A Nested Logit Model of Occupational Attainment and Labour Income Inequality in Chile during the 1990s

ALEXIS PALMA[1]

Borås University

Abstract

This paper studies the inequality of earnings in 12 occupational groups in the Chilean labour market during the 1990s. Using the estimates of the earnings regressions and the inequality decomposition of Fields and Yoo (2000), we are able to explain 44% of the inequality of earnings of white-collar self-employed workers in 2000, 38% of white-collar informal workers, and 33% of the inequality of earnings of white-collar formal workers. Education is found to be a key variable in these levels of occupation accounting for 30%, 32%, and 21%, respectively. Looking at the earnings inequality of workers of lower-level occupations, we are able to explain 38% in the self-employment sector, 41% in the informal, and 25% in the formal sector. Among these workers formal education explains a lower share of their earnings inequality, being higher among individuals of the formal sector. On the other hand, hours of work is a more important variable contributing to explain 26% of the earnings inequality among self-employed manual workers, 28% among manual informal workers, but only 5% of the earnings inequality of their counterparts in the formal sector.

Resumen

Este documento estudia la desigualdad de los ingresos de 12 grupos ocupacionales en el mercado laboral chileno durante la década de los noventa. Usando datos del año 2000, las estimaciones de las regresiones de ingresos y la descomposición de desigualdad de Fields y Yoo (2000) podemos explicar el 44%, el 38% y el 33% de la desigualdad de los ingresos de los profesionales que trabajan por cuenta propia, que trabajan en el sector informal y que trabajan en el sector formal respectivamente. La educación es una variable clave para explicar la desigualdad y explica el 30%, 32% y 21% de la desigualdad de los ingresos de estos grupos ocupacionales. Cuando

[1] Borås University, Allégatan 1, 501 90 Borås, Sweden. E-mail: alexis.palma@hb.se. I am grateful to Osvaldo Salas and to my colleagues at Borås University for useful comments.

analizamos la desigualdad de los ingresos de los trabajadores no calificados podemos explicar el 38% de la desigualdad de los trabajadores por cuenta propia, el 41% de la desigualdad de los trabajadores en el sector informal y el 25% de la desigualdad de los trabajadores en el sector formal. En estos grupos la educación explica una menor proporción de la desigualdad de ingresos, siendo mayor entre los individuos del sector formal. Por otro lado, las horas de trabajo es una variable más importante que la educación y contribuye a explicar el 26% de la desigualdad de ingresos entre los trabajadores no calificados por cuenta propia, el 28% entre los trabajadores no calificados del sector informal, pero solo el 5% de la desigualdad de los trabajadores no calificados del sector formal.

***Keywords*:** Earnings inequality, Inequality decomposition, Nested Logit.

1 Introduction

The purpose of this paper is to shed light into the factors that affect the occupational outcomes of Chilean household heads, among 12 different occupational groups, and the determinants of earnings inequality within them during the 1990s.

The development of the Chilean economy during the 1990s does not have any counterpart in other periods of Chilean history. After the dramatic turnover to a highly liberalized market economy in the early 1970s, and the deep economic crisis of the early 1980s, since late 1980s most key economic indicators improved during the 1990s. For instance, real GDP/Capita increased by 91% between 1987 and 2000, see Table 1. The growth rate ranged from 2.0% to 10.4% over the same period, although it was higher during the period 1992-1996. Unemployment was higher in 2000 than in previous years of the 1990s, most probably due to the deceleration of the economy during the late 1990s. This implied that the number of unemployed increased by over 160 thousand individuals comparing 2000 with 1992. Perhaps one of the most important achievements of the Chilean economy during this period is the substantial reduction of household poverty rates which decreased from 39.4% in 1987 to 16.6% in 2000; an achievement with few counterparts outside East Asia, Gill and Montenegro (2002).

However, one area of strong disappointment, in this otherwise quite impressive development, is the lack of downward pressure on income inequality indicators as suggested by Table 1. The Gini of the household per capita income decreased somewhat from 1987 to 1994 but stayed constant during the rest of the years in Table 1. Even the Gini of the hourly income of wage-and-salary workers was highly stable during this period. It is only for own-account workers that the Gini reports a high range of variation, decreasing during the second half of the 1990s.

The stability of the Chilean income inequality and its high level is an observation that has been stressed by several scholars and has been the focal point of many studies; see for instance Litchfield (2002). Several of these studies have put forward the rate of return to education as a key source of income inequality in Chile; see for instance Beyer (1997), Contreras (2002) and Contreras (2003). This result is important as there is evidence that Chilean workers become more and more educated and post-secondary education characterized an increased proportion of Chilean workers. From mid-1980s to mid-1990s the group of occupied workers with 13 or more years of education increased by more than 180%, compared with 40% for those with 9-12 years of education and almost no increase at all for the less educated workers, see García-Huidobro (1999).

Another observation made by García-Huidobro (1999) is the increased proportion of occupations that demand levels of education higher or equal to the average level of education of Chilean workers. For instance, the occupations that reported the largest annual growth were the following: clerical workers (6.8%), transport workers (5.9%), qualified operators (4.6%), professional and technicians (4.2%). The role that the occupational structure has on the level of earnings inequality and the determinants of inequality within different occupational groups is an analysis that has been neglected in the Chilean literature. This is surprising since some studies have put forward that analysing long series of data, the wage inequality of white-collar workers is the main component of the overall wage inequality which follows closely the behaviour of the inequality of this group of workers.

Table 1
Social and Economic Indicators, 1987-2000

Year	GDP / Capita	Growth	Unem− ployment	Primary Sector Employment	Secondary Sector Employment	Therdiary Sector Employment	Average Schooling	Household Poverty	Gini Houseolds Income	Gini Wage/salary Workers	Gini Own-account Workers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1987	292	4.8	10,9	22.9	21.1	56.0	9.3	39.4	56.8	n.a.	n.a.
1990	342	2.0	7,8	21.8	23.0	55.1	9.7	33.3	55.9	48.9	57.3
1992	401	10.4	6,6	20.2	24.7	55.1	9.7	27.7	55.7	48.1	51.0
1994	439	4.0	7,8	18.4	24.2	57.4	10.0	23.2	55.0	48.8	56.8
1996	506	5.9	6,4	17.1	24.9	58.0	10.2	19.7	56.2	50.9	54.7
1998	550	2.5	6,3	15.9	24.0	60.0	10.4	17.8	56.8	49.2	53.6
2000	558	4.0	9,2	15.7	22.1	62.2	10.6	16.6	56.9	49.5	53.0

Source: (1)-(2), (3) Banco Central de Chile and (4)-(6) Chile Social and Economic Indicators 1960-200, (8) MIDEPLAN (2001), others are own calculations.
Notes: (1) Thousands of Chilean pesos of 1986, (3) Average unemployment that year, (5)-(7) Percentage of respective sector employment on total employment, (8) Household Poverty, (9) Household per Capita Income, (10)-(11) Hourly Earnings.

The wage inequality of blue-collar workers, on the other hand, has been much lower and more stable than that of white-collar workers. This suggests that understanding the behaviour of white-collar inequality is of crucial importance to understand the behaviour of overall inequality in the Chilean labour market.

As in many developing countries, the Chilean labour market is characterized by a relatively high percentage of individuals working without a formal contract. Either because they work in the informal sector or because they work in the self-employment sector. Employment in the informal sector and in some segments of the self-employment sector is associated with several socio-economic disadvantages as no job security, lower earnings, and bad job conditions. Therefore a closer study of the workers that enter this segment of the labour market is an important component to understand economic well-being of Chilean households. An increased percentage of informal employment among wage/salary workers is also a pattern that characterizes the Chilean labour market during the 1990s. Amuedo-Dorantes (2003) suggests that between 1990 and 2000 the percentage of informal male wage/salary workers increased from 10% to 18%. The increment among female workers was even higher, rising from 12% to 26%, over the same period. Self-employed workers, on the other hand, according to CASEN data has been relatively stable at 24% during the 1990s.

This paper attempts to take into account several of the patterns outlined above analysing a more rich structure of the labour market than has been used in other studies of Chilean income inequality. We analyse 12 different occupational groups defined according to sector of employment and occupation. In a first step we classify individuals as working in the self-employment sector, in the informal sector, or in the formal sector. In a second step the individuals of each of these groups were classify as white-collar, clerical and sales, blue-collar, or manual.

The questions that we aim to answer in this paper are two: Firstly, which are the principal individual characteristics that influence the choice between the 12 different occupations in our study? Secondly, which are the main variables that explain the inequality of earning within these groups? What differ our study from others performed on Chilean data is the use of an approach that until now has been found in just a few studies of the labour market, but is a powerful tool in models with more complex choice structure, namely the nested logit model. Other examples where this model has been used in a labour market context is found in Falaris (1987), Hagstrom (1996), and Soopramanien and Johnes (2001), but none of them is used in an income inequality context.

Our results suggest that there are different degrees of substitutability between occupations in the different sectors in our study but is highest in the formal sector, followed by the informal and the self-employment sector. Furthermore, comparing 2000 with 1992 we observe a lower level of earnings inequality in the formal sector where all groups had a lower Gini or Variance of log earnings at the end of the period. In the informal and in the self-employment sector, on the other hand, the picture is more variated but tended to increase in high-level occupations and tended to decline in lower-level occupations.

Using the estimates of the earning regression and the inequality decomposition of Fields and Yoo (2000), we are able to explain 44% of the inequality of earnings within high-level self-employed workers in 2000, 38% among their informal counterparts, and 33% within formal white-collar workers. Formal education, measured by years of schooling, is found to be a key variable to explain the inequality in these occupations accounting for 30%, 32%, and 21%, respectively.

Looking at the earnings inequality of workers of lower-level occupations, we are able to explain 38% in the self-employment sector, 41% in the informal sector, and 25% in the formal sector. Among these workers formal education explains a lower share of their earnings inequality, being higher among individuals of the formal sector. On the other hand, hours of work is a more important variable contributing to explain 26% of the earnings inequality among self-employed manual workers, 28% among manual informal workers, but only 5% of the earnings inequality of their counterparts in the formal sector.

The structure of the paper is the following. In section 2 we provide a summary of previous studies on income inequality in Chile. Section 3 presents the empirical model. In section 4, the data and the different categories used in the paper are presented. Section 5 presents the estimates of the nested logit model. Section 6 presents the estimates of the earnings equations. In section 7 we report to which extent are we able to explain inequality within groups. Section 8 provides a summary and suggestions for further research.

2 Previous Results on Education and Income Inequality

This section summarizes the findings of several previous studies on income inequality in the Chilean labour market. The literature recognizes at least two distributional effects of an increase in the level of education. The first is the composition or Kuznets effect that arises as the percentage of high educated workers, and thereby also high income earners, increases generating an increase also in total inequality. However, inequality eventually decreases as the percentage of workers with low education, and thereby also low incomes, become small. The second effect arises as an increase in the relative supply of high educated workers decreases the education premium creating a compression of the between-group inequality.

Contreras (2002) and Contreras (2003) analysed the inequality of salaries, respective earnings, within the well-known Mincerian model of human capital and applying the inequality decomposition of Fields and Wo (2000). The results of Contreras suggest that the percentage of workers with primary education declined and the percentage of workers with 12 or 13-16 years of schooling increased in most periods. Both these groups increased their percentages of full time workers to approximately 3 times their value in the late 1950s. High-educated workers represented 4.1% of the total male full time workers in the late 1950s, increasing to over 13% in the 1990s.

Comparing how the educational structure evolved over the years with how the rate of return to education evolved over the same period reveals interesting results. Average values found by Contreras (2002), ranging between 0.12 and 0.15, are in line with, for instance, the study of Riveros (1990), who used the same survey but analysed the period 1960-1985.[2] The data reported discern relatively small changes in the rate of return to primary education from the late 1950s to the late 1970s. A larger change is found, roughly 20% downwards, in the early 1980s. After that, this rate of return started to increase, reaching in the last period a level similar to that at the start of the research period. More dramatic changes are reported by the rate of return to secondary education, which decreased between 1958 and 1975, increased in the following two periods, but experienced a significant drop in the late 1980s and in the early 1990s. This pattern implied that in the 1990s, the return to secondary education was less than 1/4 of the 1958-1965 value.

Return to university education, on the other hand, increased in most periods, except in the period 1971-1975 and in the last period, implying a 50% higher value in the last period compared with 1958-1965. Notice that the pattern followed by high education and its rate of return is somewhat unexpected. As the relative supply of university educated workers (represented by the group with more than 17 years of education) increased over periods, we expect the rate of return of this level of education to decrease, if demand is constant. Conversely, this rate of return increased in most periods, especially between 1976 and 1990. Thus, a compression effect seems to have occurred for secondary educated workers, but not for high-level education.

In the Table 2 below we found the share of the inequality of log earnings explained by years of schooling calculated by Contreras (2002), column (1). In the following three columns, we found the percentage of the share of the inequality, explained by years of schooling, explained by the rate of return to primary, secondary, and university education, respectively.

---

[2] In an international perspective, Heckman and Hotz (1986) report the following rates of return: Africa 0.134; Asia 0.128; Latin America 0.182; LDC average 0.144; Intermediate 0.097; Advanced 0.077.

Table 2
Contribution of Years Schooling to Explain the Inequality of Salaries

Period	Schooling (1)	Primary (2)	Secondary (3)	University (4)
1958-1965	0.3800	27%	45%	28%
1966-1970	0.3980	26%	35%	40%
1971-1975	0.3420	27%	30%	42%
1976-1980	0.4260	24%	30%	46%
1981-1985	0.3860	18%	31%	51%
1986-1990	0.3920	19%	18%	63%
1991-1996	0.3283	22%	8%	70%

Source: Contreras (2002).

The calculations of Contreras suggest that among observable factors, schooling is the most important, explaining roughly 40% in most periods, but being somewhat lower in 1971-1975 (0.3420), and in 1991-1996 (0.3283). Of these shares, university education was not only the major contributor from 1966-1970 and onwards; the contribution of this level of education experienced a trend upwards reporting a percentage in 1991-1996 that was more than two times larger than the share in the beginning of the research period. The opposite pattern was reported by secondary education, which decreases or was constant in most periods, being in the last period less than 1/5 of the percentage explained by this level of education in 1958-1965.

In the other work by Contreras, Contreras (2003), the 1990 and the 1996 versions of CASEN and a broader sample of individuals is used. In this case the dependent variable is earning. Since this type of income form the bulk of household income, his work also gives an indication on how powerful education is to explain income inequality among households. Contreras not only controls for schooling but also for experience, gender, level of participation in the labour market, self-employment, and sector of occupation. He estimated the rate of return to schooling to 0.10, which is in line with the results of, for instance, the Mincerian regression of Arellano and Braun (1999). Contreras decomposition reveals that, among the observable variables, schooling is the most important factor to explain inequality in both years accounting for 18% of the 1990 inequality, increasing to 21% in 1996. The contribution of other observable variables is much smaller: for instance, the variable that denotes if the worker is self-employed explains 7% in 1990 and 9% in 1996, while occupation accounts for 6% in 1990, compared with 5% in 1996. Another approach to inequality decomposition is the method we found in Larrañaga (1999). Using CASEN data from 1987-1996 Larrañaga found that the vast majority of the inequality is explained within the different cohorts, when the sample was classified according to gender, age, sector of employment, or region. The important exception is education, in which differences across educational cohorts account for 22.1% of the total income inequality. Compare this with the proportion explained by economic sector, 7.6%, and the share explained by age, 5.9%. The contributions of gender and region are even minor.

3 The Model

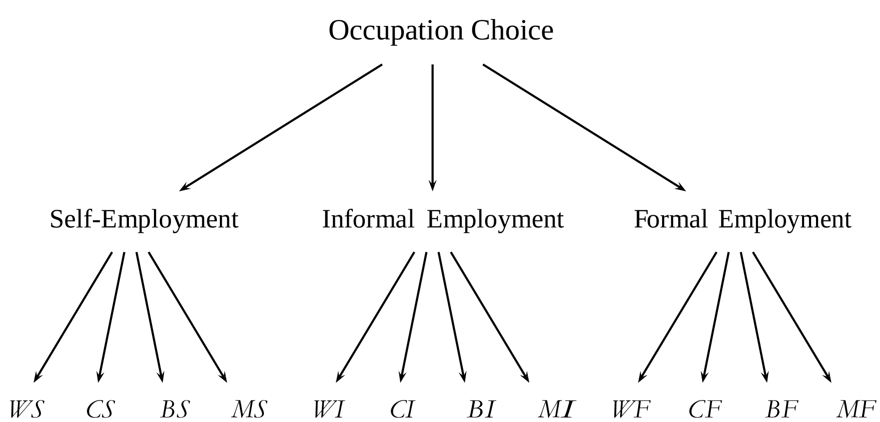
It is a common practice in the estimation of wage equations, when the sample for which the wage is observed is not a random sample of the population we aim to study, to include a variable that contains information on the sample selection process. When the wage equation is estimated for different segments of the economy, commonly a multinomial logit model is used in the estimation of the segment assignment equation; see for instance Tiefenthaler (1994). However, an important disadvantage of the multinomial logit model is that it could be too restrictive as the model demand the assumption of independence of irrelevant alternative (IIA) to be fulfilled. IIA arises from the assumption of the extreme value distribution of the error terms and implies that the ratio of the probability of 2 different alternatives is completely independent of the other alternatives in the choice set. However, there are several empirical applications where the IIA may be an inappropriate assumption. This may be the case when some alternatives of the choice set are more closely substitute to each other than other alternatives. For instance, in studies on the choice between different types of dwelling situations, substitutability between dwelling size may be higher than between tenure category (owning or renting), see for instance Borsch-Supan (1987). The degree of substitutability among different calling patterns within a service (basic or Optional Calling Plan) may be different from that between services options within a calling pattern; see Lee (1999).

In the model of Soopramanien and Jones (2001) substitutability is assumed to be different among occupations within the group of part time workers than among full time workers. This implies that part-time individuals with a certain occupation more readily move to part-time work in other occupations, than to full-time work in the same occupation. Our model suggests that there are different degrees of substitutability between occupations in the three different sectors in our study, that is the self-employment sector, the informal employment sector, and the formal employment sector. We believe that this may be a correct assumption given the characteristics of the individuals we expect to find in respective sector. If the individuals that we found in the self-employment and informal sector have stronger preferences for flexible working schedules that are possible to find in the formal sector. Workers in these sectors may be reluctant to move to occupations in the formal sectors and prefer to stay in the self-employment or informal sector where they may have the possibility to govern their working schedule. In this case an unobserved component of the utility function reflecting the preference for flexible working schedule is correlated within the self-employment and the informal sector.

When some alternatives of the choice set are closer substitutes than others, the model can be estimated by means of the nested logit model, which is less computational demanding than the multinomial probit but more flexible than the multinomial logit. The name of the model arises from its structure, as it allows a partition of the choice set into groups or 'nests' of alternatives that are similar to each other in a unobserved way. The advantage of the nested logit model is better understood when one alternative is no longer available in the choice set. According to the IIA assumption the relative probability of choosing between two different alternatives is totally independent of the other elements in the choice set. Suppose now that a third alternative of the choice set is no longer available. The multinomial logit model suggests that the relative probability of the two alternatives remain unchanged. But what happens if the removed alternative is close substitute to one of our two initial alternatives? We should expect that individuals from the now no longer available alternative would now prefer the close substitute instead of other alternatives available. This should affect the relative probability, which implies that the IIA is not fulfilled. The nested logit model avoids the IIA and hence allows a richer pattern of substitutions between alternatives. The structure of the nested logit model we use in our model is the following: first denote the choice set as *O*. We have three different sectors; the self-employment sector, *S*, the informal employment sector, *I*, and the formal employment sector, *F*. Each one of these sectors is made up of white-collars, clerical and sales, blue- collar, and manual workers. These occupations are denoted as *W*, *C*, *B*, and *M* respectively. In this way we have the following structure: $S = \{WS, CS, BS, MS\}$, $I = \{WI, CI, BI, MI\}$, and $F = \{WF, CF, BF, MF\}$. Graphically we have

Figure 1

Decision Tree of the Occupation Choice



Let U_{jm} denotes the utility experienced by an individual in sector m and occupation j.[3] This utility is made up of one systematic part and one random part unobserved to the econometrician:

[3] We omit the index for the individual.

$$U_{jm} = V_{jm} + \varepsilon_{jm}, \ \forall\, jm \in O$$

Status jm is chosen only if $U_{jm} > max\, U_{kl} \ \forall\, jm \neq kl \in O$. When the unobservable part of the utility function ε_{jm} is assumed to follow a generalized extreme value distribution (GEV), we can see the choice between the different alternatives as composed of two different parts, one is the sector employment decision, and the other is the occupation decision. In our empirical model

$$V_{jm} = \beta'_m Y_m + \gamma'_{jm} H_{jm} + \delta'_{jm} Z_{jm}$$

where Y_m represents the variables that affect the desirability for sector m, H_{jm} represent human capital variables affecting desirability for occupation jm, Z_{jm} represent other variables affecting desirability for occupation jm. For instance, the probability of choosing to be white-collar worker in the formal employment sector is given according to the nested logit in the following way:

$$P_{WF} = P_F \cdot P_{W|F}$$

$$P_F = \frac{exp(\beta'_F Y_F + \mu_F I_F)}{\sum_{m' \in \{S,I,F\}} exp(\beta'_{m'} Y_{m'} + \mu_{m'} I_{m'})}$$

$$P_{W|F} = \frac{exp(\gamma'_{W|F} H_{W|F} + \delta'_{W|F} Z_{W|F} / \mu_F)}{\sum_{j' \in \{WF,CF,BF,MF\}} exp(\gamma'_{j'} H_{j'} + \delta'_{j'} Z_{j'} / \mu_F)}$$

The first part of the right hand side is the probability of choosing nest F, while the second part is the probability of choosing alternative W, given F, and

$$I_F = ln(\sum_{j' \in \{WF, CF, BF, MF\}} exp(\gamma'_{j'} H_{j'} + \delta'_{j'} Z_{j'} / \mu_F))$$

is the sector *F* inclusive value which summarizes the attractiveness of the formal employment sector. The parameters μ_m are called inclusive value parameters and give several pieces of important information. At the one hand, they can be used to test if the model is misspecified. They should lie in the interval $0 < \mu_m \leq 1$ in order to be consistent with the random utility maximization, McFadden (1981). However, some researchers argue that requiring this condition to be fulfilled for all possible values of the observed data may be too restrictive, see for instance Borsch-Supan (1990). Several researchers have developed conditions that allow inclusive values parameters greater than one.

Hauber and Parsons (2000) summarize the findings on this issue. However, they suggest that research point out that $1 < \mu_m < 2$ will in most cases fail to fulfil the condition that make the model still consistent with utility maximization. This problem arises in the model of Hausman et al. (1995). They suggest however that the interpretation of the parameter in this case is that there is a greater correlation among the utility of the elements of different nests than the correlation of the utility of the elements within the nest. On the other hand, inclusive value parameters give an estimate of the degree of dissimilarity of the choices within the different nest. That is, the lower its value, the higher is the degree of substitutability of the elements within a nest.

In the following step we estimate the earnings equation for the 12 occupational groups in our study. When individuals are not randomly assigned to the different groups, OLS estimation will generate biased estimates of the population parameters. This problem is solved by introducing an additional variable that contains information on the sample selection into the different groups of the labour market in the earnings equation. This is done by using the information obtained in the estimation of the nested logit model to construct a sample selection correction variable that takes the form:

$$\lambda_{jm} = -\frac{\theta(\Phi^{-1}(P_{jm}))}{P_{jm}}$$

where Φ and θ represent the standard univariate normal density and distribution functions, respectively. Using this variable in the earnings equation we get

$$log\ E_{jm} = \alpha'_{jm} X_{jm} + \eta_{jm} \cdot \lambda_{jm} + u_{jm}$$

Where E_{jm} is earning in sector m and occupation j, and X_{jm} is a vector of variables that explain earnings. Estimating this model by OLS we obtain consistent estimates of the earnings equation. The ultimate goal of our model is to shed light into the factors that affect earnings inequality, especially how the distribution of years of schooling and the parameter of schooling in the Mincerian regression, contributes to explain the variation of earnings within occupational groups. We have found several studies that uses selectivity corrected earnings regressions across different occupations, see for instance Yuhong and Johnes (2003). There are also other studies that have used the nested logit model to correct for selectivity bias, see for instance Falaris (1987). Our purpose is to follow his approach but in contrast to Falaris, where the decisions are related to migration, in our model the decision is related to sector employment and occupation.

The method we use to analyse the contribution of education to explain the level of inequality is the method of Fields and Yoo (2000) which has been used in the Chilean case by Contreras (2002), Contreras (2003), and Amuedo-Dorantes (2004). The advantage of this decomposition is that is has stronger theoretical fundaments since is based on an income generating model based in the well-known human capital model. Starting from a Mincerian regression and applying the inequality decomposition of Fields and Yoo (2000), the share of the inequality of earnings that is explained by, for instance, variable ℓ given by the expression

$$s_{jm}^{\ell} = \frac{\alpha_{jm}^{\ell} \cdot \sigma_{jm}(\ell) \cdot \rho_{jm}(\ell, log\, E)}{\sigma_{jm}(log\, E)} \tag{1}$$

where $\sigma_{jm}(\ell)$ is the standard deviation of variable ℓ, $\sigma_{jm}\,(log\, E)$ is the standard deviation of log earnings, $\rho_{jm}\,(\ell,\, log\, E)$ is the correlation between ℓ and log earnings, and α^{ℓ}_{jm} is the parameter of variable ℓ in the Mincerian regression for sector m and occupation j.

4 Data and Definition of the Different Categories

The individuals that we use in this study are the household heads of the household included in the CASEN survey of 1992 and 2000. CASEN is a cross-sectional household survey that has been conducted in Chile approximately every two years since 1987. The aim of this survey is to collect detailed information on several socioeconomic variables of Chilean households such as health, housing, education, employment and income. The sample of households included in the survey is selected by multistage stratification and is of national coverage. CASEN includes also an expansion factor to take into account that the sample was taken using a multistage method and interpreted as the number of households that a specific household in the sample represents in the population. The number of households included in CASEN increased from 22,719 in 1987 to 65,036 in 2000, while the number of individuals increased from 97,044 in 1987 to 252,748 in the 2000 survey. In our study we only include household heads older than 15 years of age but younger than 65, that are working at the time of the survey. We exclude those individuals that report a missing value in at least one of the variables included in the model. Including only household heads we reduce the eventual correlation that may exists in the occupational choice of the individuals within a household. We abstract from the labour force participation decision in order to reduce the computational burden. However, we are aware of the importance of this aspect and aim to study this issue in a following work.

Traditionally, the definition of the informal sector is associated to the size of the firm, the occupation of the worker, and technology employed; see Saavedra and Chong (1999). Using these criteria wage/salary workers without a contract and non-professional self-employed are traditionally classified as informal workers. The rest is classified as formal. We use instead a larger number of segments including 12 different categories instead of the simple formal-informal classification.

First we classify individuals as self-employed, formal, or informal. Our definition of self-employment includes employers, that is, individuals that operate his/her own economic enterprise or engaged independently in a profession/trade and hires one or more employees. It includes also own account workers which fulfilled the definition in the previous lines but without hiring any employee. As Amuedo-Dorante (2004) we follow the definition of informal employment where the focus rests on the contract of the worker. We define as informal all wage/salary workers without any type of written contract.

The reason to have self-employed as a separate group, and not making them a part of the informal sector, is that the allocation of paid work between self-employment and wage/salary employment has attracted increased interest in the international literature in recent years. The reason is that this group has become an important group of workers in developed countries labour market and therefore there is an increased need to understand the functioning of this group. In developing countries the percentage of paid workers in the self-employment sector tends to be higher than in developed countries. In the Chilean case the importance of this category is reflected by the data found in Garcia-Huidobro suggesting that self-employment, excluding employers, contributed with 393 thousands of new jobs in 1996 compared with 1986 being the fastest growing category during this period, engaging 24% of the occupied labour force in 1996.

Further, individuals in each sector are classified as white-collar which includes managers and professionals; clerical and sale workers; blue-collar, which include technicians and similar; and workers without any qualification which we denote with the name of manual workers. In this way we obtain a set of 12 different occupations characterized by sector and type of occupation.

Table 3 reports the occupational structure of household heads in CASEN 1992 and in CASEN 2000. In 1992, within the self-employment sector and the formal sector, blue-collar workers is the largest occupational group.

Table 3
Occupational Structure 1992-2000

Occupation	1992 Total	1992 $S_{Employment}$	2000 Total	2000 $S_{Employment}$	Ratio Occupation	Ratio Sector
	(1)	(2)	(3)	(4)	(3)/(1)	(6)
Self-Employment						
WS	162	6,93	240	9,81	1,49	
CS	92	3,95	65	2,68	0,71	
BS	337	14,44	314	12,82	0,93	
MS	112	4,79	62	2,53	0,55	0.97
Informal Employment						
WI	9	0,38	29	1,20	3,33	
CI	19	0,83	35	1,44	1,82	
BI	66	2,83	128	5,23	1,94	
MI	96	4,12	120	4,90	1,25	1.64
Formal Employment						
WF	289	12,38	369	15,10	1,28	
CF	241	10,34	281	11,49	1,16	
BF	553	23,71	520	21,28	0,94	
MF	357	15,29	282	11,52	0,79	1.01
Total	2332	100	2445	100		

Source: Own calculations from CASEN using expansion factors.

In the informal is the manual occupation that engages the largest share of workers. Comparing 2000 with 1992 we found that white-collar workers was the group that reports the largest increment, increasing by 49% among self-employed, 230% among informal, and 29% among formal workers. Thus, our results suggest that high-level occupations become an important segment of the Chilean labour market engaging 19% of the household heads in 1992, but increasing to 26% in 2000. The percentage of household heads with clerical and sales occupation stayed almost constant between 1992 and 2000 at 15%. On the other hand, the two lower occupations engaged less household heads in 2000, decreasing from 41% to 39%, and from 24% to 19%, respectively.

As a whole the informal sector was the sectored that reports the largest increment, increasing by 64%; while the other two sectors remained virtually constant. This suggests an increment of 122,125 households head without a labour contract, comparing 2000 with 1992. There are suggestions in the literature that informal employment and self-employment are used to avoid high severance payments and to avoid the difficulty of terminating indefinite contracts in Chile. This may explain that when the labour market expands a larger proportion of the new jobs are found in the informal employment sector. However, why don't we observe also a larger percentage of employment in the self-employment sector? An alternative explanation to the pattern found in Table 3 is that wage and salary workers that aim to work less than full time prefer an employment in the informal sector where more flexible working schedules are possible. To see if this is a plausible explanation we report the percentage of household heads in respective occupational group that works less than 35 hours per week. Table 4 suggests that part-time employment is rare to find in the formal employment sector were at most 7% of the household heads worked less than 35 hours in 1992. In the other two sectors, part-time employment is more common, especially among self-employed. In 2000, larger percentages of part-time workers are found in all sectors but the increment seems to be larger among informal workers where between 14%-26% worked less than 35 hours per week. In the other two sectors part time work engaged between 14.54% and 34.16% among self-employed and between 5.06% and 8.36% among formal workers. This may suggests that sector choice is closely related to the choice of part time work.

Table 4
Percentage of Part-time Work in Each Occupational Group
(Part-time Work is < 35 hours)

Occupation	1992 (1)	2000 (2)	Ratio (2)/(1)
Self-Employment			
WS	11.03	14.54	1.32
CS	13.78	27.14	1.97
BS	13.73	22.73	1.66
MS	26.23	34.16	1.30
Informal Employment			
WI	9.12	23.79	2.61
CI	15.72	20.94	1.33
BI	6.17	14.81	2.40
MI	12.02	26.05	2.17
Formal Employment			
WF	6.96	8.36	1.20
CF	2.31	4.46	1.93
BF	1.79	5.06	2.83
MF	4.97	5.67	1.14

Source: Own calculations from CASEN using expansion factors.

A potential explanation to the downturn of low-level occupations is how the industry employment structure evolved between the two years in our study. Comparing data from the two years we found four sectors that report a clear trend. In the agricultural-hunting-fishing sector employment decreased from 18.49% in 1992 to 14.49% in 2000; in industry from 16.98% to 14.1%. On the other hand, it increased in financial services, from 5.07% to 7.91%, and in social and communal services, from 25.37% to 27.77%. The pattern reported by these sectors is in line with the pattern reported by the occupations found in Table 3 and expected to be important in these sectors. Thus, the decreased importance of agricultural-hunting-fishing and industry may have induced lower job opportunities for blue-collar and manual workers. The opposite should have occurred with the increased share of financial services and social and communal services and job opportunities for white-collar workers.

Further information is reported in Table 5 which reports the share of each group on total earnings, S_{Income}, the Gini coefficient of earnings within each group, *Gini*, and the variance of log earnings within each group.

Table 5
Share on Total Employment, Share on Total Income, and Gini for the Different Employment Status

Occupation	1992			2000		
	S_{Income}	*Gini*	*Variance of log Earnings*	S_{Income}	*Gini*	*Variance of log Earnings*
	(1)	(2)	(3)	(4)	(5)	(6)
Self-Employment						
WS	25.95	0.583	1.298	30.58	0.587	1.281
CS	4.82	0.492	0.782	2.08	0.469	0.799
BS	14.11	0.466	0.711	9.67	0.411	0.669
MS	2.79	0.411	0.659	1.09	0.378	0.617
Informal Employment						
WI	0.48	0.413	0.664	1.26	0.450	0.757
CI	0.35	0.359	0.385	0.57	0.373	0.530
BI	1.17	0.330	0.389	1.79	0.295	0.348
MI	1.23	0.283	0.342	1.18	0.273	0.342
Formal Employment						
WF	22.85	0.452	0.689	28.99	0.445	0.590
CF	6.86	0.376	0.443	7.31	0.329	0.327
BF	13.74	0.334	0.349	11.34	0.307	0.285
MF	5.64	0.255	0.245	4.14	0.229	0.165

Source: Own calculations from CASEN using expansion factors.
Notes: (1) and (4) Share of each group on total earnings, (2) and (5) Gini of earnings.

The share of respective occupational group on total income is largest for self-employed with high-level occupations. In most cases the share of this group on total income is three times larger than its share on total employment. A similar picture is reported by high-level formal workers but in this case its share on total income is approximately two times its share on total employment. Among mid-level groups, their share on total income is in most cases lower than their share on total employment. On the other hand, manual workers report a much smaller share of income than of employment. For instance in year 2000, the ratio between its share on total income and on total employment, $S_{Income}/S_{Employment}$, was 0.28 for informal manual workers but 0.52 for formal manual workers, and 0.63 for self-employed manual workers.

Table 5 also shows that, comparing within sectors, the level of inequality is clearly related to the type of occupation, being highest among with-collar workers followed by clerical and sales workers and blue-collar workers. Inequality is clearly lowest among manual workers for whom inequality, in many cases, it was half the inequality of white-collar workers. The level of inequality was higher in the self-employment sector compared with their counterparts in the other two sectors. Comparing 2000 with 1992 we notice a decline in the inequality in the formal sector where all groups rapport a lower dispersion of earnings, independently of which measure of inequality is used. In the other two sectors the picture is more variated but inequality tended to increase in higher-level occupations and tended to decline in lower-level occupations. So we conclude that it was a tendency to a lower level of earnings inequality in the research period, especially in the formal sector and in lower-level occupations.

5 Model Specification and Estimates of the Nested Logit Model

It is not easy to identify the variables that should enter the vector Y_m since there is a rich literature that handles the determinant of informal employment, but the suggestions on the variables that affect the desirability for self-employment is rare to find in the literature.

Amuedo-Dorante (2004) summarizes the two views of the determinants of informal sector employment. On the one hand, we have the hypothesis of voluntary employment, which suggests that individual preference for employment in a certain sector of the economy is completely based on the comparison of the expected earnings in the different sectors. The individual then chooses the sector that offers the highest earning, given the individual characteristics, and thereby the highest utility. This is the approach found in many studies, for instance, Tiefenthaler (1994) and Gindling (1991). According to this approach, human capital variables, such as education and experience, should be included to work as proxies for the offer wage across sectors.

The other, is the non-voluntary employment hypothesis, which sees employment in this sector as much less attractive than in the formal, attracting individuals with no other better options, even if this sector do not offer the highest earning. This is the working approach used by Amuedo-Dorante (2004), where she finds evidence that poverty increases the probability to work in the informal sector in Chile. Therefore, variables that function as proxies for poverty should enter the model. In our empirical model we include a dummy that takes the value one if there are unemployed adults in the household. In the same way we include a dummy that takes value one if there are inactive adults in the household.

The variables that the literature suggests to affect the probability of entering the self-employment sector are the following according to Le (1999). Education has two conflicting effects on the self-employment choice. The first, education improves the individual's managerial ability and thereby increases the propensity to be engaged in the self-employment sector. The second, a higher level of education improves the employment possibilities in the wage/salary sector and thereby reduces the propensity of self-employment work. Labour market experiences have been hypothesized to increase the propensity to self-employment work. The idea is that labour market experience is a proxy for accumulated general knowledge to understand the functioning of markets. Thus, using labour market experience we will be able to test the hypothesis that greater managerial and learning abilities will increase the propensity to self-employment work. However, it has alternatively been suggested that longer labour market experience implies a longer period to accumulate the necessary financial resources to enter the self-employment sector. In this study we use age as a proxy for labour market experience.

Family background such as the number of children and marital status has also been suggested to affect the self-employment decision. Married persons have the possibility to incorporate the spouse in the business, thereby reducing eventual shirking problems when non-family members are employed. Family economic conditions are also included in our model. Since self-employment is a risky project, any factor that increases the economic instability of the household may reduce the propensity to self-employment work. This is the case if there are unemployed or inactive individuals in the household that may discourage individuals to enter a risky project. On the other hand, a higher household income or owning a house may provide a security against the risky income that is associated with self-employment.

The results of the estimation reveals that at the upper level model, where formal employment is the omitted sector, the effect of years of schooling is significant for both years. The negative value for self-employment and informal employment suggests that more years of formal schooling decrease the probability of entering these sectors. This result is in line with the hypothesis that education increases the employment possibilities in the formal sector and thereby reduces the attractiveness of the self-employment sector and the informal sector. The effect of age is variated, and in most cases not significant, with exception of older workers in the informal sector. For these workers a higher age had a negative and significant effect. Males are more likely to work as self-employed in the model of 1992. In the model of 2000 this effect is positive but not significant. This result may indicate that males have an advantage in the access to the financial market to borrow the capital needed to start a business. On the other hand, males are less likely to work in the informal sector. This effect is significant in both years. An interpretation of this is that employment in the informal sector is primarily an option for female workers. Living in rural areas increases the probability to work in the informal sector. This result is expected since most probably is in these areas, and in relation to the agricultural sector, that an important part of informal activities are concentrated. The dummy variable that takes value one if the individual is married is also significant and negative for both self-employed and informal workers. The presence of a baby or a small child seems to have a significant effect on the probability of entering the informal sector. This result may indicate be the effect that female workers attempt to combine the care of children with participation in the labour market since the participation of female workers tend to be higher in the informal sector than in the other two sectors. To have unemployed adults in the household have a significant and negative effect on the probability of entering the self-employment sector, but have a significant and positive effect on entering informal employment. The effect of having inactive adults is negative and significant for self-employment and informal work. This result may indicate that in households where there is an unemployed individual the household head tend to avoid of entering a risky activity as self-employed and that the household head with an unemployed individual in the household may see informal employment as an alternative option if no employment is found in the formal sector.

The inclusive value parameters (μ_m) were estimated to 0.8546, 0.3603, and 0.5482 in 1992. In the model of 2000 their values were instead 1.2684, 0.7605, and 0.4633. In other words, the degree of substitutability within sectors decreased in the self-employment sector and in the informal sector but increased in the formal sector. That $\mu_s >$ 1 in the self-employment sector in 2000 indicates that household heads saw occupations in other sectors as closer substitutes to similar occupations in the self-employment sector than other occupations within the self-employment sector. The higher level of unemployment of that year may have driven this change. As unemployment increased, some self-employed workers started to look for opportunities as employees in other sectors instead of being driving their own business.

At the lower level model, the omitted alternative is manual occupation within each sector so we interpret the parameter as the effect of respective variable on the probability of entering the different occupations with respect to manual occupations. As it is expected, individuals with more years of schooling are more likely to work as a white-collar worker, but this effect is highest in the formal employment sector, followed by the occupations in the informal employment sector, and in the self-employment sector. The effect of the male-variable is variated but in all cases is positive for blue-collar workers and in most cases negative for clerical workers. This may reflects a strong segmentation between occupations with respect to gender in the Chilean labour market.

The effect of residence in regions other than Santiago is positive and significant for blue-collar workers in the self-employment sector and in the informal sector in 2000. In 1992, on the other hand, its effect is positive only for blue-collars in the self-employment sector although is not significant. Also the effect of living in rural areas is positive for this group in 1992. In 2000 its effect is even positive for withe-collar in the self-employment sector. For all other groups the effect of this variable is negative. The presence of children of different ages in most cases decreases the probability of working in high-level occupations with respect to manual workers.

As measure of goodness-of-fit we use the $\text{Pseudo-}R^2 = 1 - \ln L_n / \ln L_0$ where $\ln L_n$ is the value of the likelihood function of the nested logit model, and $\ln L_0$ is the value of the likelihood function when only a constant enters the model. Using this definition we calculate the $\text{Pseudo-}R^2$ for 1992 to 0.264 and for 2000 to 0.253. Table 6 reports the percentage of correctly predicted occupations. The model performs quite well workers in the formal sector, but performers quite badly for the other two, especially for those in the informal sector. We also report the percent of correctly predicted observation using a multinomial model, that is, using the restriction μ_m:s = 1. From this we draw the conclusion that from a prediction point of view the two models are identical. However, the nested logit provide additional information, relaxing the restriction on the μ_m:s. A preliminary estimation reveals that the selection variable is highly correlated with years of schooling in the earning regressions of high-level occupations. To solve this problem we estimate an alternative specification where instead of years of schooling of the individual we use the average years of schooling of the household for those alder than 15 years, excluding the household head. The model performs somewhat worse but we reduce the multicollinearity problem.

Table 6
Percentage of Correctly Predicted Occupations

	Nested Logit (1)		Multinomial Logit[1] (2)		Multinomial Logit[2] (3)	
Occupation	1992	2000	1992	2000	1992	2000
Self-Employment						
WS	8.09	8.02	9.02	11.07	5.95	7.28
CS	0.45	0.93	1.40	0.89	2.23	2.48
BS	25.52	19.32	25.91	20.08	24.06	16.80
MS	0.11	0.34	0.11	0.00	1.45	1.40
Informal Employment						
WI	0.00	0.00	0.00	0.00	0.00	0.64
CI	0.00	0.00	0.00	0.00	0.04	5.14
BI	0.00	0.00	0.00	0.00	0.05	0.40
MI	0.62	12.09	0.83	14.52	0.60	12.38
Formal Employment						
WF	66.00	73.91	65.40	72.77	51.50	55.22
CF	11.68	18.89	11.40	18.52	5.06	8.54
BF	74.10	78.14	73.19	76.98	70.57	73.61
MF	30.90	10.17	30.98	10.01	29.28	10.44
Total	35.97	35.02	35.83	35.05	32.16	29.87

Notes: (2) Years of schooling of the individual is used, (3) Average years of schooling of the household is used.

6 Estimates of the Earnings equations

The estimates of the parameters of the earnings equations are reported in the tables of the appendix. We estimate one equation for each of the 12 occupational groups we use in our model, including the selection correction variable, λ_{jm}. The left-hand side variable that we use in the regression is the log of monthly earnings from principal occupation.

According to the work of Mincer (1974), among the regressors we should include years of schooling, a measure of in work experience and its square. Additional variables included are a dummy for male workers to capture the male premium commonly found in empirical studies on the determinants of earnings. We also include a provincial dummy to capture eventual earnings disparities between Santiago and other regions.

The Mincerian specification works better on the data of 2000 when the explanatory power of the model (R^2) ranged between 22% and 44%. The variables that we include in the earnings regressions have the expected sign in most cases in the formal sector; in the other two sectors the picture is more variated. Being male has in most cases a positive, and significant, effect on earnings while living in other regions than Santiago has a negative effect on earnings. The Age variables are significant only in some cases but they indicate that in 2000 the earnings of white-collar self-employed workers peaked at 46 while the earnings of blue-collar formal workers peaked at 55.

The variable of most interest, years of schooling, is significant in most regressions of 1992 with exception of white-collar and manual informal workers. In 2000 it is significant for all groups. The effect of education on earnings is similar comparing individuals from the formal and self-employment sector. Among informal workers, on the other hand, the effect of education is substantially lower than in the other 2 sectors in 1992 but higher in 2000. For instance, in 1992 the effect of education is estimated to 0.092 for high-level formal workers, but 0.025 for high-level informal workers and 0.097 for high-level self-employed compared but 0.129, 0.155, and 0.128, respectively in 2000. These values imply that the private rate of return to education for white-collar workers was most probably between 9% and 13% during this period.

Among low-level occupations the effect of schooling is much lower than compared with other occupations, being the lowest among informal workers. Comparing across years, not only among white-collar workers we found in 2000 a higher effect of education on earnings that it was the case in 1992. Most of the other occupations report the same pattern, especially in the formal and informal sector. We find also that the coefficients on the sample selection correction are significant in both years for the following occupations: white-collar self-employed, formal and informal blue-collar and manual workers. These results indicate that there is a merit in correcting for sample selection bias in these groups. Since these values are important in the calculation of the share of inequality explained by this variable, we expected an increased in the share of inequality explained by education among white-collar workers.

7 How Much Inequality Are We Able to Explain?

In section 4 we concluded that it was a tendency to lower levels of earnings inequality during the 1990s, especially in the formal sector and in lower-level occupations but how important are the variables used in this study to explain earnings inequality within occupational groups? And has the share explained by these variables changed during the 1990s?

Using expression (1) to calculate the share of earnings inequality that is explained by years of schooling we arrive at the results reported in Table 7 and Table 8. In 1992 the share accounted for by years of education ranged from 1% among informal manual workers to 18% among self-employed white-collar workers. Moreover, the results suggest that in most cases, years of schooling explain a larger proportion of the earnings inequality in high-level occupations than among blue-collar or manual workers. In 2000 schooling become even more important to explain the inequality of earnings. Moreover, we discern two important results; the first, the share explained by this variable increased among white-collar workers in all three sectors. The second, the formal sector is the only one where all occupations report a higher share explained by education in 2000 compared with 1992. This implies that during the analysing period, education, due to an increased private rate of return to education, played an important inequality increasing effect within occupations despite that inequality declined in the occupations of this sector. These results partially corroborate the results of previous studies where they found the schooling variable as a key determinant of inequality. However, what it is surprising in our results is the fact that education is an important factor to explain inequality even in relatively homogenous groups of workers.

Table 7
Share Explained by Respective Variable in 1992

Status	Schooling (1)	Age (2)	Male (3)	Province (4)	log Hour (5)	λ_{jm} (6)	**Total** (7)	*Variance of log Earnings* (8)
Self-Employment								
WS	0.181	0.002	0.027	0.004	0.012	0.102	**0.329**	1.298
CS	0.094	-0.001	0.047	0.001	0.045	-0.012	**0.175**	0.782
BS	0.100	0.003	0.042	0.016	0.056	0.023	**0.240**	0.711
MS	0.058	0.018	0.031	0.015	0.096	0.051	**0.270**	0.659
Informal Employment								
WI	0.036	0.042	0.100	0.000	0.004	0.108	**0.291**	0.664
CI	0.106	-0.006	0.017	0.020	0.075	0.002	**0.214**	0.385
BI	0.049	0.001	0.008	0.040	0.035	0.016	**0.149**	0.389
MI	0.011	0.009	0.038	0.034	0.085	0.017	**0.194**	0.342
Formal Employment								
WF	0.141	0.028	0.034	0.050	0.039	0.041	**0.334**	0.689
CF	0.112	0.023	0.042	0.010	0.000	0.012	**0.198**	0.443
BF	0.113	0.024	0.008	0.006	0.005	-0.001	**0.156**	0.349
MF	0.036	0.004	0.045	0.019	0.018	0.051	**0.173**	0.245

Source: Own calculations from CASEN without expansion factors.

Table 8
Share Explained by Respective Variable in 2000

Status	Schooling	Age	Male	Province	log Hour	λ_{jm}	**Total**	*Variance of log Earnings*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Self-Employment								
WS	0.301	0.021	0.034	0.008	0.004	0.074	**0.443**	1.281
CS	0.072	0.009	0.030	0.000	0.129	0.026	**0.265**	0.799
BS	0.126	-0.002	0.046	0.024	0.108	0.024	**0.326**	0.669
MS	0.029	0.007	0.046	0.015	0.259	0.020	**0.377**	0.617
Informal Employment								
WI	0.318	0.030	0.002	0.011	0.029	-0.009	**0.381**	0.757
CI	0.088	0.011	-0.010	0.023	0.201	0.099	**0.411**	0.530
BI	0.077	0.002	0.018	0.042	0.151	0.040	**0.330**	0.348
MI	0.023	-0.001	-0.004	0.066	0.278	0.049	**0.410**	0.342
Formal Employment								
WF	0.209	0.014	0.032	0.022	0.035	0.017	**0.329**	0.590
CF	0.194	0.011	0.028	0.026	0.005	0.001	**0.264**	0.327
BF	0.169	0.004	0.024	0.003	0.013	0.001	**0.216**	0.285
MF	0.063	0.000	0.020	0.043	0.045	0.079	**0.248**	0.165

Source: Own calculations from CASEN without expansion factors.

An explanation of this may be that we have not taken into account in which sector of the economy the individual is employed. If individuals with similar occupations but employed in different sectors of the economy are compensated for their investments in education differently, we should obtain results as we obtained in our calculations.

During the period we analyse in our study the Chilean economy was, on the one hand, more and more oriented towards the external sector. During this period the degree of openness, measured as the sum of exports and imports as a share of GDP reached nearly 60% and the average tariff 9.5% (Palma, 2007). On the other hand, total public spending, which includes among other items, expenditures in governmental activities, defense, infrastructure, health and education increased from near 6,000,000 million of CLP of 2008 in 1990 to 13,000,000 million of CLP of 2008 in 2002 (Palma, 2014). Thus, the demand for some occupations must have been stronger in some sectors of the Chilean economy than in others. Therefore, the high share of education to explain inequality within occupations may have been driven by the uneven growth of the different sectors of the economy during this period. If this hypothesis is correct it leads to two important implications. The first, in order to get a better understanding of the factors that explain inequality within occupations it may be necessary to take into account the sector of the economy where individuals are employed to see whether the share explained by education is important even after sector of employment is accounted for. The second, it may be more difficult than expected to drive policies to reduce the level of inequality in Chile since inequality is high even within relatively homogenous occupational groups. An additional hypothesis that can be drawn from our results is that the stability of the inequality observed in the Chilean economy might be driven by inequality changes within and between occupational groups that are moving in the opposite directions generating small changes at the aggregated level.

Another important variable is log Hour. In 1992 this variable contributed to explain between 1% and 10% in the self-employment sector; between 0% and 9% in the informal sector; and between 0% and 4% in the formal employment sector. Eight years later, the share explained by this variable become substantially higher ranging between 0% and

26%, between 3% and 28%, and between 1% and 5%, respectively. In the self-employment and the informal sector, in occupations other than withe-collars, this variable tended to explain a larger share than schooling and become a key factor to explain the level of earnings inequality. In the formal sector, on the other hand, the contribution of log Hour is low (lower than 6%), and lower than the percentage explained by schooling. An implication of this result is that as the participation of female workers in the labour market increases, the share of this variable to explain inequality should continue to be important as it is most likely female workers that tend to work part-time. Since informal employment tend to be associated with precarious labour conditions, policy makers should promote more flexible working-schedule in the formal sector to improve the working conditions of female workers.

Age plays a minor roll to explain the dispersion of earnings, while the dummy variable use to control for the effect of gender seems to be more important in the self-employment sector and the formal sector. This variable explained some share of the inequality in the occupations of the informal sector in 1992, in 2000 its share almost vanished. Province has a low explanatory power but it is more important in low-level occupations and in the formal sector. Moreover, although the pattern shown by the share explained by this variable variated, it clearly increased for informal manual workers for whom the share explained by this variable explained the second largest share in 2000.

Another interesting result is the contribution of the sample selection correction variable that has a significant contribution to explain the inequality among self-employed white-collar workers for whom this variable was the second most important explanation to the inequality of earnings. Even for white-collar informal workers this variable had an important contribution in 1992 but it almost disappeared eight years later.

8 Conclusions and Suggestions for Further Research

The main purpose of this paper has been to perform an analysis of the variables that affect the occupational choice between 12 different occupational groups and the variables that explain earnings inequality within them. The structural change that the Chilean labour market went through during the 1990s with increased participation of workers with high-level education, and occupations that demand higher education than the average of Chilean workers indicates that future research should give more attention to understand this group of workers. High-level occupations not only account for a sizable percentage of total earning, their income inequality is also substantially higher than the inequality in other occupational groups.

Our results suggest that schooling is a key factor to explain inequality within high-level occupations being higher among self-employed white-collars than among their formal counterparts. In 1992, the percentage explained by education among white-collars was estimated to 18% in the self-employment sector, 4% in the informal sector, and 14% in the formal. In 2000, however, these values increased to 30%, 32%, and 21%, respectively. In low-level occupations the contribution of schooling is much lower, especially among informal workers.

In low-level occupation of the self-employment sector and in low-level occupation of the informal sector the variable that represents hours of work is the one with largest contribution to explain earnings inequality. In 1992, in the self-employment sector, hours of work explained 1%-10%, in the informal 0%-9%, and in the formal 0%-4%. In year 2000 those values were 0%-26%, 3%-28%, and 1%-5%, respectively. As a whole this year we are able to explain 44% of the inequality of earnings among white- collar self-employed workers, 38% among white-collar informal workers, and 33% among white-collar formal workers. Looking at the earnings inequality of manual workers in 2000 we are able to explain 38% in the self-employment sector, 41% in the informal sector, and 25% in the formal sector. In summary, the variables that increased their contribution to explain the dispersion of earnings were to a larger extent education, in the formal sector;

log Hours, in the self-employment sector and in the informal sector. To a minor extent, the Male-variable, in lower-level occupations of the self-employment sector and the formal sector; and Province, among manual workers of the informal sector.

Our work has not only shed light into the issues that we were initially investigating. It has contributed to arise other questions and generated some ideas for future research. We may extend the type of model used in this paper to include not only the choice between occupations but also the choice to participate in the labour market. This is a relevant issue given that Chile has been found to have low level of female labour force participation during this period (among females of age 25-45), even compared with other Latin American countries, see Inter-American Development Bank (1998). Moreover, we have found that part time work has become an important aspect of the self-employment and the informal sector. Therefore we may need to take this aspect into account in the future.

Another possible area for further investigation is to construct two different sample selection variables that could be used in the earning regression, one controlling for participation, the other for the occupation choice. There are several examples of dual selection models in the literature, see for instance Vijeberg (1993) and Tunai (1986). This research has been concentrated to the migration decision and the employment decision, but we think that this approach also may be applied to the participation and the sector employment decisions.

Finaly, as mentioned before, the inclusive value parameters are estimates of substitutability of alternatives within nests. Therefore, the nested logit model may also be used to investigate mobility between different segments of the labour market. This may contribute to the debate on the eventual existence of labour market segmentation. To our knowledge there is no application of the nested logit model in this context.

9 References

Amuedo-Dorantes, C. (2003): Determinants and Poverty Implications of Informal Sector Work in Chile, San Diego State University. Forthcoming in the Economic Development and Cultural Change.

Amuedo-Dorantes, C. (2004): Work Contracts and Earnings Inequality: The Case of Chile, San Diego State University. Forthcoming in the Economic Development and Cultural Change.

Arellano, M. S., and M. Braun (1999): Rentabilidad de la Educación Formal en Chile, Cuadernos de Economía, 36, 685-724.

Beyer, H. (1997): Distribución del Ingreso: Antecedentes para la Discusión, Estudios Públicos, 65 (Verano), 1-54.

Borsch-Supan (1987): Econometric Analysis of Discrete Choice with Application on the Demand for Housing in the U.S. and Western Germany. Lecture Notes in Economics and Mathematical Systems, 296, Heidelberg: Springer-Verlag.

Borsch-Supan (1990): On the Compatibility of Nested Logit Models with Utility Maximization, Journal of Econometrics, 43, 373-388.

Contreras, D. (2002): Explaining Wage Inequality in Chile: Does Education Really Matter? The Journal of Income Distribution, 11(1), 53-76.

Contreras, D. (2003): Poverty and Inequality in a Rapid Growth Economy: Chile 1990-1996, The Journal of Development Studies, 39(3), 181-200.

Falaris, E. M. (1987): A Nested Logit Migration Model with Selectivity, International Economic Review, 28(2), 181-200.

Fields, G., and G. Yoo (2000): Falling Labour Income Inequality in Korea's Economic Growth: Patterns and Underlying Causes, Review of Income and Wealth, 46(2), 139-160.

García-Huidobro. (1999): La Capacidad Generadora de Empleo de la Economía Chilena, Serie Reformas Económicas, CEPAL, # 31.

Gill, I. S., and C. Montenegro (2002): Responding to Earnings Differentials in Chile, Chapter 6 in "Crafting Labour Policy: Techniques and Lessons from Latin America", ed. by I. S. Gill, C. E. Montenegro and D. Dömeland, World Bank and Oxford University Press.

Gindling, T. H. (1991): Labour Market Segmentation and the Determination of Wage in the Public, Private-Formal, and Informal Sectors in San José, Costa Rica, Economic Development and Cultural Change, 586-605.

Hagstrom, P. A. (1996): The Food Stamp Participation and Labour Supply of Married Couples: An Empirical Analysis of Joint Decisions, The Journal of Human Resources, 31(2), 383-403.

Hauber, B., and G. R. Parsons (2000): The Effect of Nesting Structure Specification on Welfare Estimation in Random Utility Model of Recreation Demand: An Application to the Demand for Recreation Fishing, American Journal of Agricultural Economics, 82, 501-514.

Heckman, J., and Hotz J. (1986): An Investigation of the Labour Market of Panamanian Males Evaluating the Sources of Inequality, Journal of Human Resources, Vol. XX1(4), 507-542.

Inter-American Development Bank (1998): Facing Up to Inequality in Latin America, Economic and Social Progress in Latin America 1998-1999 Report. Maryland, The Johns Hopkins University Press.

Larrañaga, O. (1999): Distribución de Ingresos y Crecimiento Económico en Chile, Serie Reformas Económicas, CEPAL, # 35.

Lee, B. (1999): Calling Patterns and Usage of Residential Toll Service under Self Selecting Tariffs, Journal of Regulatory Economics, 16, 45-81.

Le, A. T. (1999): Empirical Studies of Self-employment, Journal of Economic Surveys, 13(4), 381-416.

Litchfield, J. (2002): Updated Income Distribution and Poverty Measures for Chile 1987-1998, Background Paper 1 to Chile's High Growth Economy, Poverty and Income Distribution 1987-1998, World Bank Country Study.

McFadden, D. (1981): Econometric Models of Probabilistic Choice in Structural Analysis of Discrete Data, C. Manski and D. McFadden (Eds.), Cambridge, MA: MIT Press.

MIDEPLAN (2001): Pobreza, Crecimiento y Distribución del Ingreso en Chile en los Noventa, Santiago: Ministerio de Planificación y Cooperación, Documento # 24.

Mincer, J. (1974): Schooling, Experience, and Earnings, Columbia University Press, New York.

Palma, A. (2007): Explaining Earnings and Income Inequality in Chile. Economic Studies, Department of Economic, School of Business, Economics and Law, Göteborg University, 169.

Palma, A. (2014): The Redistributive Impact of Public Policies in a Country With a High Level of Income Inequality: The Case of Chile. Poverty and Public Policy, Vol 6 Issue 2.

Riveros, L. A. (1990): The Economic Return to Schooling in Chile. An Analysis of its Long-term Fluctuations, Economics of Education Review, 9(2), 111-121.

Saavedra, J., and A. Chong (1999): Structural Reform, Institutions, and Earnings: Evidence from Formal and Informal Sectors in Urban Peru, The Journal of Development Studies, 35(4), 95-116.

Soopramanien, D., and G. Jones (2001): A New Lookat Gender Effects in Participation and Occupation Choice, LABOUR, 15(3), 415-443.

Tiefenthaler, J. (1994): A Multisector Model of Female Labor Force Participation: Empirical Evidence from Cebu Island, Philippines, Economic Development and Cultural Change, 719-742.

Tunai, I. (1986): A General Structure for Models of Double Selection and in Application to Joint Migration-Earnings Process with Remigration, In Research on Labor Economics, Vol. 8, Part B, ed. Ronald G. Ehrenberg, 235-282. Greenwich, Conn: JAI Press, Inc.

Yuhong, D. and Johnes, O. (2003): Influence of Expected Wages on Occupational Choice: new Evidence from Inner Mongolia, Applied Economics Letters, 10, 829–32.

Vijverberg, P. M. (1993): Educational Investments and Returns for Women and Men in Cote d’Ivore, The Journal of Human Resources, 28(4), 933-974.

Appendix

Table A1
Definition of Variables Used in the Paper

Variables	Definition
Schooling	Years of formal schooling.
Eh	Years of schooling of the household for those alder than 15 years, excluding the household head.
Age	Age
Age^2	Age squared /10.
Inc	Unearned income /1000.
Male	Dummy for males.
Rural	Dummy for residents in rural area.
Province	Dummy for residents outside Santiago.
Married	Dummy for married.
Baby	Dummy if there are children of less than 5 years of age in the household.
Child	Dummy if there are children of 6 ≤age≤ 10 in the household.
Young	Dummy if there are children of 11 ≤age≤ 15 in the household.
Unemp	Dummy if there are unemployed individuals in the household.
Inac	Dummy if there are inactive individuals in the household.
House	Dummy for house owners.

Table A2
Mean of Explanatory Variables, CASEN 1992

Variables	Self-Employment				Informal Employment				Formal Employment			
	White-collar	Clerical	Blue-collar	Manual	White-collar	Clerical	Blue-collar	Manual	White-collar	Clerical	Blue-collar	Manual
Schooling	11.94	8.80	7.58	6.39	13.42	8.92	7.71	5.65	14.86	10.96	8.61	6.93
Age	45.10	45.40	45.26	42.66	41.01	38.72	38.50	40.76	41.27	39.41	39.96	40.37
Inc	25.23	10.85	7.33	4.81	12.09	5.39	3.77	4.19	17.11	7.18	5.42	4.43
Male	89.80	72.92	95.25	85.66	94.23	70.98	96.38	78.79	87.46	79.73	96.52	86.34
Rural	6.52	7.63	28.44	29.09	0.98	7.70	17.77	44.30	4.91	4.93	12.34	30.77
Province	51.79	52.58	65.52	65.63	55.36	52.90	56.34	74.70	48.04	53.89	54.33	65.19
Married	81.04	61.99	78.52	62.90	78.40	54.19	71.97	58.86	81.06	72.21	83.44	69.88
Baby	31.36	29.49	36.77	46.18	37.22	49.94	49.93	49.61	39.60	41.76	46.56	48.57
Child	29.27	32.26	33.44	36.05	27.96	25.62	38.20	35.66	32.98	30.87	35.35	33.75
Young	26.02	21.84	28.43	27.73	19.71	23.68	27.42	27.32	24.28	27.14	28.25	28.13
Unemp	3.25	6.90	5.09	6.02	6.93	8.49	7.00	6.00	3.20	5.26	6.00	5.23
Inact	65.61	63.69	80.65	74.78	71.68	62.95	78.50	73.38	66.33	69.43	80.72	75.30
House	60.01	55.97	62.49	59.42	40.27	38.77	39.12	48.01	40.76	38.15	46.44	42.87

Source: Own calculations from CASEN.

Table A3

Mean of Explanatory Variables, CASEN 2000

Variables	Self-Employment				Informal Employment				Formal Employment			
	White-collar	Clerical	Blue-collar	Manual	White-collar	Clerical	Blue-collar	Manual	White-collar	Clerical	Blue-collar	Manual
Schooling	12.38	9.55	8.88	6.80	14.24	10.59	8.63	6.63	15.47	11.56	9.48	7.98
Age	46.06	45.72	45.87	45.05	39.64	40.81	41.41	43.21	41.71	40.02	41.03	42.28
Inc	42.77	19.79	15.78	11.66	23.61	13.75	9.77	8.68	35.74	17.50	10.53	10.66
Male	85.76	61.88	93.82	75.82	78.75	54.82	97.22	65.66	81.22	74.11	97.11	84.67
Rural	7.49	4.20	18.25	12.90	2.21	3.84	15.32	29.03	2.33	2.73	12.19	22.55
Province	51.87	47.68	63.84	57.96	47.81	56.18	61.13	64.44	47.45	49.61	58.98	63.28
Married	70.56	48.45	72.08	54.76	50.39	42.84	68.97	48.33	70.44	64.54	78.68	65.65
Baby	29.43	30.03	31.23	36.89	27.56	37.80	41.83	38.20	32.05	37.02	42.58	40.74
Child	31.18	37.88	35.02	41.62	20.83	34.23	38.89	36.49	32.74	32.87	40.21	37.29
Young	28.00	30.28	31.73	29.83	20.87	21.71	32.06	31.75	24.34	26.13	32.46	30.08
Unemp	5.04	9.96	9.22	10.47	4.74	9.75	11.94	10.98	5.32	8.60	8.79	9.89
Inact	61.43	59.37	74.11	71.53	52.04	51.39	73.22	68.22	56.58	61.19	78.16	71.41
House	52.25	52.27	55.64	55.62	25.82	38.71	42.58	47.72	33.25	35.46	40.43	41.71

Source: Own calculations from CASEN.

Table A4
Maximum Likelihood Estimates of the Nested Logit Model,
Choice among Sectors, CASEN 1992
(Omitted Sector: Formal Employment)

Variables	Self-Employment	Informal Employment	Formal Employment
Constant	-1.4325***	-0.4595***	
	(21.53)	(19.12)	
Schooling	-0.0690**	-0.0917***	
	(1.98)	(14.16)	
25-34	0.1720**	0.0748***	
	(2.09)	(3.27)	
35-44	0.3024*	0.0359	
	(1.69)	(0.96)	
45-54	0.2093	-0.2610***	
	(0.78)	(7.06)	
55-64	0.3347	-0.1511***	
	(0.98)	(5.58)	
Male	0.3006***	-0.2327***	
	(4.54)	(3.49)	
Province	0.0913***	0.1369***	
	(4.69)	(4.74)	
Rural	0.1075**	0.3988***	
	(2.56)	(12.33)	
Married	-0.5067***	-0.4709***	
	(4.56)	(24.27)	
Baby	0.1102	0.1389***	
	(1.14)	(10.10)	
Child	0.1487***	0.1058***	
	(4.19)	(9.99)	
Young	-0.0542***	-0.0506***	
	(3.14)	(4.71)	
Unemp	-0.1827***	0.2926***	
	(22.39)	(26.48)	
Inact	-0.1937***	-0.0515***	
	(27.71)	(7.48)	
Huse	0.4887***	0.0005	
	(123.09)	(0.09)	
Inc	0.0089***	-0.0098***	
	(33.11)	(21.14)	
μ_m	0.8546**	0.3603**	0.5482***
	(2.51)	(2.56)	(5.30)

Log-likelihood	-4264203
Log-likelihood constant	-5795090
n	23,544

Source: Own calculations from CASEN.
Notes: Robust z statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A5

Maximum Likelihood Estimates of the Nested Logit Model,
Choice among Occupations, CASEN 1992
(Omitted Occupation: Manual Occupation in Respective Sector)

	Self-Employment			Informal Employment			Formal Employment		
Variables	White-collar	Clerical	Blue-collar	White-collar	Clerical	Blue-collar	White-collar	Clerical	Blue-collar
Constant	-7.4967***	-3.6112***	-3.6569***	-15.4357***	-3.3206***	-3.1172***	-12.5486***	-5.1062***	-4.0867***
	(29.51)	(17.72)	(17.39)	(29.01)	(19.23)	(18.15)	(156.23)	(65.73)	(64.58)
Schooling	0.3761***	0.1696***	0.1192***	0.6795***	0.2230***	0.1288***	0.7495***	0.3233***	0.1051***
	(112.55)	(47.72)	(43.15)	(55.15)	(61.19)	(58.42)	(267.13)	(146.44)	(60.52)
Age	0.1567***	0.1126***	0.0961***	0.2166***	0.0333***	0.0280***	0.1721***	0.1112***	0.1242***
	(14.93)	(11.60)	(10.71)	(7.81)	(4.23)	(3.41)	(50.20)	(34.82)	(50.04)
Age^2	-0.0122***	-0.0091***	-0.0070***	-0.0197***	-0.0032***	-0.0034***	-0.0150***	-0.0114***	-0.0138***
	(12.30)	(8.37)	(7.23)	(5.62)	(3.37)	(3.36)	(34.64)	(31.25)	(50.17)
Male	-0.0493	-0.8718***	1.0233***	2.1274***	0.0768*	2.3884***	-0.3415***	-0.6717***	1.3778***
	(0.71)	(21.10)	(38.68)	(27.65)	(1.90)	(59.46)	(14.30)	(41.83)	(95.71)
Province	-0.1098***	-0.1549***	0.0431	-0.2082***	-0.3541***	-0.4954***	-0.3206***	-0.0803***	-0.2017***
	(3.64)	(6.97)	(1.60)	(6.47)	(15.13)	(31.35)	(27.81)	(7.46)	(21.51)
Rural	-0.8130***	-1.0544***	0.1350***	-3.4858***	-1.7391***	-1.1981***	-0.8375***	-1.3795***	-0.9616***
	(10.44)	(30.56)	(6.11)	(28.55)	(52.82)	(56.19)	(16.86)	(83.10)	(119.34)
Married	0.8592***	0.4927***	0.4487***	0.5668***	0.1344***	0.1473***	0.5848***	0.4791***	0.4081***
	(18.96)	(21.39)	(22.97)	(11.25)	(4.49)	(6.89)	(22.42)	(31.56)	(40.27)
Baby	-0.4419***	-0.5444***	-0.2747***	-0.5824***	-0.0552**	-0.3183***	-0.2716***	-0.3354***	-0.1973***
	(22.12)	(25.09)	(15.10)	(12.18)	(2.51)	(18.78)	(23.42)	(32.96)	(24.34)
Child	-0.2643***	0.0353*	-0.1049***	-0.6188***	-0.4753***	-0.0736**	-0.0418***	-0.1679***	-0.0531***
	(13.59)	(1.81)	(5.93)	(10.09)	(20.08)	(2.43)	(3.50)	(15.68)	(6.09)
Young	-0.1206***	-0.3570***	-0.0319	-0.3582***	-0.0903***	0.0065	-0.2368***	-0.0743***	-0.1232***
	(3.11)	(11.88)	(1.39)	(6.50)	(3.92)	(0.40)	(16.37)	(5.70)	(13.19)

Source: Own calculations from CASEN.

Notes: Robust z statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A6

Maximum Likelihood Estimates of the Nested Logit Model,
Choice among Sectors, CASEN 2000
(Omitted Sector: Formal Employment)

Variables	Self-Employment	Informal Employment	Formal Employment
Constant	-0.7105***	0.7917***	
	(5.26)	(7.02)	
Schooling	-0.2283*	-0.1547***	
	(1.89)	(10.63)	
25-34	0.1300	-0.0343	
	(1.43)	(0.65)	
35-44	0.3048	0.0386	
	(1.56)	(0.51)	
45-54	0.2803	-0.1317***	
	(0.94)	(2.71)	
55-64	0.1059	-0.2676***	
	(0.28)	(6.32)	
Male	-0.4797	-0.9437***	
	(1.04)	(4.71)	
Province	-0.0908**	-0.0440**	
	(2.05)	(2.02)	
Rural	-0.7577**	0.4409***	
	(1.99)	(8.26)	
Married	-0.4848***	-0.3612***	
	(3.14)	(10.51)	
Baby	0.1510	0.0567***	
	(0.70)	(3.41)	
Child	0.4170*	0.0577**	
	(1.65)	(2.30)	
Young	0.0081	0.0990***	
	(0.16)	(8.39)	
Unemp	-0.2089***	0.1348***	
	(31.16)	(17.06)	
Inact	-0.2197***	-0.0926***	
	(30.49)	(17.61)	
Huse	0.3883***	0.0557***	
	(74.40)	(11.15)	
Inc	0.0043***	-0.0062***	
	(21.56)	(30.73)	
μ_m	1.2684*	0.7605***	0.4633***
	(1.76)	(2.97)	(3.21)

Log-likelihood	-4536535
Log-likelihood constant	-6076190
n	37,107

Source: Own calculations from CASEN.

Notes: Robust z statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A7

Maximum Likelihood Estimates of the Nested Logit Model,
Choice among Occupations, CASEN 2000
(Omitted Occupation: Manual Occupation in Respective Sector)

Variables	Self-Employment			Informal Employment			Formal Employment		
	White-collar	Clerical	Blue-collar	White-collar	Clerical	Blue-collar	White-collar	Clerical	Blue-collar
Constant	-5.5688***	-1.7464**	-2.9588***	-6.5492***	0.4829***	-2.9817***	-9.8979***	-2.7935***	-3.9431***
	(9.46)	(2.42)	(4.16)	(23.05)	(3.14)	(29.28)	(111.18)	(35.46)	(61.06)
Schooling	0.4066***	0.1964***	0.1779***	0.6642***	0.3050***	0.1486***	0.8703***	0.3232***	0.1040***
	(24.84)	(15.26)	(14.98)	(60.51)	(62.67)	(53.30)	(255.79)	(106.87)	(44.15)
Age	0.0657***	0.0034	0.0492***	-0.1190***	-0.2192***	-0.0447***	-0.0249***	0.0211***	0.0978***
	(9.21)	(0.18)	(3.46)	(6.63)	(21.76)	(9.38)	(5.64)	(5.90)	(31.47)
Age^2	-0.0018**	0.0028***	-0.0024***	0.0158***	0.0273***	0.0065***	0.0074***	-0.0020***	-0.0111***
	(2.33)	(2.58)	(3.73)	(5.79)	(17.30)	(13.19)	(12.87)	(4.55)	(27.90)
Male	0.5574***	-0.5451***	1.4847***	1.3239***	-0.1948***	3.1874***	-0.4755***	-0.8163***	1.7541***
	(7.78)	(12.29)	(23.82)	(18.20)	(3.96)	(73.53)	(23.31)	(48.93)	(112.69)
Province	-0.0681	-0.2268***	0.2240***	-0.2490***	0.0592**	0.1022***	-0.4203***	-0.3389***	-0.0507***
	(1.01)	(6.14)	(3.38)	(9.97)	(2.31)	(5.81)	(34.90)	(29.16)	(4.33)
Rural	0.6455***	-0.4352***	0.6662***	-1.3540***	-1.5511***	-0.8915***	-0.6482***	-1.3119***	-0.5638***
	(7.04)	(7.70)	(9.02)	(9.63)	(26.49)	(16.41)	(24.36)	(82.02)	(56.36)
Married	0.2830***	-0.0116	0.1795***	-0.2734***	0.1037***	-0.0160	0.3374***	0.4063***	0.2093***
	(6.23)	(0.14)	(2.66)	(5.15)	(4.87)	(0.94)	(17.43)	(32.07)	(20.64)
Baby	-0.2055**	-0.1963***	-0.2235***	-0.5580***	-0.0462**	-0.0118	-0.3685***	-0.2732***	-0.0337***
	(2.10)	(3.33)	(2.98)	(22.10)	(1.97)	(0.52)	(27.46)	(23.72)	(3.27)
Child	-0.3296***	-0.0009	-0.3013***	-0.5751***	0.0301	-0.0394**	0.0617***	-0.1992***	-0.0221**
	(4.99)	(0.02)	(4.07)	(17.78)	(1.15)	(2.45)	(5.00)	(18.31)	(2.37)
Young	-0.0670	0.1275***	0.0618	-0.3762***	-0.2932***	0.0014	-0.4229***	-0.2038***	-0.0130
	(1.64)	(2.90)	(1.39)	(15.11)	(9.62)	(0.09)	(30.87)	(15.72)	(1.37)

Source: Own calculations from CASEN.

Notes: Robust z statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A8

Estimates of the Occupational Earnings Equation, CASEN 1992
(Dependent Variable: Log of Monthly Earnings)

	Self-Employment				Informal Employment				Formal Employment			
Variables	White collar	Clerical	Blue collar	Manual	White collar	Clerical	Blue collar	Manual	White collar	Clerical	Blue collar	Manual
Constant	10.589***	10.327***	6.162***	7.271***	9.782***	8.338***	7.854***	7.978***	6.830***	9.900***	8.462***	9.326***
	(10.536)	(11.373)	(15.327)	(15.725)	(3.926)	(6.096)	(9.963)	(17.102)	(11.531)	(15.788)	(28.094)	(34.017)
Schooling	0.097***	0.070***	0.063***	0.043***	0.025	0.049***	0.035***	0.016**	0.092***	0.067***	0.061***	0.021***
	(9.717)	(6.756)	(13.508)	(4.931)	(0.654)	(3.416)	(3.578)	(2.207)	(12.183)	(9.195)	(19.360)	(6.656)
Age	0.027	-0.017	0.055***	0.039**	0.122*	-0.006	-0.018	0.013	0.032**	0.057***	0.059***	-0.003
	(1.040)	(0.665)	(5.430)	(2.390)	(1.956)	(0.235)	(1.198)	(0.991)	(2.397)	(4.985)	(9.156)	(0.595)
Age^2	-0.003	0.002	-0.005***	-0.005**	-0.015**	0.001	0.002	-0.002	-0.002	-0.005***	-0.006***	0.000
	(1.015)	(0.609)	(4.059)	(2.469)	(2.185)	(0.196)	(0.984)	(1.022)	(1.406)	(3.969)	(7.562)	(0.757)
Male	0.454***	0.441***	0.881***	0.296***	1.055***	0.134	0.490***	0.285***	0.435***	0.374***	0.342***	0.289***
	(4.353)	(3.368)	(9.884)	(3.682)	(3.849)	(0.730)	(3.577)	(5.189)	(7.978)	(7.218)	(5.612)	(10.632)
Province	-0.120*	-0.054	-0.160***	-0.173***	-0.081	0.176*	-0.259***	-0.271***	-0.332***	-0.128***	-0.097***	-0.127***
	(1.885)	(0.733)	(4.451)	(3.101)	(0.427)	(1.760)	(4.310)	(5.463)	(9.573)	(3.681)	(4.643)	(5.875)
Log Hour	0.323***	0.391***	0.476***	0.430***	0.247	0.406***	0.389***	0.469***	0.701***	0.036	0.193***	0.237***
	(3.379)	(3.628)	(10.196)	(8.050)	(0.951)	(3.927)	(3.572)	(6.570)	(7.723)	(0.406)	(3.977)	(5.125)
λ_{jm}	0.704***	0.327	-0.476***	-0.490***	0.769*	-0.143	-0.607***	-0.131**	0.198***	0.331***	-0.121**	-0.333***
	(6.078)	(1.430)	(5.179)	(4.892)	(1.960)	(0.393)	(2.894)	(2.091)	(5.183)	(3.821)	(2.062)	(7.887)
n	1401	840	3728	1474	74	172	673	1359	2271	1926	5364	4241
R^2	0.328	0.176	0.239	0.260	0.296	0.226	0.153	0.188	0.331	0.198	0.155	0.168

Source: Own calculations from CASEN.

Notes: t-statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A9

Estimates of the Occupational Earnings Equation, CASEN 2000

(Dependent Variable: Log of Monthly Earnings)

Variables	Self-Employment				Informal Employment				Formal Employment			
	White collar	Clerical	Blue collar	Manual	White collar	Clerical	Blue collar	Manual	White collar	Clerical	Blue collar	Manual
Constant	9.132***	7.938***	6.710***	6.920***	7.042***	6.615***	7.493***	8.167***	7.355***	9.354***	8.572***	9.073***
	(7.844)	(6.323)	(16.341)	(9.334)	(6.227)	(10.326)	(18.110)	(33.776)	(13.212)	(22.837)	(28.844)	(33.213)
Schooling	0.128***	0.057***	0.068***	0.036***	0.155***	0.052***	0.037***	0.024***	0.129***	0.089***	0.067***	0.025***
	(8.665)	(3.855)	(14.501)	(3.814)	(7.101)	(4.407)	(5.950)	(4.994)	(11.819)	(16.492)	(21.815)	(6.153)
Age	0.101***	0.033	0.057***	0.063**	0.047	0.026	-0.015	0.002	0.035**	0.030***	0.033***	0.001
	(3.017)	(0.781)	(5.178)	(2.341)	(0.930)	(0.902)	(1.050)	(0.291)	(2.277)	(2.638)	(5.025)	(0.218)
Age^2	-0.011***	-0.004	-0.005***	-0.006**	-0.004	-0.003	0.002	-0.000	-0.003	-0.003*	-0.003***	0.000
	(3.193)	(0.814)	(4.371)	(2.272)	(0.647)	(0.979)	(1.166)	(0.069)	(1.494)	(1.931)	(4.011)	(0.069)
Male	0.445***	0.199	0.857***	0.264***	0.076	-0.046	0.721***	-0.040	0.342***	0.204***	0.612***	0.162***
	(5.486)	(1.042)	(8.347)	(2.884)	(0.577)	(0.273)	(5.320)	(1.060)	(8.140)	(4.840)	(9.361)	(6.435)
Province	-0.124	-0.043	-0.186***	-0.183***	-0.144	-0.180**	-0.220***	-0.266***	-0.202***	-0.172***	-0.069***	-0.149***
	(1.400)	(0.427)	(4.721)	(2.881)	(1.190)	(2.161)	(5.674)	(9.130)	(4.994)	(5.949)	(3.869)	(7.120)
Log Hour	0.192***	0.456***	0.391***	0.520***	0.305***	0.547***	0.443***	0.497***	0.515***	0.172***	0.208***	0.311***
	(3.658)	(6.400)	(14.020)	(11.367)	(3.177)	(6.499)	(10.378)	(18.062)	(7.278)	(2.897)	(5.111)	(7.182)
λ_{jm}	0.556***	-0.252	-0.455***	-0.194*	-0.205	-0.539*	-0.598***	-0.260***	0.103	-0.040	-0.301***	-0.392***
	(3.566)	(0.762)	(4.020)	(1.733)	(1.114)	(1.885)	(4.485)	(4.812)	(1.475)	(0.628)	(5.301)	(7.873)
n	2991	770	6246	1075	255	429	2163	2771	3153	3068	8497	5688
R^2	0.445	0.259	0.325	0.379	0.382	0.411	0.330	0.411	0.331	0.267	0.217	0.249

Source: Own calculations from CASEN.

Notes: t-statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%