Contribution of the determinants of income inequality in Guatemala

Luis Alejandro Alejos

October 2003

Instituto de Investigaciones Económicas y Sociales

Universidad Rafael Landívar

Tradición Jesuita en Guatemala

Contribution of the determinants of income inequality in Guatemala[1]

Luis Alejandro Alejos[2]
October 2003

Abstract

This paper decomposes income inequality in Guatemala in factors related to human capital, ethnic and gender discrimination, the occupational structure, and non-labour income. The method proposed by Fields (2002) is used to carry out this decomposition. The empirical results show a significant variation in the contribution between the determinants at a national level, and those of each socio-economic group in which the sample is divided. It is found that the most heterogeneous group is that of agriculture and livestock workers. Nonetheless, the role of education as one of the main determinants of income inequality is persistent across the sample.

JEL classification*: D31, J31*
Key words: Guatemala, income inequality, education, decomposition.

---

[1] The original Spanish version "Participación de los Determinantes de la Desigualdad de Ingresos" is available at http://www.url.edu.gt/idies/publica/Revistas/REVISTA70.pdf

[2] I would like to thank Craig McIntosh (University of California at Berkeley) for useful comments and suggestions. I am also grateful to Tomás Rosada, IDIES Director, and my colleagues at IDIES for their support during this research.

Contribution of the determinants of income inequality in Guatemala

1. Introduction

One of the main objectives of any social policy directed at the development of a society is its efficiency in achieving the benefits it pursues. In the case of a policy oriented to reduce income inequality, it is necessary to understand what the determinants of this inequality are, as well as their contribution to the income differentials observed across the sample, in order to increase the efficiency of such policy. This is particularly important for countries like Guatemala, in which the resources to implement these type of policies are limited. Moreover, it is also important to determine what the impact of social policies may be for different socio-economic groups in society.

Two questions arise from the previous remarks: (1) how much income inequality in the country is accounted for by each explanatory factor? (2) what is the contribution of these determinants in the inequality of income within a given socio-economic group in society? The answers to these two questions could be useful in the design of social policies aiming to reduce income inequality both at national level and within social groups[3].

Both questions outline the need to point out, with relative precision, the contribution that each determinant has in the distribution of income, so that more importance can be given to those factors which have a greater effect on it.

Among the different methods found in the literature of inequality decomposition, the technique developed by Fields (2002) has been chosen, since it allows an axiomatic decomposition which does not suffer from the path dependence problem found in other decompositions. This method proposes to derive the contribution of each explanatory variable to the variance of the dependent variable, in the framework of a multiple regression model. Therefore, this methodology allows us to answer the two questions that this paper focuses on.

As it will be seen in its contents, the main contribution of this paper is not in itself to answer the questions previously stated, but to focus in the process of doing so in order to identify the problems which it involves. Particular emphasis is given to the heterogeneity of results shown at the group level, especially for the group of agriculture and livestock workers. Such heterogeneity represents a significant difficulty for the analysis of the contribution of each determinant of income inequality at the national level, due to the huge variations within each group, particularly in terms of educational attainment. Therefore, the formulation of social policies that allow the reduction of income inequality both nationally and within groups is not an easy task. This leads us to the conclusion that it is necessary to concentrate our analytical efforts in the

[3] The idea tried to be captured in this paragraph is that there is a possibility for an income inequality reduction-policy (focused on one factor) to lower national and within group inequality, while reaffirming between group differentials. This may lead to the creation of social tension between the groups involved, which is highly undesirable from a social point of view. If a clear conception of the structure of inequality is available, this possibility can be detected more easily.

distribution of income for different socio-economic groups within society, in order to obtain a clearer picture of reality. The present paper carries out such an analysis in a superficial way, acknowledging the limitations imposed by the use of cross-section data.

The analysis is structured as follows: Section 2 is a brief summary of the wage differentials literature. Section 3 focuses in Fields decomposition technique and presents the regression model used to carry out such decomposition. Section 4 shows the empirical results derived from applying the methodology outlined in Section 3. Finally, the conclusions of this analysis are presented in Section 5. Furthermore, two appendices, one statistical and the other methodological, complement the investigation.

2. Theoretical Approach

When understanding the importance of income inequality, it is crucial to mention that this inequality, in itself, is not a worrying fact. The concern arises because it creates inequality of opportunities[4]. It is for this reason that we concentrate in accounting for labour and non-labour income differentials, in an attempt to explain the inequality of income and the factors that contribute to its existence.

In the literature of earnings differentials we can find various theories that attempt to clarify this phenomenon by focusing either on the workers' characteristics or on the structure of the labour market in which they participate[5]. We can classify within the first group the theories of human capital and labour market discrimination. In the second, we find the theories of compensating wage differentials and efficiency wages, and the dual sector framework.

Moreover, we can also think of factors not enclosed in these theories. For instance, non-labour income is an important source of income for many individuals. The focus here will be mainly on variations in asset holdings and the receipt of remittances from abroad, as well as transfer payments.

There are several other events that can determine income differentials at a point in time: floods, hurricanes, climate changes, nepotism, corruption, differences in social capital, and economic and political crises can play a major role in determining the distribution of income.

I shall focus only on some of the factors mentioned above, especially because of measurement problems, which will be discussed in Section 3.

2.1 Differences in human capital

The pioneering works by Mincer (1958, 1974), Becker (1962) and Schultz (1960) revealed the importance of human capital in workers' productivity, and hence, in the determination of labour income. Modern studies have confirmed this fact and relate it with the growth and development of societies throughout higher productivity.

There are many sources of human capital. Becker identifies four main sources: schooling, on-the-job training, health and information. All of them are said to improve the physical and mental abilities of the individuals, thereby raising their productivity and wages. Of these, research has focused mostly on schooling[6].

---

[4] See Hild and Voorhoeve (2003) for a discussion on the concept of inequality of opportunities. The definition of inequality of opportunities to which this paper refers is formally represented by the concept of 'opportunity dominance', which these authors develop in their discussion.

[5] The list that follows is not exhaustive. Indeed, it is usual to find in empirical studies a large percentage of unexplained variation in income levels among individuals (see Cowell and Jenkins, 1995).

[6] As pointed out by Carneiro and Heckman (forthcoming, 2003): "Human capital accumulation is a dynamic process. The skills acquired in one stage of the life cycle affect both the initial conditions and the technology of learning at the next stage. Human capital is produced over the life cycle by families, schools, and firms, although most discussions of skill formation focus on schools as the major producer of abilities and skills, despite a substantial body of evidence that families and firms are also major producers of abilities and skills".

In the case of Guatemala, a country with great disparities in educational attainment among its social groups, it seems reasonable to expect that education is a major determinant of income inequality[7]. It is important to notice that when considering education, we have a problem of two dimensions. On the one hand, we are concern about the quantity of education. On the other, we focus on its quality. Let us assume for a moment that we observe two individuals with identical characteristics. If wages still differ, we may presume that this gap is formed by differences in the quality of education received, because a better education would proportionate greater levels of human capital to the individual, having an incidence in his productivity.

On-the-job training has also two significant dimensions. The distinction between the two elements relies on whether on-the-job training leads to firm-specific or general skills. This has important implications in within-jobs and between-jobs labour mobility and its consequent earnings gains or losses.

The degree of information about wages offered by different employers can also play a role in the inequality of earnings. The most informed among workers are expected to be able to obtain higher returns to their skills, simply because they are more likely to find a firm that is willing to pay higher wages for their abilities.

Finally, the state of emotional and physical health can cause variations in worker's productivity and therefore determine income differentials.

2.2 Discrimination in the labour market[8]

Guatemala is a multicultural country and, particularly, a country clearly divided between indigenous and non-indigenous populations. This division is a consequence of Guatemala's colonial past under the Spanish domination. However, the indigenous civilizations that inhabited the region before the Spanish conquest were also deeply segregated[9]. Therefore, taking into account the historic stratification reigning in the country, it does make sense to think about the existence of significant ethnic discrimination in the labour market.

Nevertheless, discrimination has its origins not only in the ethnical segregation, but also in the social culture, which narrows opportunities for women. There is also evidence which leads us to consider gender discrimination as an important determinant of income inequality[10].

---

[7] See Chart A.1.2 in Appendix 1.

[8] Here we are concerned only in the direct effects that discrimination may have in earnings. It is acknowledged, however, that discrimination plays a major role in determining other variables affecting income (e.g. worker's occupation, area of living, etc.).

[9] See Chapter II of the Guatemalan Human Development Report 2000 – "Guatemala: la fuerza incluyente del desarrollo humano", for an analysis of the historical causes of the present social stratification of the Guatemalan society. Martínez Peláez (1971) also presents a profound historical analysis of this fact.

[10] See Jusidman (2002) for a reflection on topics concerning gender discrimination in Guatemala. It is important to mention that, from the employer's point of view, employing female workers may involve higher costs due to the implications of maternity periods. This can act as a negative incentive for female employment, limiting women's participation in the labour market.

2.3 Compensating wage differentials and efficiency wages

The theory of compensating differentials justifies the variation in earnings existing across different industries, regarding them as a consequence of distinct working environments. For instance, a job that may involve potential health damage needs to compensate individuals in order to attract them to this industry. This can be achieved by paying workers a premium which compensates for the disagreeable working conditions[11].

On the other hand, the theory of efficiency wages is based in the idea that higher-than-market wages can increase worker productivity and/or decrease costs for the firm. This works mainly through the channels of better nutrition (physical health), reduction of adverse selection and moral hazard problems, psychological factors (e.g. motivation), and a decrease in turnover and screening costs (Krueger and Summers, 1988).

Both the theory of compensating differentials and the theory of efficiency wages invite us to consider the earnings differentials arising from different economic activities (i.e. inter-sectoral differentials). In Guatemala, a large part of the workforce is employed by the agricultural sector, where salaries are low. Hence, we expect worker occupation to be one of the main determinants of the inequality of income.

2.4 Dual labour market

Lewis (1954) was the first who presented a dual sector model, considering a 'subsistence' agricultural sector, with low wages and low productivity, and a modern 'capitalist' sector that could attract workers from rural areas due to the presence of higher average wages and higher productivity. The presence of an 'unlimited' supply of labour would provide an explanation for the subsistence-wage levels[12]. Modern models have incorporated a dual-dual framework, which distinguishes modern (formal) and informal sector enterprises in both urban and rural areas (Bigsten and Levin, 2000)[13].

In the case of Guatemala, we can assure the importance of the agricultural sector. Furthermore, we can also find relatively big structural differences between urban and rural areas[14], as well as between the formal and informal sector.

---

[11] The idea of compensating wage differentials comes from the writings of Adam Smith (Lanfranchi et al., 2001). Among other empirical studies, Lanfranchi et al. (2001) find significant results of the presence of compensating differentials in the French labour market.

[12] Lewis points out to "subsistence agriculture, casual labour, petty trade, domestic service, wives and daughters in the household, and the increase of population" as the main sources of this unlimited labour supply. According to this argument, the potential for an unlimited supply of labour in Guatemala is large.

[13] Referring to Thorbecke and Stiefel (1999), Bigsten and Levin say that using the dual-dual framework "they show that population shifts between socio-economic groups are an important factor in explaining changes in poverty", which implies that these shifts may also be important in accounting for income differences over time, even when changes in poverty and inequality may not be in the same direction (see Justino et al., 2003).

[14] The "Assessment of Rural Poverty. Latin America and the Caribbean" (IFAD, 2001), relates earnings variations in rural areas with geographic isolation, highlighting that poor villages are established in remote locations, with precarious communication and service systems. Indeed, many poor villages in Guatemala's North-West have these characteristics.

3. Methodology

The data used in this analysis comes from the National Survey of Employment and Income (ENEI), 2002. This survey was designed and carried out by the Guatemalan National Institute of Statistics (INE).

Although the ENEI 2002 accounts for some declared self-consumption (mainly food), the monetary values attached to it are given by the individual. This can lead to inflated figures or to an underestimation of the actual self-consumption. Therefore, the present analysis does not account for self-consumption values in the income aggregate used.

The exclusion of self-consumption is also due to the conceptual and methodological difficulties its measurement involves. For instance, when considering developing countries, Jusidman (2002) identifies food, housing, working tools, firewood and building materials as the most important self-consumption goods, while pointing out at domestic work and child rising as self-consumption services. To account for these variables is impossible in many cases given the available data.

Consequently, individuals who reported being family workers or non-family workers without receiving monetary wages were excluded from the sample. It is acknowledged that the exclusion of this group and the omission of self-consumption can be important sources of bias, which may restrict the validity of our results.

The precise structure of the income aggregate used throughout this paper is presented in Appendix 2. This definition includes both labour and non-labour income. Moreover, the detailed procedure used in the creation and filtering of data is also described in Appendix 2.

3.1 - Fields decomposition technique[15]

As it has been previously pointed out, Fields decomposition technique has been used in order to carry out the proposed analysis of income inequality. The said decomposition is based on an income-generating function. Such function can be written as a linear regression model of the following structure:

$$\ln Y_{it} = a_t' Z_{it} \tag{1.a}$$

where,

$$a_t = [\alpha_t \quad \beta_{1t} \quad \beta_{2t} \; . \; . \; . \; \beta_{Jt} \quad 1] \tag{1.b}$$

and

$$Z_{it} = [1 \quad x_{i1t} \quad x_{i2t} \; . \; . \; . \; x_{iJt} \quad \varepsilon_{it}] \; . \tag{1.c}$$

[15] The derivation that follows is presented in Fields (2002).

Assuming that the model is correctly specified, in the sense that good estimates are obtained for the regression coefficients, this method proposes to decompose the log-variance of income.
Taking the variances of both sides of equation (1.a), we obtain:

$$\text{var}\,(\ln Y_{it}) \;=\; \text{var}\,(a_t{}'Z_{it})$$

LHS is the log-variance, a simple measure of inequality. RHS in this equation can be manipulated, applying Mood, Graybill and Boes theorem[16] to

$$\ln Y \;=\; \sum_{j=1}^{J+2} a_j\, Z_j$$

obtaining

$$\text{cov}\left[\sum_{j=1}^{J+2} a_j\, Z_j,\, \ln Y\right] \;=\; \sum_{j=1}^{J+2} \text{cov}\left[a_j\, Z_j,\ \ln Y\right] \tag{2}$$

Note that LHS of equation (2) is the covariance of *ln Y* and itself. Hence, it is the variance of *ln Y*:

$$\text{var}\,(\ln Y) \;=\; \sum_{j=1}^{J+2} \text{cov}\left[a_j\, Z_j,\ \ln Y\right] \tag{3}$$

Dividing (3) by var (ln *Y*):

$$100\% \;=\; \frac{\sum_{j=1}^{J+2} \text{cov}\left[a_j\, Z_j,\ \ln Y\right]}{\text{var}\,(\ln Y)} \;\equiv\; \sum_{j=1}^{J+2} s_j\,(\ln Y) \tag{4.a}$$

Fields calls each $s_j(\ln Y)$ a *relative factor inequality weight*, where

$$\begin{aligned} s_j(\ln Y) \;&=\; \text{cov}\,[a_j Z_j, \ln Y] \;/\; \text{var}(\ln Y) \\ &=\; a_j {*} \sigma(Z_j) {*} \text{cor}[Z_j, \ln Y] \,/\, \sigma(\ln Y) \end{aligned} \tag{4.b}$$

Note that

$$\sum_{j=1}^{J+1} \text{cov}\left[a_j\, Z_j,\ \ln Y\right] / \,\text{var}\,(\ln Y) \;=\; R^2(\ln Y) \tag{4.c}$$

Then, the fraction explained by the jth explanatory factor $[p_j(\ln Y)]$, within the model is:

[16] Mood, Graybill and Boes theorem says: let $A_1, \ldots, A_P$ y $B_1, \ldots, B_Q$ be two sets of random variables, and $a_1, \ldots, a_P$ and $b_1, \ldots, b_Q$ two sets of constants. Then

$$\text{cov}\left[\sum_{p=1}^{P} a_p\, A_p,\, \sum_{q=1}^{Q} b_q\, B_q\right] \;=\; \sum_{p=1}^{P}\sum_{q=1}^{Q} a_p b_q\ \text{cov}\left[A_p, B_q\right]$$

$$p_j(\ln Y) \equiv \frac{s_j(\ln Y)}{R^2\ (\ln Y)} \qquad (4.d)$$

Given the problems presented by the log-variance as a measure of inequality, it would be useful to extend these results to other inequality measures which satisfy the axioms listed in the literature[17]. Fields (2002) shows this possibility using Shorrocks theorem (1982). His results show that under the six assumptions enumerated by Shorrocks, equations (4.a), (4.b), (4.c) and (4.d) are valid for any index $I(\ln Y_1, \ldots, \ln Y_N)$ that is continuous and symmetric[18] and for which $I(\mu, \mu, \ldots, \mu) = 0$, where μ is mean income. The Gini coefficient, the Atkinson index and the generalised entropy family, among others, fulfil these conditions.

3.2 - Regression model

The following is the linear regression model corresponding to equation (1.a), on which this analysis is based:

$$\begin{aligned}\ln(\text{INCOME}_i) = \; & b_0 + b_1\text{EDUC}_i + b_2\text{AGE}_i + b_3\text{AGE2}_i + b_4\text{CAPACIT}_i + b_5\text{GETNICO}_i + \\ & b_6\text{GENDER}_i + \textstyle\sum b_{j+6}\,\text{OCCUPATION}_{ij} + b_{15}\text{FACTOR}_i + b_{16}\text{REMESAS}_i + \\ & b_{17}\text{OTHER\_NL}_i + b_{18}\text{ASALAR}_i + b_{19}\text{FORMAL}_i + b_{20}\text{AREA}_i + e_i\end{aligned}$$

where,

$\ln(\text{INCOME}_i)$:	natural logarithm of income[19]
EDUC_i:	years of education
AGE_i:	age of the individual
AGE2_i:	AGE squared
CAPACIT_i:	1 = has participated in a training programme, 0 = otherwise
GETNICO_i:	1 = non-indigenous ethnical background, 0 = indigenous background
GENDER_i:	1 = male, 0 = female
OCCUPATION_{ij}:	eight dummies compose this factor[20] (j = 1 to 8)
REMESAS_i:	1 = recipient of remittances, 0 = otherwise
OTHER_NL_i:	1 = recipient of transfer payments and/or inheritance, 0 = otherwise
FORMAL_i[21]:	1 = formal-sector worker, 0 = informal-sector worker
ASALAR_i:	1 = salaried worker, 0 = self-employed (1st. job)
FACTOR_i:	1 = recipient of rent and/or interest payments, 0 = otherwise
AREA_i:	1 = urban, 0 = rural
e_i:	unobserved factors.

[17] See Foster and Sen (1997) for a detailed discussion on the problems presented by the log-variance as an inequality measure and to review the axiomatic approach described in the literature.
[18] By symmetry it is understood that the inequality index satisfies the condition I(Y') = I(Y), where Y' is a permutation of Y (Litchfield 1999).
[19] See Appendix 2 for a detailed composition of the income aggregate.
[20] The categories used are based on data from ENEI 2002 for the worker's first job. These categories are described in the following paragraphs.
[21] According to the classification used in the "Guatemala Poverty Assessment" (GUAPA). This approach classifies workers as formal if they work in the public sector, in a private-sector firm with more than 5 employees, or in a farm with more than 5 workers. The rest of the workers are classified as informal.

The variables EDUC, CAPACIT, AGE and AGE2, partially account for the effect of human capital on productivity, and hence, on labour income, as it was argued earlier.

Due to the lack of specific data on years of experience for each worker, it was considered the use of a variable capturing potential experience ($\text{AGE}_i - 6 - \text{EDUC}_i$) as a proxy[22]. Nevertheless, it is important to observe that potential experience is not a good proxy in the case of women, because it tends to overestimate their participation (in years) in the labour market[23]. This has a direct implication in the underestimation of the returns to experience, and hence, in a systematic underestimation of the contribution of differences in labour experience to the explanation of income inequality[24]. Furthermore, it is common to observe in the Guatemalan society many individuals who decide to work and study at the same time. The dataset does not allow us to identify these individuals. Therefore the use of potential experience was avoided.

Consequently, in order to capture the effect of labour experience, AGE and AGE2 were included in the model. The presence of education and worker training in the set of chosen explanatory variables, allows us to interpret the effect of age as a proxy for experience[25]. The inclusion of the variable AGE2 allows us to consider the hypothesis of diminishing returns to experience. This is to say that human capital depreciates with the years.

There are also problems when considering the effects of education. As it was pointed out in earlier paragraphs, it is important to distinguish between the effects of education quality and the years of education. Thus, it would be optimal for our analysis if we could account for both variables in the educational factor, in order to have a clearer picture of their relative impact in income disparities. However, the lack of specific information in ENEI 2002 regarding the type of education received, in terms of it being public or private, or with respect to any other proxy for education quality, does not allow us to separate the roles of quantity and quality of education[26].

The possible effects of discrimination in the labour markets are captured by the variables GETNICO and GENDER, thereby separating ethnic from gender discrimination.

The potential differentials arising from the existence of compensating wage differentials and/or efficiency wages are attempted to be captured by the use of eight dummy variables representing different occupational categories. These categories, based on the

[22] The ENEI 2002 database has data on worker tenure. Connolly and Gottschalk (2000) include both tenure and experience, in order to separate the effects of general labour experience and the accumulation of industry-specific human capital. For our purposes, we disregard the use of tenure and experience simultaneously due to multicollinearity problems, specifically among agricultural and livestock workers.

[23] This is due to the interruption in labour participation experienced by many women during pre-natal and post-natal periods.

[24] Equation (4.b) clearly shows the dependence of the relative factor inequality weight on the regression coefficient. If the latter is underestimated, *ceteris paribus*, so is the former.

[25] Mincer (1981) proposes the use of age and age-squared as a proxy of experience. Based on evidence from empirical studies that include both age and labour experience, Mincer concludes that the concave profile of the earnings function is mostly due to experience.

[26] In some empirical studies, the student-teacher ratio is used as a proxy for quality of education. ENEI 2002 does not have variables that permit us to calculate this ratio. Other pertinent problems, though more difficult to account for, are late school enrolment and the temporary interruption of the education process. Both of these problems can have significant effects on the returns to education.

classification presented by ENEI 2002, are: (1) Government members and Public Administration personnel, (2) Professional scientists and intellectuals, (3) High school level technicians and professionals, (4) Office employees, (5) Service workers and storekeepers, (6) agricultural, livestock and fishery workers, (7) Officials, operators and artists of mechanical arts and other professions, (8) Machine and installation operators and assemblers, (9) Other (mainly, unskilled workers). Category (9) was used as the excluded one.

The variable ASALAR was used to control for variations between salaried workers and self-employed, in particular to capture the effect of monetary bonuses to which salaried workers are entitled by law.

Moreover, three dummy variables were included to account for non-labour income. FACTOR refers to income coming from rent and interest payments, REMESAS groups income received from remittances, and the variable OTHER_NL gathers income originated from transfers and inheritances.

Finally, the dummy variable FORMAL was included, to take into account the strong dual structure existing in the Guatemalan labour market. The variable AREA also contributes to this framework, allowing us to obtain a general perspective of the importance of urbanisation related factors in explaining income differentials.

A criticism to the previous model is that it assumes equal returns to education and experience for men and women, indigenous and non-indigenous individuals, and for urban and rural areas, despite empirical evidence of the contrary[27]. This criticism seems valid in the case of Guatemala, because there are significant differences in educational attainment between population subgroups[28].

A way of accounting for differences in returns would be to specify interaction variables (e.g. GENDER*EDUC, OCCUPATION*EDUC, etc.). Nevertheless, the inclusion of such a type of variables restricts our interpretation of the contribution percentages. The reason why is because Fields decomposition technique explains the variation of each explanatory variable separately, but when the variable is an interaction variable, the effect measured cannot be differentiated between the two factors. Hence, it would be possible to talk about the joint effect of gender and education, but it would not be possible to separate both effects.

Since we are interested in the individual effect of each factor, the use of interaction variables was neglected. As an alternative, we considered to repeat the analysis for different population subgroups. This is done in Section 4.

[27] The estimated returns to education in Guatemala are 3% for men and 6% for women (World Bank, 2003).

[28] See Chart A.1.2, in Appendix 1.

4. Empirical Results

Before presenting the empirical findings of this analysis, it is important to remark that the use of cross-section data limits the conclusions provided by this sort of study. Particularly, it is impossible to have a vision of the *evolution* of the contribution of each determinant of income inequality, which would be useful to distinguish structural inequality from temporal inequality.

This section proceeds as follows: first, the results of the analysis at the national level are presented; then, four subgroups of the population are analysed independently. The subgroups are based on economic activity and employment characteristics.

4.1 - Analysis at the national level

Using the model presented in section 3.2, the following results were obtained:

Observations:	4691	R-squared:	0.4792
F(20, 4671) :	148.98	Dependent variable:	ln(income)
Prob > F :	0.0000		

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.059008	0.005018	11.759	0.000	0.049170	0.068846
age	0.052801	0.005490	9.618	0.000	0.042038	0.063563
age2	-0.000553	0.000068	-8.171	0.000	-0.000685	-0.000420
capacit	0.106941	0.053552	1.997	0.046	0.001953	0.211928
getnico	0.210378	0.039189	5.368	0.000	0.133549	0.287207
gender	0.561376	0.036862	15.229	0.000	0.489108	0.633643
occu1	1.033923	0.074838	13.815	0.000	0.887205	1.180642
occu2	0.577242	0.065699	8.786	0.000	0.448441	0.706043
occu3	0.365201	0.080650	4.528	0.000	0.207090	0.523312
occu4	0.365830	0.065209	5.610	0.000	0.237990	0.493669
occu5	0.329050	0.050193	6.556	0.000	0.230647	0.427452
occu6	-0.496401	0.076123	-6.521	0.000	-0.645638	-0.347163
occu7	0.181827	0.052282	3.478	0.001	0.079329	0.284325
occu8	0.349774	0.069629	5.023	0.000	0.213268	0.486279
factor	0.497714	0.084192	5.912	0.000	0.332658	0.662770
remesas	0.706222	0.095249	7.414	0.000	0.519490	0.892955
other_nl	0.307027	0.065028	4.721	0.000	0.179540	0.434513
asalar	0.185267	0.042620	4.347	0.000	0.101713	0.268822
formal	0.396152	0.037657	10.520	0.000	0.322325	0.469978
area	0.217002	0.034981	6.203	0.000	0.148422	0.285581
_constant	4.454802	0.118032	37.742	0.000	4.223404	4.686200

Note: Calculations using the individual sample weights calculated by INE in the ENEI 2002. Robust standard errors are reported.

In general, the regression coefficients show the expected trends: conditional on the other factors, men earn more than women, non-indigenous workers more than indigenous workers, independent workers less than non-independent employees, workers in the formal sector more than those in the informal sector, and finally, income appears to be higher for workers living in urban areas as opposed to those living in rural areas.

Moreover, the existence of a concave age-earnings profile, in which human capital depreciates gradually, is justified by the results.

Table 4.1 shows the contribution of the determinants of income inequality included in the model, for the whole sample.

Table 4.1 – Contribution of the determinants of income inequality at a national level

Education	12.35%
Experience	1.84%
Training	0.77%
Ethnicity	2.70%
Gender	2.98%
Occupation	15.41%
Non-labour income	2.37%
Salaried workers vs. Self-employed	2.52%
Formal vs. Informal	3.59%
Area	3.39%
Total explained (= R^2)	**47.92%**

According to the calculations made based on equation (4.b), which was presented in section 3.1, the factor with greatest contribution is the individual's occupation. The differences in occupation explain up to 15.4% of income variation. The second most important factor is education, which accounts for 12.4%. This is a somewhat surprising result, because a greater contribution was expected based on empirical evidence for developing countries[29]. Nonetheless, it is necessary to indicate that education is an important determinant of worker's occupation[30].

The joint contribution of labour experience and training is very modest according to the results. Only 2.6% of income inequality appears to be explained by these factors. Adding the effect of education, we find out that the accounted differences in human capital are significant, contributing to 15% of income differentials.

Discrimination is also significant, having a contribution of approximately 6%, which is shared equally between ethnic and gender discrimination. Although a larger participation may have been expected, the shown results only account for the direct effect of discrimination, leaving aside its indirect effects.

The income gap between salaried workers and self-employed is responsible of 2.5% of income variations in the sample, while 3.6% is attributed to the existing disparities between the formal and informal sectors.

[29] Fields et al. (1998), for instance, find a contribution of 20% in urban Bolivia, representing 79% of the inequality explained by their model. A similar result is presented by Gindling and Trejos (2003), who also find a contribution of 20%, this time with data from Costa Rica.

[30] Without attempting to provide a precise explanation of the determinants of occupation in Guatemala, a logit model was run to observe some patterns. The results, presented in Table A.1.1, Appendix 1, show that the more educated a worker is the lower the probability that he will be employed in the agricultural sector, which has the lowest income average in the sample and accounts for more than one third of the percentage attributed to occupational variation.

Furthermore, the differences in the distribution of non-labour income contribute with 2.4%. Among the three categories used to decompose non-labour income, factor payments have the largest participation, explaining 1.1% of income differentials.

Finally, the area of living also shows to be important, determining 3.4% of the variation in income.

4.2 - Analysis by socio-economic groups

4.2.1 - Non-agricultural salaried workers in the formal sector

As it can be observed in the results below, the total explanatory capability of the model seems to improve significantly when analysing the sample of non-agricultural salaried workers in the formal sector only.

Observations:	757	R-squared:	0.5391
F(17, 739) :	28.84	Dependent Variable:	ln(income)
Prob > F :	0.0000		

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.059056	0.007243	8.154	0.000	0.044838	0.073275
age	0.032586	0.012682	2.569	0.010	0.007689	0.057482
age2	*-0.000292*	*0.000157*	*-1.860*	*0.063*	*-0.000599*	*0.000016*
capacit	*0.040041*	*0.052300*	*0.766*	*0.444*	*-0.062633*	*0.142714*
getnico	*0.020967*	*0.071687*	*0.292*	*0.770*	*-0.119768*	*0.161702*
gender	0.110104	0.052786	2.086	0.037	0.006475	0.213733
occu1	0.792344	0.136420	5.808	0.000	0.524526	1.060161
occu2	0.418521	0.123728	3.383	0.001	0.175620	0.661421
occu3	0.333337	0.113598	2.934	0.003	0.110323	0.556351
occu4	*0.106488*	*0.113569*	*0.938*	*0.349*	*-0.116468*	*0.329443*
occu5	*0.150660*	*0.097808*	*1.540*	*0.124*	*-0.041353*	*0.342674*
occu7	*0.113829*	*0.090277*	*1.261*	*0.208*	*-0.063402*	*0.291059*
occu8	*0.179211*	*0.143053*	*1.253*	*0.211*	*-0.101627*	*0.460049*
factor	0.386240	0.102879	3.754	0.000	0.184269	0.588210
remesas	0.437920	0.120738	3.627	0.000	0.200889	0.674950
other_nl	0.263444	0.073860	3.567	0.000	0.118443	0.408444
area	0.152873	0.069855	2.188	0.029	0.015735	0.290010
_constant	6.023536	0.260571	23.117	0.000	5.511989	6.535084

Note: Calculations using the individual sample weight calculated by INE in the ENEI 2002. Robust standard errors are reported. Non-statistically significant variables at the 5% level appear in *italics*.

Nonetheless, many variables appear to be non-statistically significant. This is the case for worker's participation in training programmes and various occupational categories. An interesting result is the apparent non-significance of the variable capturing ethnic discrimination, as well as the rejection of the hypothesis stating that human capital depreciates with the years.

These patterns are clearly captured by the percentage participation of these variables, as shown in Table 4.2.1:

Table 4.2.1 – Contribution of the determinants of income inequality for non-agricultural salaried workers in the formal sector

Education	25.29%
Experience	5.92%
Training	0.80%
Ethnicity	0.24%
Gender	-0.02%
Occupation	13.84%
Non-labour Income	4.06%
Area	3.78%
Total explained (= R^2)	**53.91%**

In fact, the contribution of ethnic discrimination is insignificant. Likewise, we surprisingly find that gender discrimination does not explain income differentials for this group.

Education is clearly the most important determinant within individuals belonging to this group. As much as 25.3% of the total variation appears to be due to this factor. Worker experience also has a bigger contribution, which adds up to 5.9%, significantly larger than the national estimate.

The occupational factor remains of great importance, capturing 13.8% of income differences, while the level of urbanisation accounts for 3.8%.

It is interesting to observe the jump in the contribution of non-labour income to total inequality. Of the 4.1% it represents, 2.2% are due to differences in the reception of rent and interest payments. This suggests that this source of income has a larger concentration among these workers.

It is important to remark the importance of human capital as a whole. Education, experience and worker training explain together 32.0% of the measured income inequality for this group, more than twice the significance of the same factor at a country level.

4.2.2 - Non-agricultural salaried workers in the informal sector

Contrary to the regression results shown for the previous group, all the explanatory variables, with the exception of non-labour income coming from transfer payments, are statistically significant.

The returns to education and its participation are lower, while the contribution of experience and worker training rise significantly. Therefore, human capital accounts for 21.6% of income inequality.

Observations:		1914		R-squared:		0.4511
F(17, 1896) :		61.95		Dependent Variable:		ln(income)
Prob > F :		0.0000				

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.042412	0.006585	6.440	0.000	0.029497	0.055328
age	0.072562	0.007802	9.301	0.000	0.057261	0.087862
age2	-0.000821	0.000101	-8.102	0.000	-0.001020	-0.000622
capacit	0.208408	0.068648	3.036	0.002	0.073774	0.343042
getnico	0.246964	0.045897	5.381	0.000	0.156951	0.336978
gender	0.370408	0.050785	7.294	0.000	0.270808	0.470008
occu1	0.999250	0.091044	10.975	0.000	0.820692	1.177807
occu2	0.582435	0.079534	7.323	0.000	0.426451	0.738418
occu3	0.463300	0.108290	4.278	0.000	0.250920	0.675679
occu4	0.511299	0.077781	6.574	0.000	0.358754	0.663844
occu5	0.393602	0.064388	6.113	0.000	0.267324	0.519881
occu7	0.158000	0.070390	2.245	0.025	0.019950	0.296051
occu8	0.368746	0.078301	4.709	0.000	0.215180	0.522312
factor	0.361660	0.115482	3.132	0.002	0.135175	0.588146
remesas	0.448778	0.119749	3.748	0.000	0.213925	0.683631
other_nl	*0.108176*	*0.087060*	*1.243*	*0.214*	*-0.062567*	*0.278920*
area	0.176986	0.045009	3.932	0.000	0.088715	0.265258
_constant	4.538772	0.139574	32.519	0.000	4.265038	4.812506

Note: Calculations using the individual sample weight calculated by INE in the ENEI 2002. Robust standard errors are reported. Non-statistically significant variables at the 5% level appear in *italics*.

Once more, occupation is the single most important factor, explaining 13.8% of the differentials. Ethnic discrimination accounts for 3.7%, its largest level among the groups studied, while gender discrimination has a lower contribution (1.3%).

Table 4.2.2 – Contribution of the determinants of income inequality for non-agricultural salaried workers in the informal sector

Education	10.72%
Experience	8.58%
Training	2.43%
Ethnicity	3.68%
Gender	1.29%
Occupation	13.75%
Non-Labour Income	1.74%
Area	2.92%
Total explained (= R^2)	**45.11%**

The participation of non-labour income falls to only 1.7%, whereas 2.9% appear to be due to urbanisation-related factors.

4.2.3 - Non-agricultural self-employed workers

This subgroup shows interesting results. The huge percentage attributed to gender inequality stands out[31].

Observations:	1448	R-squared:	0.4616
F(17, 1430) :	52.30	Dependent Variable:	ln(income)
Prob > F :	0.0000		

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.063636	0.009319	6.829	0.000	0.045356	0.081917
age	0.043918	0.011014	3.987	0.000	0.022313	0.065523
age2	-0.000488	0.000126	-3.861	0.000	-0.000735	-0.000240
capacit	*0.213295*	*0.114779*	*1.858*	*0.063*	*-0.011857*	*0.438447*
getnico	*0.132043*	*0.072243*	*1.828*	*0.068*	*-0.009671*	*0.273756*
gender	0.923422	0.069127	13.358	0.000	0.787820	1.059024
occu1	1.138629	0.141799	8.030	0.000	0.860472	1.416786
occu2	0.868359	0.199771	4.347	0.000	0.476484	1.260234
occu3	*0.222371*	*0.240388*	*0.925*	*0.355*	*-0.249179*	*0.693922*
occu4	*0.321530*	*0.466272*	*0.690*	*0.491*	*-0.593121*	*1.236180*
occu5	0.398567	0.109104	3.653	0.000	0.184546	0.612587
occu7	*0.204285*	*0.116129*	*1.759*	*0.079*	*-0.023516*	*0.432086*
occu8	0.390535	0.158429	2.465	0.014	0.079756	0.701313
factor	0.545115	0.094519	5.767	0.000	0.359705	0.730525
remesas	0.883942	0.175214	5.045	0.000	0.540238	1.227645
other_nl	0.354219	0.084884	4.173	0.000	0.187709	0.520729
area	0.234836	0.069559	3.376	0.001	0.098388	0.371285
_constant	4.493640	0.237406	18.928	0.000	4.027938	4.959342

Note: Calculations using the individual sample weight calculated by INE in the ENEI 2002. Variable FORMAL was omitted due to too small variation. Robust standard errors are reported. Non-statistically significant variables at the 5% level appear in *italics*.

Educational disparities explain 11.1%, a percentage similar to that of the previous group, though education is more important than occupation in this context.

Table 4.2.3 – Contribution of the determinants of income inequality for non-agricultural self-employed workers

Education	11.14%
Experience	1.35%
Training	0.73%
Ethnicity	1.30%
Gender	16.89%
Occupation	8.33%
Non-labour Income	3.73%
Area	2.69%
Total explained (= R^2)	**46.16%**

[31] This sudden result seems suspicious. However, the contribution of this variable remains large even when changing the specification of the model.

The impact of experience (1.4%) and training (0.7%) drops down to national levels, lowering the contribution of human capital (13.2%).

Ethnical factors are responsible of a low 1.3%, non-labour income differentials account for 3.7% and the area of living explains 2.7% of income variation.

4.2.4 - Agricultural, livestock and fishery workers

As it can be seen in the results below, this group is the most heterogeneous from all groups studied. The capability of our model to explain income differentials is very limited. Only 17.0% of this variation is captured by the explanatory variables in the model[32].

Observations:	572	R-squared:	0.1695
F(7, 564) :	16.44	Dependent Variable:	ln(income)
Prob > F :	0.0000		

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.072952	0.021566	3.383	0.001	0.030593	0.115310
age	0.066727	0.016313	4.090	0.000	0.034686	0.098769
age2	-0.000670	0.000170	-3.951	0.000	-0.001003	-0.000337
getnico	0.305560	0.117727	2.596	0.010	0.074323	0.536796
gender	0.723051	0.162798	4.441	0.000	0.403287	1.042816
nl	0.769387	0.162074	4.747	0.000	0.451045	1.087730
area	0.403809	0.111002	3.638	0.000	0.185781	0.621836
_constant	3.338512	0.415046	8.044	0.000	2.523288	4.153736

Note: Calculations using the individual sample weight calculated by INE in the ENEI 2002. Variables FORMAL, ASALAR, FACTOR, REMESAS and OTHER_NL were omitted due to small levels of variation. Variable NL (1=recipient of non-labour income, 0=otherwise) was generated to capture the effect of non-labour income. Robust standard errors are reported.

As shown in Table 4.2.4, one of the main sources of the low capability of explanation is the drop in the contribution attributable to education (2.9%). This is not surprising, because the average worker in this group has only two years of formal education, with many reporting not to have had any sort of formal education whatsoever.

Consequently, the low level of human capital accumulated through education is reflected in its significantly low contribution. Nonetheless, education and experience jointly explain 5.6%, which accounts for approximately one third of the total explained by the model.

[32] It is important to notice that income for agricultural, livestock and fishery workers may be very volatile. Therefore, the use of cross-section data to analyse income variations in this group imposes serious restrictions in our results. Furthermore, the sample size we observe is relatively small. This combined with the omission of self-consumption may significantly bias the estimates presented. Thus, the results shown in this section are only an attempt to characterise the inequality structure within this group and should be carefully interpreted.

Non-labour income, on the other hand, reaches its maximum contribution among the studied groups (5.5%) and is the single most important determinant of the income inequality shown among these individuals.

Ethnic differences account for 2.0% and gender disparities 3.0%. Urbanisation related factors, also unsurprisingly, contribute with only 1.2% of the total variation.

Table 4.2.4 – Contribution of the determinants of income inequality for agricultural, livestock and fishery workers

Education	2.91%
Experience	2.69%
Ethnicity	1.99%
Gender	2.98%
Non-labour Income	5.14%
Area	1.24%
Total explained (= R^2)	**16.95%**

It is interesting the fact that the variables CAPACIT and FACTOR could not be included in the regression to small variation[33]. In the first case, this may reflect the limited access to (and perhaps supply of) training programmes for these economic activities. In the second case, it confirms the trend expected *a priori*: factors generating rent and interest payments are not concentrated in this group, hence its lack proving an obstacle for the development of this group.

[33] In the sample of agricultural, livestock and fishery workers only 25 individuals reported to have participated in training programmes, while only 9 says to be a recipient of rent and interest payments.

5. Conclusions

The previous results are summarised in Table 5.1. Furthermore, Table 5.2 shows similar results of the same analysis carried out for gender, ethnic and area groups[34]. These outcomes exemplify the complexity of analysing the relative contribution of the determinants of income inequality. As it was observed in Section 1, it is important to focus not only in the inequality at the national level, but also within each particular group. The differences in contribution of each determinant among these groups are evident. The most heterogeneous group, with respect to the rest of the sample, is that of agricultural, livestock and fishery workers, in which the capability of the model to explain income variation is very limited.

This heterogeneity is of great relevance for the design of policies to reduce the inequality of income across the country. For instance, while in some of the groups the proportion of the differentials attributed to ethnic and gender discrimination are important, in other –as it is the case of non-agricultural salaried workers employed in the formal sector–, this discrimination seems not to exist.

It is necessary to interpret these results carefully. The low contribution of education for the agricultural, livestock and fishery workers group does not mean that its importance for their development is limited, neither that its fomentation will not help to reduce the gap between these workers and the other groups.

Rather, the mentioned low sharing indicates the relative equality in educational attainment within this group. Unfortunately, this relative equality is not due to high enrolment rates, but to the low average schooling[35].

Table 5.1 – Contribution of the determinants of income inequality by occupational categories

Variable	National	Non-agricultural workers			Agricultural, livestock and fishery workers
		Salaried Workers		Self-employed	
		Formal	Informal		
Education	12.35%	25.29%	10.72%	11.14%	2.91%
Experience	1.84%	5.92%	8.58%	1.35%	2.69%
Training	0.77%	0.80%	2.43%	0.73%	
Ethnicity	2.70%	0.24%	3.68%	1.30%	1.99%
Gender	2.98%	-0.02%	1.29%	16.89%	2.98%
Occupation	15.41%	13.84%	13.75%	8.33%	
Non-labour Income	2.37%	4.06%	1.74%	3.73%	5.14%
Salaried workers vs. Self-employed	2.52%				
Formal vs. Informal	3.59%				
Area	3.39%	3.78%	2.92%	2.69%	1.24%
Total explained (= R^2)	**47.92%**	**53.91%**	**45.11%**	**46.16%**	**16.95%**
Variance (lnY)	*1.3583*	*0.5538*	*0.7942*	*1.5144*	*1.5123*

[34] The respective regressions leading to Table 5.2 are presented in Appendix 1.

[35] Only two years per capita, according to the sample. Almost 50% of the workers in this group reports not to have received any sort of formal education.

Moreover, it is important to say that education, through its role as a determinant of occupation, may have a larger participation as a potential agent of change in the reduction of income inequality in Guatemala.

In general, the variation in the percentages for the same factor across groups shows to a certain degree its concentration in these groups. For instance, the relatively high levels found for non-labour income among salaried workers, in the formal sector, reflect the higher concentration of factors generating rent and interest in this group, while it suggests the growing importance of remittances for agricultural, livestock and fishery workers.

Although the presented results seem consistent with the trends expected *a priori*[36], this analysis, rather than procuring a final answer and a precise guide to the problem of the relative contribution of income inequality determinants, pretends to open a discussion on the inter-sectoral analysis of this phenomenon. Moreover, it places particular emphasis in the structural differences of this inequality within the various socio-economic groups that coexist in the Guatemalan economy, as well as the significance of their understanding for the design of social policies.

Table 5.2 – Contribution of the determinants of income inequality by area, gender and ethnicity

Variable	Area		Gender		Ethnicity	
	Urban	Rural	Male	Female	Indigenous	Non-indigenous
Education	16.19%	5.04%	10.47%	14.23%	5.84%	15.28%
Experience	4.95%	0.53%	2.68%	0.88%	0.49%	3.41%
Training	1.54%	0.19%	0.71%	1.13%	0.27%	1.22%
Ethnicity	1.67%	2.52%	2.58%	1.99%		
Gender	4.08%	5.73%			5.30%	2.32%
Occupation	12.07%	12.47%	24.36%	11.47%	15.09%	14.01%
Non-labour Income	2.72%	2.61%	2.57%	3.09%	2.19%	2.52%
Salaried workers vs. Self-employed	1.48%	2.84%	-0.18%	6.06%	1.74%	2.59%
Formal vs. Informal	4.24%	2.45%	3.19%	5.38%	3.13%	3.54%
Area			2.22%	5.16%	3.06%	2.86%
Total explained (= R^2)	**48.95%**	**34.37%**	**48.60%**	**49.39%**	**37.11%**	**47.75%**
Variance (lnY)	*1.1184*	*1.2248*	*1.2744*	*1.4540*	*1.2843*	*1.1971*

[36] The precision of the percentages shown is doubtful, given the limitations imposed by the available data. However, the relative trends they imply may be useful to provide some guidance for future research.

References

Baumeister, E. (2003) "*Tierra, Empleo e Ingresos de la Población Rural en Guatemala*". Cuadernos de Desarrollo Humano / N° 2002-8. Guatemala: Sistema de Naciones Unidas.

Becker, G. S. (1962) "Investment in Human Capital: A theoretical analysis". *Journal of Political Economy.*

Bigsten, A. and Levin, J. (2000) "*Growth, Income Distribution, and Poverty: A Review*". Working Paper in Economics No 32. Göteborg: Göteborg University

Carneiro, P. and Heckman, J. (2003) "Human Capital Policy", forthcoming in J. Heckman, A. Krueger, *Inequality in America: What Role for Human Capital Policies*, MIT Press.

Connolly, H. and Gottschalk, P. (2000) "*Returns to Tenure and Experience Revisited – Do Less Educated Workers Gain Less from Work Experience?*"
http://fmwww.bc.edu/ec-p/wp473.pdf

Cowell, F. A. and Jenkins, S. P. (1995) "How much inequality can we explain? A methodology and an application to the United States". *The Economic Journal.*

Deininger, K. and Squire, L. (1996) "*Measuring Income Inequality: A New Database*" HIID Development Discussion Paper No. 537, Harvard University.

Fields, G. S., Leary, J., López Calva, L., and Pérez de Rada, E. (1998) "*Education's Crucial Role in Explaining Labor Income Inequality in Urban Bolivia*". HIID Development Discussion Paper No. 658. Harvard University.

Fields, G. S. (2002) "*Accounting for income inequality and its change: a new method, with application to the distribution of earnings in the United States*".
http://www.ilr.cornell.edu/extension/files/20021024104432-pub785.pdf

Foster, J. and Sen, A. (1997) "*On economic inequality*". Expanded edition. Oxford: Oxford University Press.

García Mancilla, C. (2003) "*Los determinantes del empleo en las microempresas en Guatemala*". Preliminary version. Forthcoming, Instituto de Investigaciones Económicas y Sociales (IDIES).

Gindling, T.H. and Trejos, J.D. (2003) "*Accounting for Changing Earnings Inequality in Costa Rica in the Final Quarter of the 20th Century*".
http://research.umbc.edu/~tgindlin/GindlingTrejos.doc

Hild, M. and Voorhoeve, A. (2003) "*Equality of Opportunity and Opportunity Dominance*".
http://www.ucl.ac.uk/~uctyaev/hild3.pdf

IFAD (2001) "*Assessment of Rural Poverty. Latin America and the Caribbean*". Santiago, Chile.

Informe de Desarrollo Humano 2000. "*Guatemala: la fuerza incluyente del desarrollo humano*". Guatemala: Sistema de Naciones Unidas.

Jusidman, C. (2002) "*Equidad de género y transformación productiva*". Cuadernos de Desarrollo Humano / N° 2002-5. Guatemala: Sistema de Naciones Unidas.

Lanfranchi, J., Ohlsson, H. and Skalli, A. (2001) "*Compensating wage differentials and shift work preferences: Evidence from France*". Working Papers in Economics No. 55. Göteborg: Göteborg University

Lewis, W.A. (1954) "*Economic Development with Unlimited Supplies of Labor*". Manchester School, 22: 139-191.

Litchfield, J. (1999) "*Inequality: Methods and Tools*". Text for World Bank's Web Site on Inequality, Poverty, and Socio-economic Performance: http://www.worldbank.org/poverty/inequal/index.htm

Martínez Peláez, Severo (1971) "*La patria del criollo*". Guatemala: Editorial Universitaria.

McConnell, Campbell, R. and Brue, S. L. (1997) "Economía Laboral Contemporánea" 4ª. edición. McGraw-Hill.

Mincer, J. (1958) "Investment in Human Capital and Personal Income Distribution". *Journal of Political Economy*.

Mincer, J. (1974) "*Schooling, Experience, and Earnings*". New York: National Bureau of Economic Research.

Mincer, J. (1981) "*Human Capital and Economic Growth*". Working Paper No. 803. Cambridge: National Bureau of Economic Research.

Psacharopoulos, G. (1995) "*The Profitability of Investment in Education: Concepts and Methods*". Policy HCO Working Paper. Human Capital Development and Operations.
http://www.worldbank.org/education/economicsed/research/wbpub/psacharopoulos.pdf

Shorrocks, A.F. (1982) "Inequality Decomposition by Factor Components". *Econometrica*.

Schultz, T.W. (1960) "Capital Formation by Education". *Journal of Political Economy.*

Spiegel, M. R. (1997) "*Estadística*". Segunda Edición. México: McGraw-Hill.

World Bank (2003) "*La Pobreza en Guatemala*". Informe N°24221-GU.

A.1 - Statistical Appendix

Table A.1.1 – Demographic and urbanisation characteristics of the sample studied

Area	Gender		Ethnicity	
	Male	Female	Indigenous	Non-indigenous
Urban	58.31%	41.69%	28.45%	71.55%
Rural	72.65%	27.35%	50.07%	49.93%
Total	66.17%	33.83%	40.30%	59.70%

Note: Calculations using the individual sample weights calculated by INE for the ENEI 2002.

Table A.1.2 – Net enrolment rates, by level and group, in Guatemala

	Pre-Primary			Primary			Secondary		
	All	Male	Female	All	Male	Female	All	Male	Female
Total (%)	23	22	25	79	81	76	25	26	24
Non-indigenous	27	27	28	84	71	86	32	32	33
Indigenous	18	16	20	75	82	67	14	18	11
Urban	35	32	38	85	88	82	46	48	44
Rural	17	17	18	75	78	72	12	14	10

Source: World Bank calculations using the ENCOVI 2000, Guatemalan National Institute of Statistics – 2000

Table A.1.3 – Occupational characteristics within the sample

Variable	Occupation	(%)
occu1	Government members and Public Administration personnel	3.25
occu2	Professional scientists and intellectuals	6.40
occu3	High school level technicians and professionals	2.58
occu4	Office employees	4.17
occu5	Service workers and merchants	17.64
occu6	Agricultural, livestock and fishery workers	19.62
occu7	Officials, operators and artists of mechanical arts and other professions	20.80
occu8	Machine and installation operators and assemblers	4.42
occu9	Other (unskilled workers)	21.12
	Total	100.00

Note: Calculations using the individual sample weights calculated by INE for the ENEI 2002.

Table A.1.4 – The role of education as a determinant of occupation

Observations:	4715	Pseudo R-squared:	0.2810
LR chi2(5) :	1323.13	Dependent Variable[a]:	occu6
Prob > F :	0.0000		

Variable	Coefficient	Standard error	z	P>z	95% Confidence interval	
educ	-0.131186	0.015063	-8.709	0.000	-0.160710	-0.101663
age	0.038792	0.002909	13.336	0.000	0.033090	0.044493
getnico	-0.547820	0.087942	-6.229	0.000	-0.720184	-0.375456
gender	1.789826	0.121530	14.728	0.000	1.551633	2.028020
area	-1.687921	0.120078	-14.057	0.000	-1.923271	-1.452572
_constant	-3.002110	0.172181	-17.436	0.000	-3.339578	-2.664641

Note: Calculations using the individual sample weight calculated by INE for the ENEI 2002.
[a] occu6 = agricultural, livestock or fishery worker (1 = yes / 0 = no)

A.1.5 – Regression analysis of individuals by gender, ethnicity and area of living

I – Analysis by Gender

A. Male

Observations:	2853	R-squared:	0.4860
F(20, 4671) :	99.13	Dependent variable:	ln(income)
Prob > F :	0.0000		

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.051608	0.006309	8.18	0.000	0.039237	0.063978
age	0.062535	0.006801	9.195	0.000	0.049200	0.075871
age2	-0.000655	0.000084	-7.812	0.000	-0.000819	-0.000490
capacit	*0.102827*	*0.074976*	*1.371*	*0.170*	*-0.044187*	*0.249841*
getnico	0.205481	0.047076	4.365	0.000	0.113174	0.297789
occu1	1.066646	0.090389	11.801	0.000	0.889411	1.243881
occu2	0.612282	0.080671	7.590	0.000	0.454102	0.770463
occu3	0.364484	0.098568	3.698	0.000	0.171211	0.557756
occu4	0.323080	0.079148	4.082	0.000	0.167887	0.478273
occu5	0.418616	0.061402	6.818	0.000	0.298218	0.539013
occu6	-0.653410	0.083591	-7.817	0.000	-0.817315	-0.489506
occu7	0.316963	0.054765	5.788	0.000	0.209581	0.424346
occu8	0.387982	0.075819	5.117	0.000	0.239317	0.536647
factor	0.481745	0.102172	4.715	0.000	0.281406	0.682084
remesas	0.572320	0.131410	4.355	0.000	0.314652	0.829988
other_nl	0.377352	0.095051	3.970	0.000	0.190976	0.563727
asalar	*-0.015563*	*0.048959*	*-0.318*	*0.751*	*-0.111562*	*0.080437*
formal	0.351289	0.045425	7.733	0.000	0.262219	0.440358
Area	0.129701	0.039167	3.312	0.001	0.052904	0.206499
_constant	5.017493	0.141157	35.546	0.000	4.740712	5.294273

Note: Calculations using the individual sample weight calculated by INE in the ENEI 2002. Robust standard errors are reported. Non-statistically significant variables at the 5% level appear in *italics*.

B. Female

Observations:		1838	R-squared:			0.4939
F(20, 4671) :		70.59	Dependent variable:			ln(income)
Prob > F :		0.0000				

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.060536	0.008494	7.127	0.000	0.043876	0.077195
age	0.040943	0.009746	4.201	0.000	0.021829	0.060057
age2	-0.000428	0.000123	-3.471	0.001	-0.000670	-0.000186
capacit	0.133718	0.062952	2.124	0.034	0.010252	0.257184
getnico	0.140205	0.065798	2.131	0.033	0.011158	0.269252
occu1	0.954172	0.123887	7.702	0.000	0.711197	1.197147
occu2	0.442113	0.107792	4.102	0.000	0.230704	0.653523
occu3	0.332161	0.126378	2.628	0.009	0.084299	0.580023
occu4	0.314087	0.102654	3.060	0.002	0.112754	0.515419
occu5	0.212995	0.087458	2.435	0.015	0.041467	0.384523
occu6	-0.551844	0.180930	-3.050	0.002	-0.906697	-0.196991
occu7	*-0.073821*	*0.103692*	*-0.712*	*0.477*	*-0.277189*	*0.129548*
occu8	*0.282500*	*0.164765*	*1.715*	*0.087*	*-0.040650*	*0.605649*
factor	0.655504	0.109596	5.981	0.000	0.440557	0.870451
remesas	0.983484	0.143264	6.865	0.000	0.702505	1.264463
other_nl	0.246338	0.082575	2.983	0.003	0.084387	0.408289
asalar	0.376690	0.077356	4.870	0.000	0.224974	0.528406
formal	0.595222	0.070085	8.493	0.000	0.457767	0.732676
area	0.320296	0.063543	5.041	0.000	0.195672	0.444920
_constant	4.659558	0.204088	22.831	0.000	4.259286	5.059830

Note: Calculations using the individual sample weight calculated by INE in the ENEI 2002. Robust standard errors are reported. Non-statistically significant variables at the 5% level appear in *italics*.

II – Analysis by Ethnicity

A. Indigenous individuals

Observations:	1529	R-squared:	0.3711
F(20, 4671) :	44.65	Dependent variable:	ln(income)
Prob > F :	0.0000		

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.050031	0.010283	4.865	0.000	0.029861	0.070202
age	0.036121	0.009409	3.839	0.000	0.017665	0.054577
age2	-0.000372	0.000116	-3.208	0.001	-0.000600	-0.000145
capacit	*0.077447*	*0.144751*	*0.535*	*0.593*	*-0.206489*	*0.361382*
gender	0.673843	0.069006	9.765	0.000	0.538486	0.809201
occu1	1.146133	0.125120	9.160	0.000	0.900706	1.391560
occu2	0.687055	0.122174	5.624	0.000	0.447407	0.926704
occu3	*0.275662*	*0.171299*	*1.609*	*0.108*	*-0.060347*	*0.611672*
occu4	0.426654	0.141446	3.016	0.003	0.149203	0.704106
occu5	0.460431	0.082440	5.585	0.000	0.298721	0.622140
occu6	-0.552586	0.109763	-5.034	0.000	-0.767890	-0.337281
occu7	*0.153598*	*0.083543*	*1.839*	*0.066*	*-0.010274*	*0.317471*
occu8	0.587116	0.102876	5.707	0.000	0.385322	0.788910
factor	0.617956	0.133859	4.616	0.000	0.355388	0.880525
remesas	0.915520	0.172569	5.305	0.000	0.577019	1.254021
other_nl	0.260814	0.106046	2.459	0.014	0.052801	0.468827
asalar	0.142807	0.071484	1.998	0.046	0.002588	0.283026
formal	0.576789	0.083116	6.940	0.000	0.413753	0.739825
area	0.274309	0.056458	4.859	0.000	0.163564	0.385053
_constant	4.701628	0.193058	24.353	0.000	4.322937	5.080319

Note: Calculations using the individual sample weight calculated by INE in the ENEI 2002. Robust standard errors are reported. Non-statistically significant variables at the 5% level appear in *italics*.

B. Non-indigenous individuals

Observations:		3162		R-squared:		0.4775
F(20, 4671) :		93.86		Dependent variable:		ln(income)
Prob > F :		0.0000				

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.064153	0.005704	11.248	0.000	0.052970	0.075336
age	0.064012	0.006552	9.770	0.000	0.051166	0.076858
age2	-0.000676	0.000079	-8.592	0.000	-0.000830	-0.000521
capacit	0.132313	0.053072	2.493	0.013	0.028253	0.236373
gender	0.485749	0.040816	11.901	0.000	0.405720	0.565778
occu1	0.974457	0.092292	10.558	0.000	0.793499	1.155416
occu2	0.506122	0.080207	6.310	0.000	0.348860	0.663384
occu3	0.368415	0.091629	4.021	0.000	0.188757	0.548073
occu4	0.329841	0.075672	4.359	0.000	0.181469	0.478213
occu5	0.254941	0.063594	4.009	0.000	0.130251	0.379631
occu6	-0.405465	0.111053	-3.651	0.000	-0.623209	-0.187720
occu7	0.217704	0.064931	3.353	0.001	0.090394	0.345015
occu8	0.286437	0.089691	3.194	0.001	0.110578	0.462295
factor	0.466932	0.095774	4.875	0.000	0.279146	0.654717
remesas	0.592450	0.105665	5.607	0.000	0.385271	0.799630
other_nl	0.331632	0.078048	4.249	0.000	0.178602	0.484661
asalar	0.218185	0.053414	4.085	0.000	0.113456	0.322915
formal	0.338535	0.041890	8.082	0.000	0.256402	0.420669
area	0.176526	0.043632	4.046	0.000	0.090976	0.262077
_constant	4.486752	0.150207	29.871	0.000	4.192239	4.781265

Note: Calculations using the individual sample weight calculated by INE in the ENEI 2002. Robust standard errors are reported. Non-statistically significant variables at the 5% level appear in *italics*.

II – Analysis by Area of living

A. Urban area

Observations:	3487	R-squared:	0.4895
F(20, 4671) :	127.54	Dependent variable:	ln(income)
Prob > F :	0.0000		

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.062645	0.004529	13.833	0.000	0.053766	0.071524
age	0.066452	0.005548	11.977	0.000	0.055573	0.077330
age2	-0.000715	0.000067	-10.689	0.000	-0.000846	-0.000584
capacit	0.140579	0.037881	3.711	0.000	0.066308	0.214851
getnico	0.145474	0.036470	3.989	0.000	0.073970	0.216978
gender	0.418557	0.033268	12.581	0.000	0.353330	0.483783
occu1	0.980548	0.086525	11.333	0.000	0.810904	1.150193
occu2	0.546518	0.067161	8.137	0.000	0.414839	0.678197
occu3	0.404818	0.083479	4.849	0.000	0.241144	0.568491
occu4	0.372226	0.060900	6.112	0.000	0.252822	0.491630
occu5	0.283988	0.053438	5.314	0.000	0.179215	0.388761
occu6	-0.312321	0.099837	-3.128	0.002	-0.508066	-0.116575
occu7	0.153559	0.054253	2.830	0.005	0.047188	0.259929
occu8	0.336264	0.065848	5.107	0.000	0.207160	0.465368
factor	0.483525	0.066752	7.244	0.000	0.352647	0.614402
remesas	0.556318	0.074551	7.462	0.000	0.410150	0.702487
other_nl	0.272032	0.056701	4.798	0.000	0.160860	0.383203
asalar	0.156082	0.043249	3.609	0.000	0.071286	0.240878
formal	0.367725	0.032655	11.261	0.000	0.303700	0.431751
_constant	4.569091	0.1232444	37.073	0.000	4.327452	4.81073

Note: Calculations using the individual sample weight calculated by INE in the ENEI 2002. Robust standard errors are reported.

B. Rural area

Observations:	1204	R-squared:	0.3437
F(20, 4671) :	35.26	Dependent variable:	ln(income)
Prob > F :	0.0000		

Variable	Coefficient	Standard error	t	P>t	95% Confidence interval	
educ	0.053405	0.011198	4.769	0.000	0.031434	0.075375
age	0.041950	0.008631	4.861	0.000	0.025017	0.058882
age2	-0.000433	0.000103	-4.190	0.000	-0.000636	-0.000231
capacit	*0.061184*	*0.134080*	*0.456*	*0.648*	*-0.201877*	*0.324244*
getnico	0.241239	0.060015	4.02	0.000	0.123491	0.358986
gender	0.724410	0.070907	10.216	0.000	0.585293	0.863528
occu1	1.147522	0.131439	8.730	0.000	0.889644	1.405401
occu2	0.656394	0.145398	4.514	0.000	0.371128	0.941661
occu3	*0.222572*	*0.189606*	*1.174*	*0.241*	*-0.149430*	*0.594573*
occu4	*0.358003*	*0.258748*	*1.384*	*0.167*	*-0.149653*	*0.865658*
occu5	0.412185	0.087108	4.732	0.000	0.241281	0.583088
occu6	-0.500777	0.104811	-4.778	0.000	-0.706412	-0.295141
occu7	0.280776	0.082566	3.401	0.001	0.118785	0.442768
occu8	0.460893	0.134277	3.432	0.001	0.197447	0.724340
factor	0.579084	0.239680	2.416	0.016	0.108839	1.049329
remesas	0.784560	0.141647	5.539	0.000	0.506652	1.062467
other_nl	0.381283	0.122386	3.115	0.002	0.141166	0.621400
asalar	0.200758	0.077644	2.586	0.010	0.048424	0.353092
formal	0.439515	0.083322	5.275	0.000	0.276039	0.602991
_constant	4.49511	0.1884991	23.847	0.000	4.12528	4.864939

Note: Calculations using the individual sample weight calculated by INE in the ENEI 2002. Robust standard errors are reported. Non-statistically significant variables at the 5% level appear in *italics*.

A.2 – Methodological Appendix

A.2.1 Income aggregate

The income aggregate used along this analysis is based in the income classification made in the ENEI 2002. Salaried and self-employed workers were included in the sample, while unpaid workers were excluded. Reported self-consumption figures were omitted for reasons previously stated.

Composition of the income aggregate

The income aggregate is composed by two large elements: labour income and non-labour income.

Labour income	Salaried workers	Income from the first job • Wage or salary • Bonus 14 • Christmas bonus • Vacation bonus } Monetary Bonuses Income from the second job • Wage or salary • Monetary bonuses • Earnings from working (if working as a self-employed)
	Self-employed	Income from the first job • Earnings from working Income from the second job • Wage or salary (if working for as a salaried worker) • Monetary bonuses • Earnings from working
Non-labour income	Factor payments	• Rent • Interest
	Transfers	• Donations • Scholarships • Food allowances • Transport bonus
	Remittances	• Remittances from abroad
	Other transfers	• Pensions • Indemnities
	Other income	• Inheritances

Source: ENEI 2002, Guatemalan National Institute of Statistics – 2002
Note: Self-consumption items have been removed from this income aggregate, as they were not used.

A.2.2 Variables creation and filtering

As mentioned in Section 3, given the problems implicated in the concept, measurement and capitalisation of self-consumption (non-monetary income), this source of individuals' income was excluded from the sample. For similar reasons, family and non-family workers who did not receive monetary payments were also excluded. In addition to the potential source of bias suffered from a loss in the representativity of the sample, we also acknowledge that agricultural, livestock and fishery workers, report significant levels of self-consumption[37].

In order to adapt the information found in ENEI 2002 to the requirements of the regression model presented in Section 3, the following variables were generated:

EDUC: The variables *p03a09a* and *p03a09b* were used to create this variable. Each category in *p03a09a* was given its related value in years, and then it was added up to *p03a09b* to obtain the years of education for each individual.

AGE2: This variable is simply the squared of the individual's age.

GETNICO: Using the variable *p03a04*, individuals were classified as non-indigenous if the values were 'Ladinos' or 'Extranjeros' (foreigners), and as indigenous in the rest of the cases.

OCCUPATION: Based on the variable *p05a02d1*, which tabulates the occupational categories (1st. job), eight dummy variables were generated as described in Section 3.

CAPACIT: According to the variable *p08a01*, a value of 1 was inputted if the individual had participated in a training course, and 0 otherwise.

FACTOR: It takes a value of 1 if the variable *factores* in the ENEI 2002 (which represents income received from rent and interest payments) has positive values and 0 otherwise.

REMESAS: Takes value of 1 if the variable *remesas* (remittances) in the ENEI 2002 has positive values and 0 otherwise.

OTHER_NL: Classifies the observations according to the variables *transfer*, *trans2* and *otros*, which stand for the remaining sources of non-labour income present in the income aggregate shown above. It takes value of 1 if the sum of these three variables is positive, and 0 otherwise.

NL: Classifies the observations according to the variable *no_labor*, which aggregates all monetary non-labour income. It takes value of 1 if the variable is a positive number, and 0 otherwise.

ASALAR: Using the variable *p05a08*, salaried workers were identify as those who are(1) public sector workers, (2) private sector employees, (3) day workers, and (4) domestic

[37] According to calculations derived from the sample, 70% of the agricultural, livestock and fishery workers, report food self-consumption, with a self-declared monthly average of Q.280 (≈ US$34) per capita.

workers. The remaining categories (5) self-employed (6) landlords, employers or associates, were classified as self-employed..

FORMAL: An individual was defined as a formal worker if he/she works for the government, the private sector in a firm with more than 5 employees, or a farm with more than 5 workers employed. In the rest of the cases, individuals were classified as informal workers. The variables *p05a08* and *p05a30*, were used to carry out this classification.

AREA: The area of living was differentiated between urban (value 1) and rural, using the variable *dominio*.