gamma_3^m (X_t^m > 0) + \gamma_4^m (d_t = d_{t-1}) + \eta_t^m$$

where W_{tk}^m is the wage offered to individual type k at age t in sector m and d_{it} is the choice individual makes at age t. In a model with risk neutral individuals and returns from the stock markets similar to the interest rate, it does not matter the timing when social security contributions such as pension retirement are accounted. To the extent that there is full capitalization, as in the Chilean case, I consider gross wages, which in the formal sector already account for contributions. Several additional terms affect sector preferences: $\gamma_{1,k}^m$ reflects worker type-specific unobserved fixed costs of work. They capture the idea that the formal and the informal sector have different amenities like flexibility, autonomy and organizational structure.. γ_2^m captures the wage valuation, which might vary across sectors, while the remaining terms intend to capture mobility costs.γ_2^m reflect entry costs to work in a sector where the individual does not have previous experience, and γ_6^m represents switching costs between sectors from $t-1$ to $t.\eta_t^m$ are preference shocks related to the sector choice.

Finally, the utility of unemployment (leisure/home production) is

$$U_{tk}^U = \sum_{k=1}^{K} \gamma_{1,k}^U \mu_k + \gamma_4^U I(t > \underline{t}) + \eta_t^U$$

The reward that individual type k obtains from staying at home, which depends on unobserved skills captured by γ_{1k}^U, and on age effects γ_4^U. The latter parameter captures the fact that above certain age threshold $\underline{t}$, longer unemployment spells can damage the arrival of new job opportunities, perhaps because of skill depreciation. The term η_t^U is a random component reflecting uncertainty in the valuation of leisure or home production. For example, pregnancy can increase the valuation of unemployment for women.

4.3 Wages

Every time individuals choose to study they forego earnings from work. Since we have different working sectors, every time individuals choose a sector they also forego earnings in another sector. Since we cannot observe the counter-factual wage across sectors, a sector-specific log-wage equation is specified in the following fashion, where the objective is to estimate each of the parameters reflecting rental prices,

$$\begin{aligned} lnW_{t,k}^m &= \sum_{k=1} \alpha_{0k}^m \mu_k + \alpha_1^m * (Ed_t = HS) + \alpha_2^m * (Ed_t = Col) \\ &+ \alpha_3^m ln(1 + X_t^E) + \alpha_4^m ln(1 + X_t^S) + \\ &+ \alpha_5^m (X_t^m > 0) + \alpha_6^m (d_t = d_{t-1}) + \epsilon_t^m \end{aligned}$$

The set of coefficients α_{0k}^m represent the rental price for unobserved ability in sector m for individual type k. These parameters explicitly control for self-selection, and they are allowed to vary freely by sector. α_1^m and α_2^m capture the average returns to schooling. A richer version of the model also includes heterogeneous returns to schooling (Carneiro and Heckman (2003)). Individuals may have different sector abilities which may interact with schooling, and which are reflected in their productivity. α_3^mcaptures non-linear returns to experience as a salaried employee, whileα_4^m the non-linear the returns to self-employment. α_5^m and α_6^m represent skill depreciation factors in the wage equations reflecting the fact that accumulated sector experience do not have the same life-cycle profiles across sectors.

Shocks to wages are allowed to be correlated across sectors, then $\epsilon_t = (\epsilon_t^{FE}, \epsilon_t^{FS}, \epsilon_t^{IE}, \epsilon_t^{IS}) \sim N(0, \Sigma)$.

4.4 Uncertainty

The source of uncertainty in the model comes from the random shocks specified at each of the choices. Preference shocks are modeled logistic while wage shocks are normally distributed. Shocks are important to produce mobility across all choices in t because they shape expected utilities in each of the alternative choices from $t+1$ to T, affecting rewards from current choices. It is likely that shocks to productivity are correlated across choices; that is why I draw wage shocks from a multivariate normal distribution which are estimated along the rest of parameters of the model. I also assume that shocks are serially uncorrelated. This assumption does not decrease how rich the model is in producing dynamics, because permanent components are incorporated in the form of unobserved heterogeneity and mobility costs.

4.5 Value Functions

Self-selection into schooling and labor choices based on expected earnings which are state-dependent to current choices imposes a problem which is non-separable over time. The dynamics is as following: the model starts at $t = 0$ equivalent to age 14, when all individuals have finished primary schooling. At every subsequent year they must choose whether to continue studying across another year of the secondary schooling, or to dropout and start working, or to be at home. If they decide to finish high school, at age 18 they must decide whether to continue to College level or to drop out of education. This decision is taken year by year. If they drop out of school before the 4th level of high school then $Ed_t = LHS$ (Less than High School), if they drop immediately after the 4th level of secondary schooling then $Ed_t = HS$ (High School degree), and if they continue studying to College level then $Ed_t = Col$ (College). Maximum College schooling is standardized to last 5 years. If an individual drops at any schooling level, he/she cannot retake studying in the future, which is supported by the data. Therefore, at any t agents face five alternative choices for the computation of the expected value function in the next period, but for the cases of being at Less than High School or High School, in which case they face six possible alternatives in $t+1$, including the following schooling level. For example, the value of education at level $Ed = HS$ will be

$$V_{t,k}^{HS}(Ed_t, X_t^E, X_t^S, \mu_k, \eta_t, \epsilon_t) = U_{t,k,R}^{HS} + \beta Emax \left\{ \begin{array}{c} V_{t+1,k,R}^{Col}(HS, X_t^E, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{FE}(HS, X_t^E, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{FS}(HS, X_t^E, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{IE}(HS, X_t^E, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{IS}(HS, X_t^E, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{U}(HS, X_t^E, X_t^S, \mu_k, \eta_{it+1}, \epsilon_{t+1}) \end{array} \right\}$$

By choosing schooling individuals obtain the instantaneous utility U_{it}^{Ed} plus the discounted expected maximum value over available alternatives at $t+1$: continuing to study at College, working in one of the four sectors $\{FE, FS, IE, IS\}$, or becoming unemployed. Expectations are taken over the the distribution of the two shocks. Notice that $X_{it+1} = X_{it}$ when agents choose schooling.

Similarly, and just for illustration the value of working as a formal-employee at t is,

$$V_{t,k}^{FE}(Ed_t, X_t^E, X_t^S, \mu_k, \eta_t, \epsilon_t) = U_{t,k}^{FE} + \beta Emax \left\{ \begin{array}{c} V_{t+1,k}^{FE}(Ed_t, X_t^E+1, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{FS}(Ed_t, X_t^E+1, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{IE}(Ed_t, X_t^E+1, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{IS}(Ed_t, X_t^E+1, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{U}(Ed_t, X_t^E+1, X_t^S, \mu_k, \eta_{it+1}, \epsilon_{t+1}) \end{array} \right\}$$

while the value of being unemployed is

$$V_{t,k}^{U}(Ed_t, X_t^E, X_t^S, \mu_k, \eta_t, \epsilon_t) = U_{t,k}^{U} + \beta Emax \left\{ \begin{array}{c} V_{t+1,k}^{FE}(Ed_t, X_t^E, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{FS}(Ed_t, X_t^E, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{IE}(Ed_t, X_t^E, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{IS}(Ed_t, X_t^E, X_t^S, \mu_k, \eta_{t+1}, \epsilon_{t+1}) \\ V_{t+1,k}^{U}(Ed_t, X_t^E, X_t^S, \mu_k, \eta_{it+1}, \epsilon_{t+1}) \end{array} \right\}$$

In the former case, given the choice $m = FE$ at t, agents accumulate one more year of experience as employees. This changes the accumulated experience as an employee, which in turns change the valuation of each of the possible choices at $t+1$. A similar rule is applied for the other sectors, while the choice of unemployment does not alter the state space for the next period.

Mobility

In the model, mobility across sectors is generated by three sources: First, sector-specific wages are affected by random shocks to incomes. A individual may switch from sector m to $\tilde{m}$ if the shock in the latter sector is larger and it more than compensate the lost in returns to experience from the former. Second, even if the shocks across two sectors in two periods are exactly the same, the individual might still switch as I allow sector and cross-sector returns to experience to vary freely. Finally, I explicitly model mobility costs, which also vary by sector. They include non-monetary transition costs or skill depreciation factors, as well as the implied cost of not having previous experience in one particular sector.

5 Model Solution and Calibration Exercise

5.1 Solution Procedure

Dynamic discrete choice models with risk neutral agents do not have analytical solution. With finite periods, the model must be solved numerically using backward recursion. At period T, each individual draw random shocks from the multidimensional error vector (η_T, ϵ_T) and chooses the alternative that yields the maximum instantaneous utility evaluated at every possible state space combination of schooling and labor histories. I assume that terminal value function over the life-cycle is $V_{iT+1} = 0$. Denote $d^*_{it} = \{Ed, m, U\}$ the optimal choice at every period. Then, at period T individuals solve

$$d^*_{iT} = argmax(U^{Ed}_{iT}, U^{m}_{iT}, U^{U}_{iT})$$

At period $T-1$, two steps are required to compute the value functions. First, they need to evaluate expectations over T computing the Emax functions, where expectations are taken over (η_T, ϵ_T), evaluated at every possible choice and state space combination at $T-1$. This involves a multidimensional numerical integration in the following way,

$$Emax[V^{Ed}_{iT}, V^{m}_{iT}, V^{U}_{iT}] = \int\limits_{\eta}\int\limits_{\epsilon} max[V^{Ed}_{iT}, V^{m}_{iT}, V^{U}_{iT}/d^*_{iT-1}, \Omega_{iT-1}]f(\epsilon)d\epsilon f(\eta)d\eta$$

Notice that both η_T and ϵ_T are vectors itself, so the dimension of the integration is the sum of the components within them. Nonetheless, the advantage of modeling preference shocks η_{it} as logistic is that the expected value ($Emax$) has a closed form expression so we can avoid the numerical integration over each of its components. Therefore, the evaluation of the $Emax$ function is reduced to a four dimensional numerical integration corresponding to the dimension of the wage shocks, which are drawn from a multivariate normal distribution and therefore can be correlated across the four working sectors. To deal with multidimensional numerical integration in a tractable way I use Gauss-Hermite Quadrature Rules (Judd (1992)).

Second, they must evaluate the instantaneous utilities at $T-1$, again for every possible combination of the steady state at that period, drawing the error vectors $(\eta_{T-1}, \epsilon_{T-1})$ and compute the value functions at $T-1$: $(V^{Ed}_{T-1}, V^{m}_{T-1}, V^{U}_{T-1})$. The optimal choice at $T-1$ is then obtained from

$$d^{*}_{T-1} = argmax(V^{Ed}_{T-1}, V^{m}_{T-1}, V^{U}_{T-1})$$

The process then is repeated in the same fashion until $t = 0$.

5.2 Calibration exercise

Previous to the estimation process I have conducted a simple calibration exercise to find out whether a reasonably calibrated Human Capital investment model of this fashion is able to reflect the patterns of the main moments from the data previous to the estimation process. It's important to note that the final version of the paper will not include the calibration but the full estimation of the model and policy predictions based on those estimates.

In order to achieve a reasonable calibration of the model I have used the wage returns found by Todd, Mukhopadhyay and Bravo (2008), who estimate a dynamic discrete choice model for males using the same database, but do not estimate sector-specific returns. Thus, my calibration strategy was to preserve the average returns they find splitting them into formal/informal and self-employed/employee sectors. Furthermore, I adjust preference parameters related to non-monetary payoffs arbitrarily so as to reproduce basic moments from the data: the proportion of individuals attending schooling, working in some sector or remaining unemployed at every period during the life-cycle, informality rates by education level, and wages by education level, all of them in means and variances. Especially important is the case of the discount factor. This cannot be assumed arbitrarily. I use a coefficient $\beta = 0.96$ which has already been used in the literature for the Chilean case. In simulations, secondary schooling has been standardized to 4 years of duration and college to 5 years. The data on tuition fees comprises 15 years and more than 300 municipalities.

In terms of magnitude, I simulate the model for a cohort of 10.000 individuals for a length of time of 51 years. Recall that it is assumed that everyone finishes primary schooling, so before period $t = 0$ everybody attend schooling. Figure 10a shows the data on life-cycle proportions of individuals by choices, and Figure 10b its counterpart from simulation results. The model does a good job at representing life-cycle choices except for the final part of the life-cycle, in which formality is decreasing in the data while in the model remains constant.

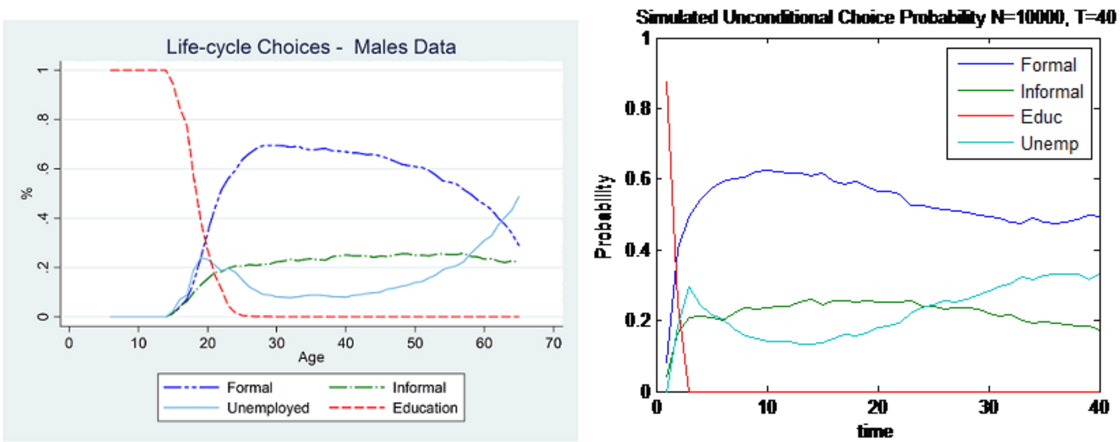

Figure 10: Proportions of life-cycle choices from a) Data; b) Simulations

Figure 11a shows simulated Informality Rates. Recall Figure 1a, which shows informality rates from the data. The model is able to replicate higher rates for the youth and for the elderly, but informality rates fall in the last part of the life-cycle. Figure 11b opens informality rates by schooling level, doing a good job in obtaining decreasing rates on schooling level.

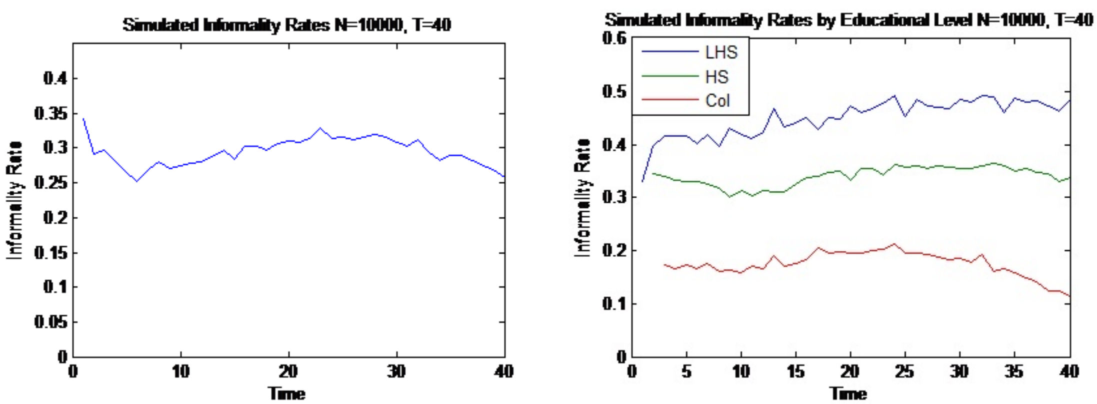

Figure 11: Informality Rates. a) Overall; b) by Educational Level

Figures 12 and Figure 13 represent simulated life-cycle wages. The calibrated model is able to simulate increasing and concave wages as it is expected from the human capital investment theory and the data. Wages in the formal sector are larger than in the informal one and they are increasing in schooling (Figures 12a and 12b respectively). They show similar patterns than the wages from data. Once I open formal and informal wages by educational level as in Figure 13, simulations results are consistent with the data. Formal wages are larger than informal ones once we control for the level of education, consistently with the hypothesis that formal firms operate at higher levels of productivity, paying on average higher wages to formal workers.

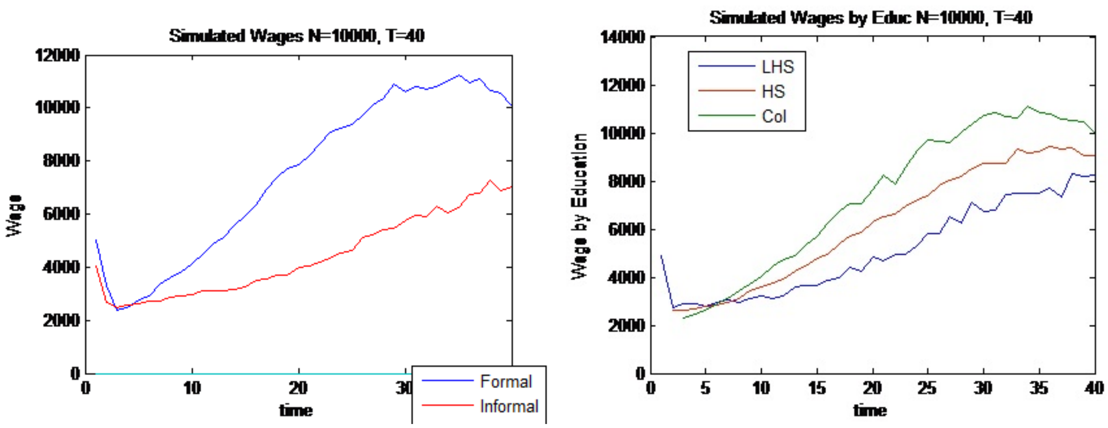

Figure 12: Life-cycle wages. a) Open by Formal-Informal; b) Open by Educational Level

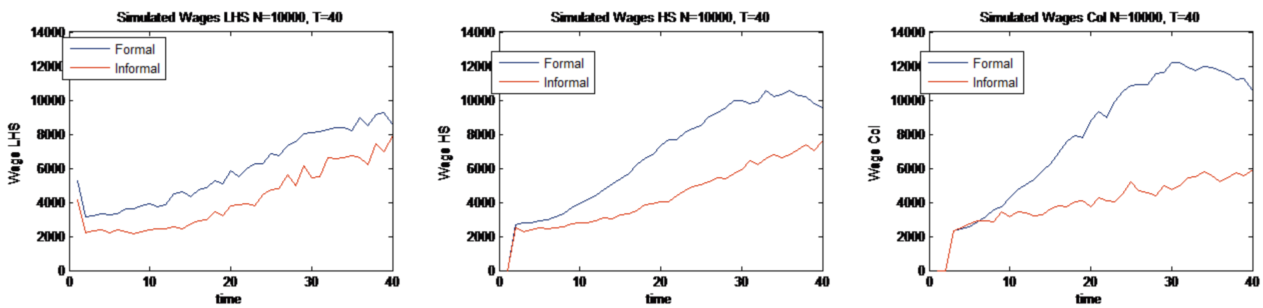

Figure 13: Life-cycle wages open by formal-Informal and educational level. a) Less than High School; b) High School; c) College

Finally, Figures 14a and 14b show simulated mobility between the formal and the informal sector in the same fashion as Figure 8. The model manages to predict a bimodal distribution of the percentage of formal experience over the life-cycle, where about 40% of the workers have almost 100% formal experience, less than 10% almost pure informal experience, and half of the workers have mixed experience between the formal and the informal sector. Interestingly, the model also does a good job in reflecting decreasing mobility rates in schooling.

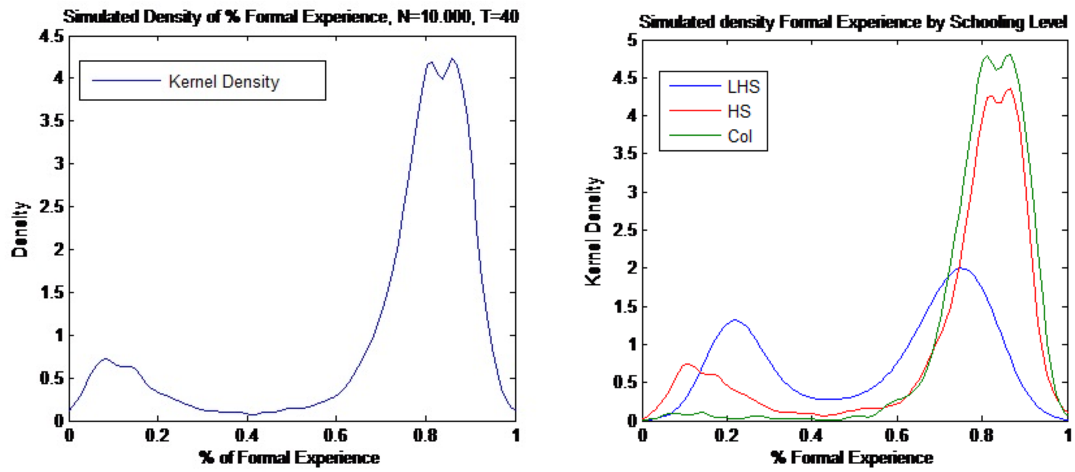

Figure 14: Mobility Formal-Informal sector; a) Overall, b) by Schooling

In conclusion, the calibration results show that an augmented human capital investment model has a potential explanatory power in explaining the main patterns of the co-variation between life-cycle schooling, informality and self-employment choices. In fact, I am able to reproduce quite accurately the proportions of life-cycle choices, informality rates, wages and mobility across sectors at different levels of schooling.

6 Estimation (Work in Progress)

I estimate the model by Indirect Inference (Gourieroux, Monfort and Renault (1993)) using a gradient-based optimization algorithm. As the dynamic programming problem requires to be solved as many times as the number of unobserved types accounted, I introduced types with discrete and finite support. In my model, k types are considered, where $k = 1, .., 4$, and the probability of being each type is estimated along the rest of structural parameters.

6.1 Indirect Inference

Meghir and Rivkin (2010) emphasize the use of simulation methods for structural estimation. The main advantage of these methods is that they do not use all the information and restrictions implied by the model, given the available data, as MLE methods, speeding up the estimation process. The accuracy of the estimated parameters depends only on a good specification of the moments identifying the data, which is relatively simple in linear models. As the model solution does not have an analytic representation, data moments must be simulated through a data generation process starting from arbitrary parameters.

The idea of indirect inference is to simulate data with the model starting from some initial vector of structural parameters (θ), and pass both the actual and simulated data by a certain auxiliary model, usually a system of linear regressions, to generate a set of data auxiliary parameters β and simulated auxiliary parameters $\beta^S(\theta)$. At each iteration of the structural parameter θ_i, Indirect Inference finds the structural estimates at next iteration $\theta_{i+1}$by minimizing the distance between the data and simulated auxiliary parameters.

The model is then used to generate simulated moments and to estimate structural parameters θ by convergence over a specified number of simulation iterations. In each of them, random shocks are drawn. Denote s the number of iterations and θ_0 an initial arbitrary set of parameters. In the first iteration, the set of initial simulated auxiliary estimates is $\beta^s(\theta_0)$ and the following set of converging parameters $\theta_1$is found by minimizing the distance between simulated and data auxiliary parameters.

The objective function at each simulation s is then is given by the metric

$$Min_\theta(\beta - \beta^s(\theta))'W(\beta - \beta^s(\theta))$$

where $W = VCV(\beta)^{-1}$ is the optimal weighting matrix obtained from the data auxiliary estimates. The process is repeated either until convergence or until the maximum number of pre-specified simulations S is reached.

6.2 Smoothing the Objective Function

Keane and Smith (2003) propose that standard selection model with wages and choices should be estimated matching a set of linear regressions including log wages and LPM for choices like

$$\begin{bmatrix} lnW_{it} \\ d_{it}^{J} \end{bmatrix} = Z_{it}'\delta + \nu_{it} \sim N(0, \sum)$$

where Z_{it} is a vector of observable regressors in the data including the schooling level, sector experience and age. In the case of schooling participation, this vector also includes as covariates the data on tuition fees for each educational level, so that I can identify structural parameters related to direct costs to schooling $(\gamma_2^{HS}, \gamma_2^{Col})$. The vector of auxiliary parameters β would then compress then the mean coefficients of the auxiliary regressors (δ), and the VCV of the residuals from all the regressions $\sum$. The latter captures second moments from the data identifying structural parameters related to the VCV matrix of wage shocks.

Despite the conceptual simplicity of the method, different authors stress out that objective functions are step functions in structural parameters in dynamic discrete choice models, which makes impossible the use of derivative-based methods (Magnac, Robin, and Visser (1995), An and Liu (2000), and Nagyp´al (2000)). Derivative-based methods are generally preferred to local or global search methods because of speed and accuracy considerations. To correct for this issue, Keane and Smith (2003) propose an alternative system of auxiliary regressions in the simulated data replacing the choice d_{ia}^{J} by a smooth function of the structural parameters obtained from the simulated value functions

$$g^{J}(\theta) = \frac{exp\left(V^{J}(\theta)/\lambda\right)}{\sum_j exp\left(V^{J}(\theta)/\lambda\right)}$$

where $g^{J}(\theta)$ can be interpreted as the probability of picking alternative J, is smooth in θ, and$\lambda$is a calibrated smoothing parameter. In that way, mirroring the set of data auxiliary regressions, the following system is estimated from the simulated data at each iteration

$$\begin{bmatrix} \sum_{J} g_{iat}^{J} * lnW_{it}^{J} \\ g_{it}^{J} \end{bmatrix} = Z_{it}'\beta^{S}(\theta) + \nu_{it}^{S} \sim N(0, \sum)$$

Notice that in the simulated auxiliary system we can observe each of the counter-factual wages, while in the actual data we only have the wage at the chosen sector. Therefore, in the simulated auxiliary system the wage regression has in the LHS the expected simulated log-wage, which is the sum of the product of the simulated sector wage and the probability of choosing that sector.

6.3 Matching Moments

In order to gain identifying power of the structural parameters, I do not only include auxiliary regressions for Log Wages and Participation, and the respective VCV of the residuals of these regressions. These regressions were found to be useful to identify the means of the structural parameters and the variance of wages. Nonetheless, the covariances of wages across sectors were only identified after including regressions for Transition probabilities and Growth of log wages across sectors conditional on observables. Furthermore, I also needed to gain identifying power of the type-specific returns to ability and psychic costs to schooling. I was able to identify them by also matching the Proportions of people below wage percentiles {10,25,50,75 and 90} by education level and sector, and by including Transition probabilities regressions to unemployment conditional on being below certain wage percentile in the previous period. The total set of moments accounts for 254, for the estimation of 45 structural parameters.

Table 3 describes the set of macthing moments and the corresponding identified structural parameters. The full set of equations of moments is described in Appendix B.

Group	Subgroup	Parameter	Moments
Wages	A. Returns to Schooling and Experience	$\alpha_1^m, \alpha_2^m, \alpha_3^m, \alpha_4^m$	A1. Log wage regressions on $\{Ed, X^m, age\}$
	B. Variance of Wages	$\sum$	B1. VCV of residuals of log wages and participations
	C. Covariances, Mobility Costs	σ_{mn}, γ_3^m	C2. Wage growth across sectors on $\{Ed, age\}$
			C3. Transition probabilities on $\{Ed, X^m, age\}$
	D. Type-specific returns to ability	α_{0k}^m	D1. Proportions Log wages by sector
			D2. Mean Log wages by sector
Sector Participation	F. Wage valuation	γ_2^m	F1. Participation regressions on $\{Ed, X^m, age\}$
	G. Type-specific fixed costs to work	$\gamma_{1,k}^U, \gamma_{1,k}^m$	G1. Transition from Sectors to Unemployment conditional on past wage quintiles
			G2. Transition between Sectors conditional on wages
Schooling	I. Tuition Costs valuation	γ_2^{Ed}	I1.Schooling regression on TC data and age
	J. Psychic costs	$\gamma_{1,k}^{Ed}$	J1. Transitions Schooling to Sectors

Table 3: Set of Moments

6.4 Estimation Exercise

The full estimation with the real data is still work in progress. Nonetheless, in order to check whether the estimation procedure works, I have performed the following identification exercise. I generated an artificial dataset from the model, and I estimate it starting from an initial guess deviated 20% from the "true" parameter. If the set of matching moments is the right one, the model should show convergence to the true parameters. The exercise has been done so far for a simplified version of the model with only two sectors: the formal and the informal. The results in Table 4 show full convergence under the described procedure.

Group	Parameters	β_o	β_1	β_{est}	dif β_o, β_1	dif β_0, β_{est}
Psychic costs to schooling	$\gamma_{1,k=1}^{HS}$	1	1.200	0.998	20.0%	-0.2%
	$\gamma_{1,k=1}^{Col}$	1	1.200	0.975	20.0%	-2.5%
	$\gamma_{1,k=2}^{HS}$	0.9	1.180	0.901	20.0%	0.2%
	$\gamma_{1,k=2}^{Col}$	1.1	0.88	1.08	-20.0%	-2%
Returns to Ability	$\alpha_{0,k=1}^{F}$	8.189	9.827	8.192	20.0%	0.0%
	$\alpha_{0,k=1}^{I}$	8.007	6.406	8.005	-20.0%	0.0%
	$\alpha_{0,k=2}^{F}$	8.189	9.827	8.192	20.0%	0.0%
	$\alpha_{0,k=2}^{I}$	8.007	6.406	8.005	-20.0%	0.0%
Returns to High School	α_1^F	0.326	0.391	0.324	20.0%	-0.7%
	α_1^I	0.362	0.290	0.363	-20.0%	0.4%
Returns to College	α_2^F	1.307	1.568	1.312	20.0%	0.4%
	α_2^I	1.208	0.966	1.208	-20.0%	0.0%
Returns to Experience same sector	α_3^F	0.2	0.280	0.200	40.0%	0.0%
	α_3^I	0.1	0.060	0.100	-40.0%	0.4%
Returns to Cross-sector experience	α_4^F	-0.1	-0.080	-0.1011	-20.0%	1.1%
	α_4^I	0.1	0.120	0.101	20.0%	1.0%
VCV wages	σ_{FI}	0.1	0.080	0.110	-20.0%	10.0%
	σ_F^2	0.7	0.840	0.699	20.0%	-0.2%
	σ_I^2	0.7	0.560	0.701	-20.0%	0.1%
Tuition Costs	γ_2^{HS}	1	1.200	1.012	20.0%	1.2%
	γ_2^{Col}	1	1.200	1.010	20.0%	1.0%

Table 4: Estimation Exercise

7 Concluding Remarks

In this paper I propose a structural model to study the relationship of human capital and labor informality using the Social Protection Survey, a panel data set from Chile. The modeling framework is based on a Human Capital Investment models in the fashion of Keane and Wolpin (1997) and Attanasio, Meghir and Santiago (2010), in which heterogeneous agents in skills and preferences self-select into schooling and labor choices based on current and expected gains. Up to the extent that formal firms operate in higher productivity ranges as a result of commitment with larger labor costs, and some firms have incentives to hire informally to reduce those labor costs, I contribute to the literature of informality by considering endogenous schooling choices and by considering labor informality as the result of a utility maximization problem, in which agents decide in terms of their comparative advantage, evaluating their payoffs in terms of expected sector gains. This is at the core of the comparative advantage theory, which reconciles the traditionally opposite views of segmented vs. competitive markets to explain how informality arises.

Some model features must be remarked. First, I added complexity to the model by incorporating the self-employed and employee sub-sectors within the formal and informal sectors. Adding them seem to be necessary if I am about to explain life-cycle patterns of labor informality which a two-sector model could not explain. Second, the model presented is the simplest version of this type of models, in which mobility across sectors is basically explained by the interaction of shocks across choices, and the returns to experience in the same sector and across sectors. The shocks across choices, although contemporaneously correlated, have no persistence. The incorporation of transition costs in the form of mobility costs and skill depreciation factors in the solution could add more dynamics to the model in order to informality and self-employment rates more accurately. Third, the assumption of risk neutral individuals has been taken to simplify the modeling and estimation, but an interesting extension is the study of additional effects of liquidity constraints and pension savings on the informality decision in a context of risk averse individuals. Finally, a partial equilibrium approach has been chosen as a first stage modeling of human capital and informality because I want to understand the nature of decisions and the incentives that workers face. This is also a good exercise to understand the first impulses produced by schooling policies, showing how sensitive is informality to the schooling decision. However, long-term policy effects will be studied in a GE framework, which is the next stage of the research agenda.

Preliminary calibration results using estimated parameters from similar models using the same database show that the augmented human capital investment model does a good job in explaining the main patterns of the most representative data moments. In fact, I am able to reproduce quite accurately the proportions of life-cycle choices, informality rates, wages and mobility across sectors at different levels of schooling. Estimation results are work in progress but preliminary estimation exercises with artificial data generated from the model shows promising results for the full estimation procedure.

References

Marcelo Arbex , Antonio F. Galvao, Jr. and and Fabio Gomes (2010) “Heterogeneity in the returns to education and informal activities”, Ibmec Working Papers wpe_216, Insper Working Paper, Insper Instituto de Ensino e Pesquisa.

Aguirregabiria, V., and P. Mira. 2007. “Sequential Estimation of Dynamic Discrete Games”, *Econometrica*, 75, 1-54.

Amaral, Pedro, and Erwan Quintin (2006) “A Competitive Model of the Informal Sector.”Journal of Monetary Economics, 53(7), 1541-53

Arias, Omar, P. Fajnzylber, W. Maloney, A. Mason, G. Perry, and J. Saavedra-Chanduvi. “Informality: Exit and Exclusion”. Washington: The World Bank.

Attanasio, Orazio, C. Meghir and A. Santiago. 2005. “Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progresa”, IFS Working Papers W10/14, Institute for Fiscal Studies.

Bosch, Mariano, and W. Maloney. 2006. “Gross Worker Flows in the Presence of Informal Labor Markets: The Mexican Experience”. Washington: The World Bank.

Carneiro, Pedro, and J. Heckman 2003. “Human Capital Policy” (July 2003). IZA Discussion Paper No. 821.

Contreras, D., L. de Mello and E. Puentes. 2008. “Tackling Business and Labor Informality in Chile”. OECD Economics Department, No. 607, OECD Publishing.

Gasparini, L. and L. Tornarolli. 2007. “Labour Informality in Latin America and the Caribbean: Patterns and Trends from Household Survey Microdata”, Working Paper , No. 0, CEDLAS, Universidad Nacional de La Plata, La Plata, Argentina.

“Indirect Inference”. Journal of Applied Econometrics, Vol. 8, Supplement: Special Issue on Econometric Inference Using Simulation Techniques. (Dec., 1993), pp. S85-S118. Stable

Guha-Khasnobis, Basudeb, R. Kanbur, and E.Ostrom. 2006a. “Beyond Formality and Informality.” In *Linking the Formal and Informal Economy: Concepts and Policies*, edited by B. Guha-Khasnobis, R. Kanbur, and E. Ostrom. Oxford University Press

Harris, John, and M. Todaro. 1970. “Migration, Unemployment and Development: A two-Sector Analysis.” *American Economic Review* 60, pp. 126-42.

Heckman, James, and G. Sedlacek. 1985. “Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market”. *Journal of Political Economy* 93 (6): 1077-125.

Heckman, James, and C. Pagés. 2004. “Law and Employment: Lessons from Latin America and the Caribbean.” University of Chicago Press for the National Bureau of Economic Research.

Heckman, J. J. and B. Singer (1984): “The Identifiability of the Proportional Hazard Model,” Review of Economic Studies, 60, pp. 231–243.

Judd, K., (1998). Numerical Methods in Economics. MIT Press, Cambridge, MA.

Anthony A. Smith, Jr. & Michael Keane, 2004. "Generalized Indirect Inference for Discrete Choice Models," Econometric Society 2004 North American Winter Meetings 512, Econometric Society.

Keane, Michael and Kenneth Wolpin. 1997. “The Career Decisions of Young Men”. *Journal of Political Economy*, 105, 473-522.

Levy, Santiago. 2008. “Good Intentions, Bad Outcomes”. Brookings Institution Press, Washington, D.C.

Magnac, Thierry. 1991. “Segmented or Competitive Labor Markets”. *Econometrica*, Vol.59, No. 1 (Jan., 1991), pp. 165-187.

Maloney, William. 1999. “Does Informality imply Segmentation in Urban Labor Markets? Evidence from Sectorial Transitions in Mexico.” *World Bank Economic Review* 13, pp. 275-302.

Maloney, William and J. Nunez Mendez, 2003. "Measuring the Impact of Minimum Wages: Evidence from Latin America," NBER Working Papers 9800, National Bureau of Economic Research, Inc.

Meghir, C. , R. Narita and J.M. Robin. 2012. “Wages and Informality in Developing Countries”, NBER Working Paper No. 18347.

Meghir, C. and Steven G. Rivkin. 2010. "Econometric Methods for Research in Education," NBER Working Papers 16003, National Bureau of Economic Research, Inc.

Todd, P., S. Mukhopadhyay and D. Bravo (2008): “How Universal School Vouchers Affect Educational and Labour Market Outcomes: Evidence from Chile”, forthcoming in *The Journal of Political Economy*, 2011.

Willis, R. and S. Rosen: “Education and Self-selection” Journal of Political Economy, 1979.

Thomas, Jim (1992). Informal Economic Activity, London School of Economics Handbooks in Economics, London School of Economics: London.

Appendix

A. Figures

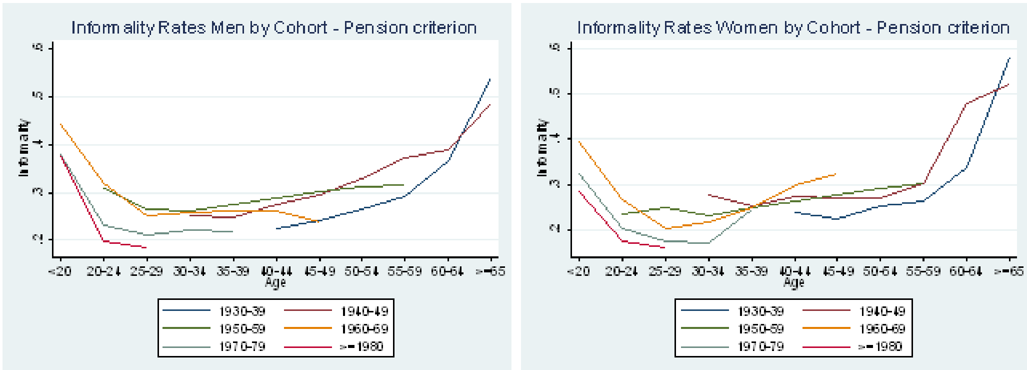

Figure A.1 Informality rates controling by cohort effects.

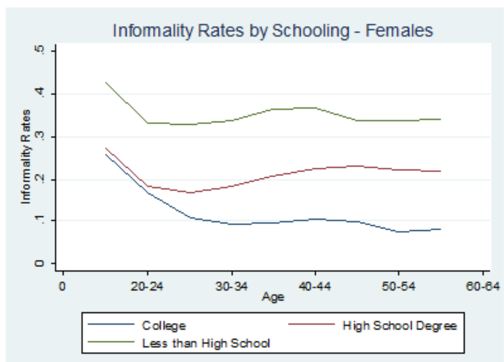

Figure A.2 Informality rates by schooling females

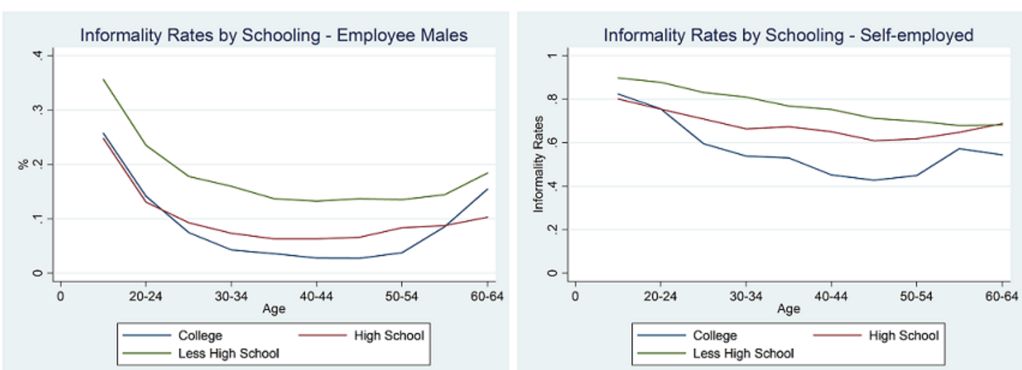

Figure A.3 Informality rates by schooling

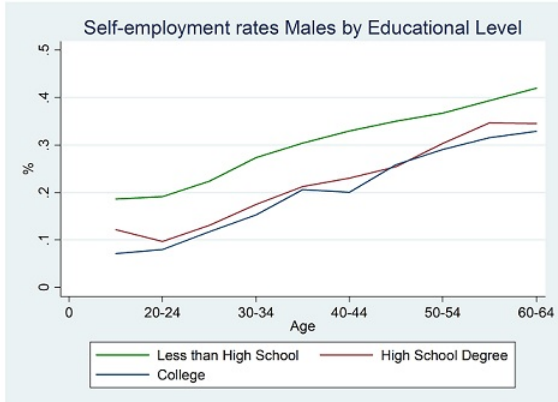

Figure 5: Self-Employment rates by Education Level

B. Auxiliary Regressions

1. Wages and Participation regressions

- $d_t^J = \{d_t^{FE}, d_t^{FS}, d_t^{IE}, d_t^{IS}, d_t^U\}$ the labor choices
- $d_{t=[1,4]}^{HS}, d_{t=[5,9]}^{Col}$ schooling choices

(a) Data

$$
\begin{aligned}
lnW_t &= Z_t^1 \delta^W + \upsilon_t^W \\
d_t^J &= Z_t^1 \delta^J + \upsilon_t^J \\
d_t^{Ed} &= Z_t^2 \delta^{Ed} + \upsilon_t^{Ed}
\end{aligned}
$$

where

- $Z_t^1 = \{Ed, log(1 + X_t^E), log(1 + X_t^S), age, age^2\}$
- $Z_t^2 = \{TC_R^{Ed}, age, age^2\}$

(b) Smoothing: Objective functions step functions in β

- $g^J(\beta)$ smooth proxy for choice probabilities

$$g^J(\beta) = \frac{exp\left(V^J(\beta)/\lambda\right)}{\sum_j exp\left(V^J(\beta)/\lambda\right)}$$

(a) Simulations regressions

$$\sum_m g_t^m * lnW_t^m = Z_t^1 \delta^W(\theta) + \nu_t^W$$

$$g_a^J = Z_t^1 \delta^J(\theta) + \upsilon_t^J$$

$$g_a^{Ed} = Z_t^2 \delta^{Ed}(\theta) + \upsilon_t^{Ed}$$

2. VCV matrix of residuals$\upsilon \sim N(0, \Sigma)$

3. Transition Probabilities

(a) Data

$$P(d_t^J = 1 \mid d_{t-1}^{J'} = 1) = Z_t^1 \delta^{JJ'} + \nu_t^{JJ'}$$

(b) Simulations

$$g_t^J \mid (d_{t-1}^{J'} = 1) = Z_t^1 \delta^{JJ'}(\theta) + \nu_t^{JJ'}$$

4. Growth of Log Wages

(a) Data

$$lnW_t^J - lnW_{t-1}^{J'} = Z_t^1 \delta_w^{JJ'} + \xi_t^{JJ'}$$

(b) Simulations

$$g_t^J lnW_t^J - g_t^{J'} lnW_t^{J'} = Z_a^1 \delta_w^{JJ'}(\theta) + \xi_t^{JJ'}$$