

# Accounting for Mexican Income Inequality during the 1990s

Rafael E. De Hoyos\*

Development Prospects Group  
*World Bank*

## Abstract

We implement several inequality decomposition methods to measure the extent to which total household income disparities can be attributable to sectoral asymmetries and differences in skill endowments. The results show that at least half of total household inequality in Mexico is attributable to incomes derived from entrepreneurial activities, an income source rarely scrutinized in the inequality literature. We show that education (skills) endowments are unevenly distributed among the Mexican population, with positive shifts in the market returns to schooling being associated with increases in inequality. Asymmetries in the allocation of education explain around 20 percent of overall household income disparities in Mexico during the 1990s. Moreover, the proportion of inequality attributable to education endowments increases during stable periods and reduces during the crisis. This pattern is mostly explained by shifts in returns to schooling rather than changes in the distribution of skills. Applying the same techniques to decompose within-sector income differences, we find that skill endowments can account for as much as 25 percent of earnings disparities but as little as 5 percent of dispersion in other income sources.

*JEL Classification:* D33, F1

*Keywords:* Inequality, regression-based decomposition, Mexico

World Bank Policy Research Working Paper 4224, May 2007

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.*

\*I am grateful to the comments and suggestions made by Paul Kattuman, Ajit Singh, Jaime Ruiz-Tagle, Hans Timmer, and Adrian Wood. Email address: [rdehoyos@worldbank.org](mailto:rdehoyos@worldbank.org).

# 1 Introduction

During the 1990s, Mexico experienced profound economic changes. The decade started with a new economic order placing the market at the center of the development strategy and reducing the role played by the state. In the mid 1990s, the North American Free Trade Agreement (NAFTA), a trilateral tariff reduction agreement between Canada, Mexico and the US, was enacted. The agreement was seen as the starting point of a long and sustained period of economic growth benefiting the less skilled laborers in particular. At the end of 1994, the same year when NAFTA was enacted, the Mexican peso suffered a massive devaluation which triggered the economic crisis of 1995. The shrinking domestic economy, combined with an expensive foreign currency within the framework of the new trade agreement turned the Mexican economy into an export-oriented one. The sectoral composition of the economy changed significantly during those years (see Figure 3). The proportion of the economy in the non-tradable service sector grew steadily between 1990 and 1995. Nevertheless, after the peso crisis and NAFTA, a redistribution equal to 2 points of the GDP occurred with the service sector shrinking and the manufacturing sector expanding.<sup>1</sup>

The economic reforms undertaken at the beginning of the 1990s in addition to the profound economic crisis of 1995 and its subsequent sectoral redistribution, could have had a significant impact upon income distribution. As it is argued by Székely (1995), the market-oriented reforms undertaken in Mexico during the second half of the 1980s had an adverse effect on household income distribution. Undertaking conventional inequality decomposition analysis Székely (1995) concludes that the inequality-enhancing impact is explained by the reduction in the government's scope for implementing redistributive policies after the privatization/liberalization of the Mexican economy during those years. Other studies that use decomposition methods in order to identify the factors that account for Mexican income inequality during the late 1980s and early 1990s had stressed the importance played by the distribution of skills in reshaping distribution (Lopez-Acevedo, 2001 and Legovini et al., 2005). Another strand of the literature had concentrated in the impact that trade liberalization has had on relative wages in the manufacturing sector. The results of this research point towards a worldwide skilled-biased technological change as the explanation behind the unfavorable distributional impact on countries that are relatively abundant in unskilled labor.<sup>2</sup>

Given the sectoral changes in the Mexican economy and the findings from previous

---

<sup>1</sup>The importance of the agricultural sector in GDP remained practically unchanged during this period.

<sup>2</sup>See Feenstra and Hanson (1997), Revenga (1997), Wood (1997) and Harrison and Hanson (1999).

studies, we argue that a sound decomposition analysis aiming to account for inequality in Mexico during the 1990s should, at least, consider the following aspects: (1) sectoral asymmetries, (2) the distribution of skills, and (3) the market rewards for those skills. By combining orthodox non-parametric inequality decomposition with a more recent semi-parametric approach, the present study measures the amount of total household income inequality that can be accounted by these three components.

We revisit the existing methods of inequality decomposition, pointing out their strengths and pitfalls. Following our methodological discussion, as a first approach, we undertake orthodox, non-parametric inequality decompositions. The advantage of this method over other decomposition techniques is that it does not impose any *a-priori* functional form on the income generating process. Moreover using the square of the coefficient of variation as our inequality index, we show that the proportion of inequality that is attributable to a particular income factor by the orthodox non-parametric decomposition rule is actually equal to the more integrated Shapley value decomposition method developed in Shorrocks (1999). The main shortcoming of the orthodox approach is the lack of economic structure behind the decomposition, which makes its interpretation somewhat difficult. Therefore to complement the non-parametric analysis, we implement a recent methodology developed by Morduch and Sicular (2002) and Fields (2003) which combines the traditional inequality decomposition by income factors with a regression-based estimation determining household income.

Our results show that earnings in the manufacturing and service sectors, together with incomes from agricultural and informal self-employment activities account for less than 50 percent of total Mexican household income inequality. These leaves more than half of total inequality without any structural interpretation, opening a mandatory line for future research. Regarding skill endowments, the regression-based decomposition shows that this asset can account for, at most, 25 percent of total income dispersion. We found that returns to schooling have a positive impact on, both, total inequality and income disparities within-sectors. Moreover, the sectoral asymmetries brought about by NAFTA and the peso devaluation, made skill endowments a more important factor contributing—positively—to income inequality in the tradable sectors. These results corroborate previous findings suggesting that the liberalizing reforms had a positive effect on relative (skilled/unskilled) wages in the tradable sector (Hanson, 2003).

This paper contributes to our understanding of the relative importance played by the most relevant factors explaining income disparities during times of liberalizing reforms.<sup>3</sup>

---

<sup>3</sup>As we stated above, although the present study includes the factors that, according to the literature are the most important determinants of inequality, we can only account for 50 percent of total

For instance, a considerably large literature had concentrated its analysis in the impact of trade reforms on relative wages in the manufacturing sector, using the methodology presented in this paper, we can quantify the proportion of total household income inequality that can be explained by relative wages in the manufacturing sector. Uncovering these issues can inform policy makers on the extent to which household income inequality can be deemed to be a problem associated with the distribution of skills or an outcome of sectoral asymmetries. Furthermore, our results can prompt scholars to move towards a new research agenda where income factors different from relative wages are used to explore the distributional impact of liberalizing or other types of economic reforms.

The paper is organized as follows. To give the reader an idea of the macroeconomic context prevailing during the 1990s, in Section 2 we present a brief description of the major macroeconomic changes which took place in Mexico during that period. In the same section, we show inequality trends and major changes in tradable and non-tradable labor markets. A revision and discussion of existing decomposition methodologies is presented in Section 3. The inequality decomposition results are shown in Section 4. Finally the conclusions can be found in the last section.

## 2 Macroeconomic and Inequality Performance

### 2.1 Macroeconomic Changes

Mexico had suffered from instability and a lack of sustained growth starting from the debt crisis in 1982 and continuing to the peso crisis of 1994-95. The macroeconomic performance during those years had ups and downs. Figure 1 shows the rate of growth of GDP and yearly inflation from 1981 to 2000. During the 1980s there were 2 recessions, the first one deriving from the debt crisis of 1982 and the second one starting with the devaluation of the Mexican peso in 1985. From 1989 to 1994 the macroeconomic performance was stable. Inflation reached a peak of 153 percent in 1988, but by 1994 it was within the one-digit figures. In December 1994 the Mexican peso devaluated by more than 60 percent against the US dollar. This was the beginning of the 1995 crisis where GDP decreased by almost 8 per cent and inflation jumped to 42 percent. The last period under analysis (1996-2000) was characterized by a rapid recovery with GDP growing at an average yearly rate of more than 5 percent and annual price changes reducing to 7 percent.

---

household income variation.

Figure 1: Macroeconomic Overview

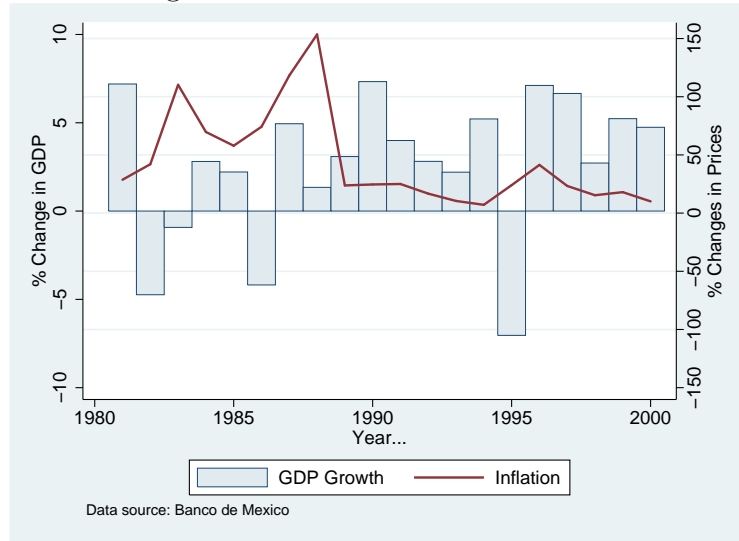
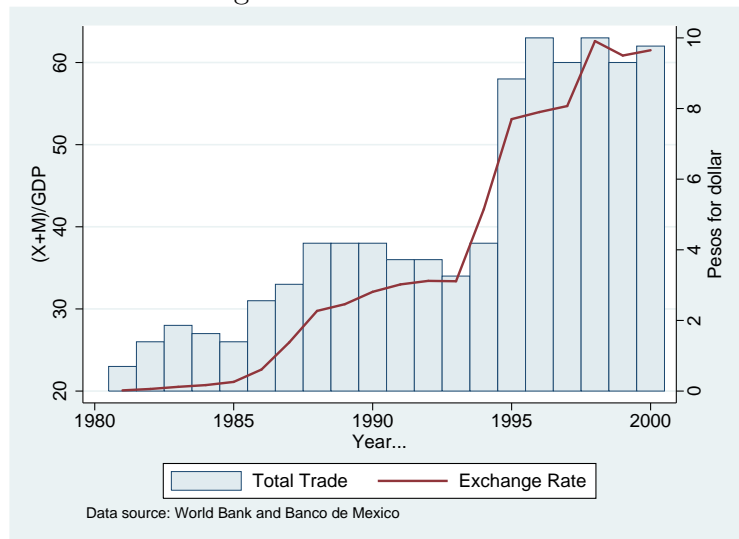


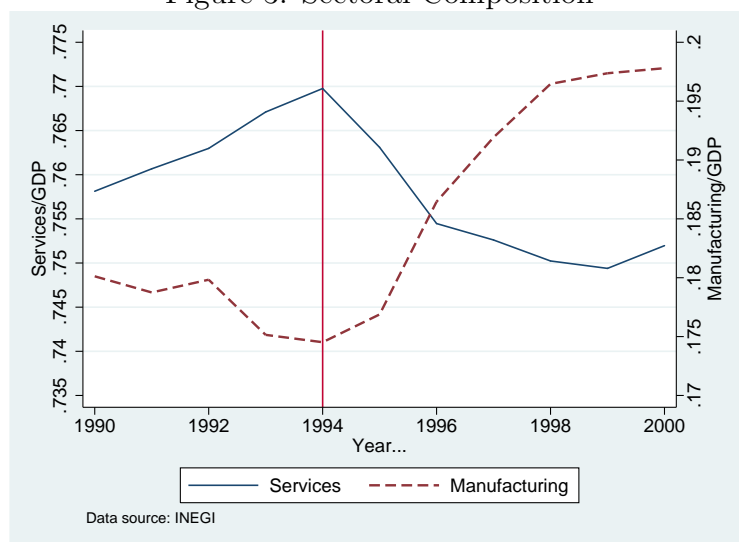
Figure 2: External Sector



During the same period major market-oriented reforms took place. In 1985 Mexico joined the General Agreement on Tariff and Trade (GATT). Between 1985 and 1988, the average tariff dropped from 25 percent to 11 percent and the coverage of import licensing decreased from 92.2 percent to 23.3 percent of total tradable goods. The reduction on trade barriers made Mexican products more competitive in international markets. In Figure 2 we can see that total trade (measured as the sum of exports plus imports as the proportion of GDP) increased from 26 percent in 1985 to 36 percent in 1988. The second major trade reform came in 1994 when the North American Trade Agreement (NAFTA) was enacted. As we mentioned above, in December of that year

the Mexican peso suffered a major devaluation making the exporting sector (mostly the manufacturing one) the most dynamic sector in the economy.<sup>4</sup> Between 1994 and 1996 the importance of international trade in the Mexican economy almost doubled, passing from a pre-crisis/NAFTA level of 38 percent to 63 percent in 1996.

Figure 3: Sectoral Composition



The peso devaluation together with the trade opportunities brought about by NAFTA, had a significant sectoral redistribution impact on the Mexican economy. The proportion of GDP that was generated in the tradable manufacturing sector experienced an average annual expansion of 3 percent between 1994 and 1998. By the year 2000 the manufacturing sector accounted for 20 percent of the economy compared with a ratio of 17.5 percent in 1994. The counterpart of this increase was a reduction of the same scale in the proportion of GDP that was generated in the service sector, passing from 77 per cent in 1994 to 75 percent in 2000 (see Figure 3).

## 2.2 Inequality Levels

The huge macroeconomic turbulence affected microeconomic agents (i.e. households, firms and individuals) especially through changes in overall income distribution. In this section, we will describe the distributional changes that took place during the 1990s and propose possible explanations for them. In the following section, we will quantify how much inequality is accounted for by these possible explanations.

<sup>4</sup>Between 1994 and 2000 manufacturing exports accounted for 95 percent of total exports.

To compute and decompose total income inequality, we use micro data from household income surveys (ENIGH) conducted by INEGI, the National Institute of Statistics in Mexico. ENIGH surveyed more than 11,000 households during years 1989, 1992, 1994, 1996, 1998 and 2000. The survey allows us to identify every individual in the household and their position within it (head, spouse, etc). We can also learn the source of income of every member of the household, their occupation, the industry they work in, hours worked, age, gender and many other variables of interest.

While constructing all statistics and inequality indexes ENIGH's survey design was taken into account. Therefore all the figures presented in this paper account for ENIGH's stratification, clustering and expansion factors. Following the literature, our inequality measures used the household as the unit of analysis (Cowell 2000). Given the difficulty of identifying intra-household distribution, our preferred income measure is household per capita incomes hence we assume no intra-household economies of scale and constant *costs* across adults and children within the household.<sup>5</sup>

Table 1: Income Inequality Indexes

	1989	1992	1994	1996	1998	2000
Gini	0.518	0.537	0.534	0.515	0.527	0.528
Theil	0.593	0.598	0.568	0.531	0.559	0.548
Entropy <sub>(<math>\epsilon=-1</math>)</sub>	0.724	0.769	0.751	0.694	0.796	0.782
Entropy <sub>(<math>\epsilon=2</math>)</sub>	2.855	1.655	1.211	1.284	1.369	1.121

Data source: Own calculations with data from ENIGH

Table 1 shows the value of four popular income inequality indexes: the Gini coefficient, the Theil and the General Entropy with inequality aversion parameter equal to -1 and 2, respectively. Table 1 shows that, although there was a distributional improvement between 1992 and 1994, we can't conclude anything about inequality changes occurring between 1989 and 1994. Our inference about the distributional changes observed between 1989 and 1994 depend on the weight given to the different parts of the income distribution, in other words there is no stochastic- or Lorenz-dominance.<sup>6</sup> Surprisingly, three out of our four inequality measures imply that the severe peso crisis of 1994-95 had a favorable distributive effect.<sup>7</sup> Between 1996 and 1998, there was a consider-

<sup>5</sup>For a more detailed description of ENIGH's survey design and the methodology followed to construct inequality indexes see De Hoyos (2005a).

<sup>6</sup>Distribution  $D_1$  Lorenz-dominates  $D_2$  if and only if *all* the points in the Lorenz curve corresponding to  $D_1$  lie closer to the 45° line than the points corresponding to  $D_2$ .

<sup>7</sup>Lopez-Acevedo and Salinas (1999) documented the possible causes behind the reduction in inequality during the 1995 economic crisis.

able increase in income dispersion (regardless of the inequality measure used) and it remained like that up to the year 2000. To summarize, distribution was more or less stable between 1989 and 1994, showing some marginal deterioration; the 1994-95 crises had a favorable distributive effect which was eliminated during the recovery phase 1996-2000.

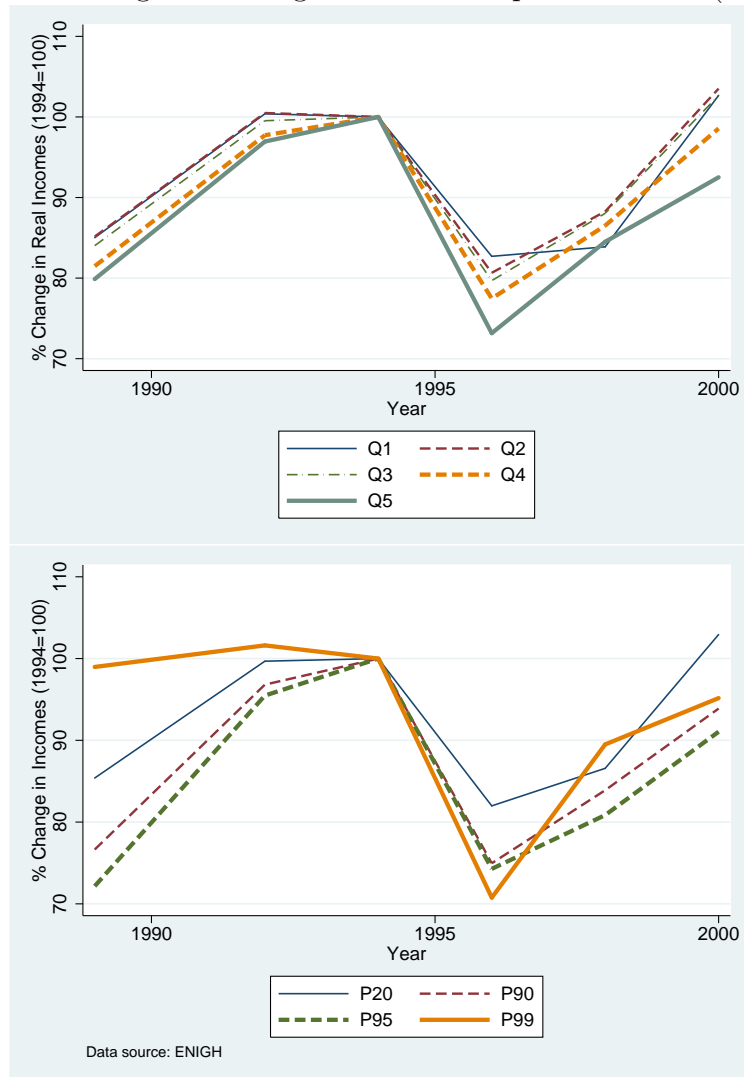
Empirical evidence shows that inequality indexes tend to be rather stable over time. This does not necessarily mean that the underlying elements behind the inequality measures are stable. To have a closer look at the changes in distribution, in the top part of Figure 4 we show the percentage change in real average per capita household income for the different income quintiles. Between 1989 and 1994, the poorest income cohorts (Q1, Q2 and Q3) did not experience an increase in real incomes as fast as the one taking place in Q4 and Q5. However, during the crisis period (1994-96), the richest 20 percent of the population (Q5) were affected more than proportionally by the negative shock with Q1 being the least affected one. This is the reason why all inequality measures show a distributional improvement between 1994 and 1996. On the other hand, between the years 1996 and 1998 the real incomes of the poorest 20 percent (Q1) remained practically constant, unlike the rising real incomes experienced in all other income cohorts. This explains the documented deterioration in distribution between 1996 and 1998. Despite the fact that by the year 2000, after a period of strong growth, Q1 to Q3 had a real income level *higher* than that observed in 1994, real income of the 'richest' 40 percent of the population remained below the 1994 level. Given the large proportion of the Mexican population living under poverty, the upper 40 percent of the income distribution actually covers the middle class.<sup>8</sup> Trying to capture heterogenous changes within the upper income cohorts, in the bottom part of Figure 4 we show income changes for several upperpercentiles. We can see that the income changes of the richest 10 and 5 percent are quite different from what happened at the top 1 percent of the distribution. Notice that the heterogeneity in performance among the richest 5 percent of the population is significantly higher than the dissimilarities among all the population. Between 1989 and 1994, real income of the richest 1 percent of the population hardly changes, however, after the 1994-96 crisis, incomes at the very top of the distribution showed a much more volatile behaviour.

---

<sup>8</sup>Taking the poverty lines provided by the Mexican Ministry for Social Development, in the year 2000, 20 percent of the population could not satisfy their nutritional needs; 26 percent did not earn enough to cover educational and health costs; and 49 percent could not cover transport or housing costs (see De Hoyos 2005a).

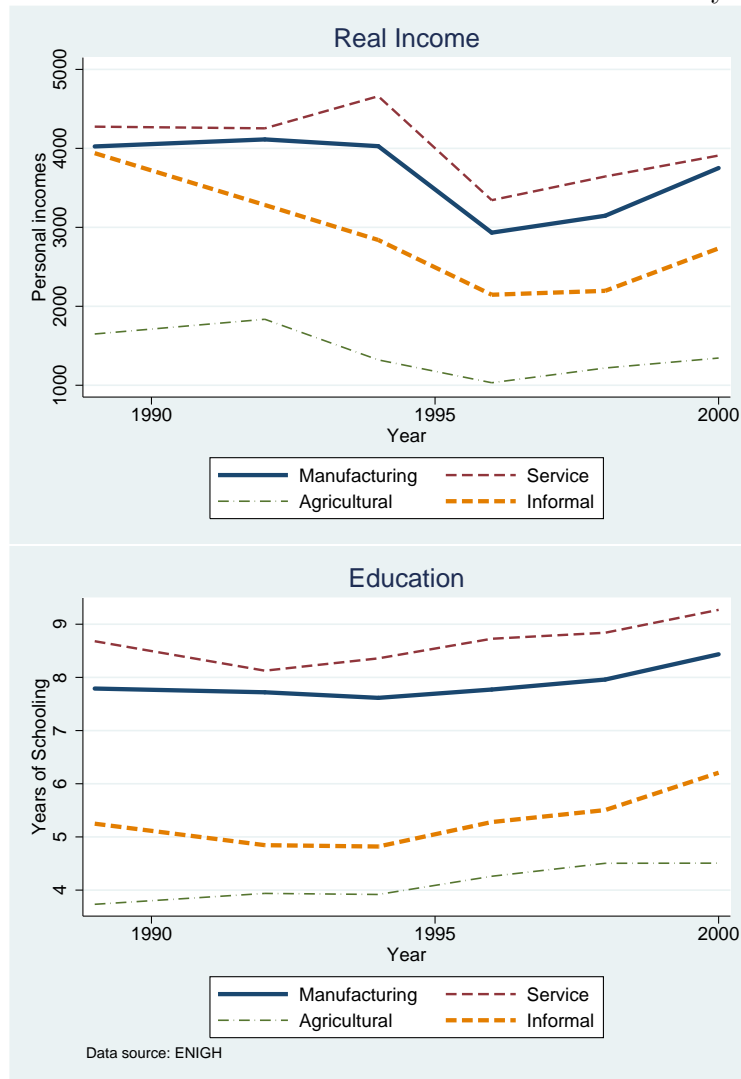


Figure 4: % Change in Average Real Income per Percentile (1994=100)



The persistent high income inequality and the changes registered after the peso crisis might be explained by sectoral redistributions caused by the 1994-95 macro shocks (see Figure 3). For instance, real incomes of heads of household working in the agricultural sector could have been stagnant during the general recovery period of the economy (1996-2000) and this might explain the increasing gap between the household incomes of the poorest and richest cohorts during those years. To explore this possibility, let us define 4 sectors in the economy: the urban manufacturing sector (tradable sector), urban service sector (non-tradable sector), agricultural sector (rural tradable) and the informal sector. In Figure 5 we show the average real personal income and the average years of formal education of heads of household in each sector of the economy.

Figure 5: Real Personal Incomes and Formal Education by Sectors



A couple of important points arise from Figure 5. Firstly, average income in the informal and agricultural sectors (where poverty concentrates) were decreasing even before the peso crisis. As a matter of fact, during the peso crisis (1994-96) real incomes of heads of households working in the agricultural sector, the lowest remunerated sector, decreased proportionally less than the large income reduction experienced by heads of household working in the relatively better off manufacturing and service sectors. This helps explain the improvements in distribution after the 1994-96 peso crisis. Secondly, after the combination of the peso crisis and NAFTA (1994), in the middle of a recession that caused an increase in the rate of unemployment, the years of education of the representative laborer increased in all sectors, suggesting that the ones who kept their jobs were the relatively better educated workers.

The descriptive statistics showed the important role played by disparities in sectoral remunerations and education endowments as possible explanations of the persistently high levels of household income inequality observed during the 1990s in Mexico. The relatively stable incomes in the agricultural and informal sectors even in the presence of a negative shock combined with a more-than-proportional reduction of incomes in the highest income cohorts explains the reduction in inequality after the 1994-95 peso crisis. On the other hand, the distribution of education endowments might be the source behind the post-crisis/NAFTA deterioration in distribution. In order to explore these hypotheses further, in the next section we will implement conventional non-parametric decomposition methods to quantify the importance of sectoral disparities in accounting for overall household income dispersion. Furthermore, we will combine orthodox non-parametric decomposition by factor components with regression analysis to quantify the importance of skills distribution and their market contribution in explaining the high levels of household income inequality observed in Mexico during the 1990s.

### 3 Methodological Aspects

Let us define  $\mathbf{Y}$  as a vector containing  $(1, \dots, N)$  household incomes as elements. In turn, income of household  $i$  is defined as the sum of  $K$  different household income components  $Y_i = \sum_{k=1}^K Y_{ik}$ . Assuming that the income components are mutually exclusive, an income inequality index,  $I(\mathbf{Y})$ , measuring household income dispersion can be defined as the sum of the contribution,  $S_k(\mathbf{Y}_k)$ , made by the  $K$  different income components:

$$I(\mathbf{Y}) = \sum_{k=1}^K S_k(\mathbf{Y}_k) \quad \mathbf{Y}_k = (Y_{1k}, \dots, Y_{Nk}) \quad (1)$$

This type of decomposition can answer the question: *what proportion of total income inequality,  $I(\mathbf{Y})$ , is explained by income factor  $\mathbf{Y}_k$ ?* Several decomposition methods had been developed, ranging from pure non-parametric ones (Shorrocks, 1982) to the more sophisticated ones based on microeconomic modeling with endogenous behavior (Bourguignon, Fournier and Gurgand, 2001). Shorrocks' (1982) seminal paper shows that, given a set of desired decomposition properties and under several assumptions, there is a unique factor decomposition rule. This decomposition rule is independent of the inequality index used and defines the *proportion* of total inequality that is attributable to income factor  $k$  in the following way:

$$s_k = \frac{\text{cov}(\mathbf{Y}, \mathbf{Y}_k)}{\sigma^2(\mathbf{Y})} \quad (2)$$

Where  $\text{cov}(\mathbf{Y}, \mathbf{Y}_k)$  is the covariance between total income and income from source  $k$  and  $\sigma^2(\mathbf{Y})$  is the variance of total income. The main advantage of this non-parametric technique lies in the absence of assumptions about structural relationships, i.e. no formal model or econometric estimation is involved. This advantage is, however, the source of its weakness. In the absence of economic structure very little can be said about the economic mechanisms driving the results. Two recent studies try to overcome this problem while keeping Shorrocks' (1982) decomposition principle. Fields (2003) develops a semi-parametric method combining Shorrocks (1982) technique with regression analysis. The author shows that income sources,  $\mathbf{Y}_k$ , can be analogous to the market value of personal characteristics within a human capital regression framework. For instance, define the logarithm of income as a function of a matrix of observable characteristics  $\mathbf{X}$ , a vector of regression parameters,  $\boldsymbol{\beta}$  and a set of unobservable components  $\boldsymbol{\epsilon}$ :

$$\ln(\mathbf{Y}) = g(\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\epsilon}) \quad (3)$$

Say that we are interested in the proportion of total income inequality that is attributable to characteristic  $k$ ,  $s_k$ , then after the parameters in equation 3 have been estimated,  $s_k$  is defined as (Fields 2003):<sup>9</sup>

$$s_k = \frac{\hat{\beta}_k \sigma(\mathbf{X}_k) \text{cor}(\ln(\mathbf{Y}), \mathbf{X}_k)}{\sigma[\ln(\mathbf{Y})]} \quad (4)$$

Where  $\hat{\beta}_k$  is the market return to characteristic  $k$  estimated from model (3) and  $\text{cor}(\ln(\mathbf{Y}), \mathbf{X}_k)$  is the correlation between the log of income and element  $k$ . Based on a formula used to decompose inequality into different income *factors*, equation 4 quantifies the proportion of inequality explained by *characteristic*  $\mathbf{X}_k$ . The advantage of a specification like (4) is that we can express proportions  $s_k$  as a function of both the

---

<sup>9</sup>Assume the functional form of equation 3 is given by  $\ln(Y) = \alpha + \beta_k X_k + \epsilon$ , define  $Y_k = \hat{\beta}_k X_k$  and substitute it into equation 2:

$$\begin{aligned} s_k &= \frac{\text{cov}[\ln(Y), (\hat{\beta}_k X_k)]}{\sigma^2[\ln(Y)]} = \frac{\hat{\beta}_k \text{cov}(\ln(Y), X_k)}{\sigma^2[\ln(Y)]} \\ &= \frac{\hat{\beta}_k \text{cor}(\ln(Y), X_k) \sigma[\ln(Y)] \sigma(X_k)}{\sigma^2[\ln(Y)]} = \frac{\hat{\beta}_k \sigma(X_k) \text{cor}(\ln(Y), X_k)}{\sigma[\ln(Y)]} \end{aligned}$$

distribution of characteristics  $\mathbf{X}$  and their market reward  $\hat{\beta}$ . Fields (2003) shows that this result applies for any inequality index, however it is only valid for decomposing inequality of the log of incomes, which makes it a rather unattractive method.<sup>10</sup> In a closely related study, Morduch and Sicular (2002) integrate inequality decomposition by factor components and population subgroups using a semi-parametric methodology. The logic behind this method is similar to the one described by equation 4, with two major modifications. First, the authors decompose the inequality of household incomes—as opposed to the log of incomes as was done by Fields (2002). Second, with the use of an axiomatic result defining the property of uniform additions, the authors show that the regression-based decomposition formula will vary both with the inequality index used and the factor decomposition rule used.<sup>11</sup> In the particular case where the inequality measure that is used is the squared coefficient of variation ( $\text{Entropy}_{(\varepsilon=2)}$  in Table 1) the *natural* decomposition rule is equation 2, and therefore the regression-based decomposition expression is equal to equation 4 but using the incomes in levels instead of its logarithm (Morduch and Sicular 2002):

$$s_k = \frac{\hat{\beta}_k \sigma(\mathbf{X}_k) \text{cor}(\mathbf{Y}, \mathbf{X}_k)}{\sigma(\mathbf{Y})} \quad (5)$$

Morduch and Sicular (2002) show that the coefficient of variation’s decomposition rule does not satisfy the property of uniform additions.<sup>12</sup>

The numerous decomposition rules found in the literature denote the lack of a unified framework. This lack of consensus had led to several pitfalls in existing techniques. For instance, the different decomposition rules will impose constraints on the inequality

---

<sup>10</sup>The variance of the log of incomes is an inequality index that violates the transfer principle (Jenkins, 1991).

<sup>11</sup>Notice that this second aspect was first pointed out in Shorrocks (1982 and 1983). Shorrocks (1982) shows that given the large range of possible decomposition rules, the contribution assigned to any factor can be made to take any value from minus to plus infinity.

<sup>12</sup>Departing from the transfer and the scale invariant axioms, Morduch and Sicular (2002) define the property of uniform additions as the reduction in an inequality index after a positive transfer of equal size to each member of the population had occurred. The authors claim that a decomposition rule will satisfy this property if the proportion of inequality attributable to factor  $k$ ,  $s_k$ , is negative when  $\mathbf{Y}_k$  is equally distributed. This *property* is not satisfied by Shorrocks’ (1982) decomposition rule (equation 2) because in such a case  $\text{cov}(\mathbf{Y}, \mathbf{Y}_k) = 0$ . However the desirability of such a condition is, at least, debatable. The condition to have  $s_k < 0$  under equation 2 is  $\text{cov}(\mathbf{Y}, \mathbf{Y}_k) < 0$ . This will be the case of, for instance, well-targeted government anti-poverty transfers, which should and will have a negative effect upon overall inequality under equation 2. The discussion comes down to the welfare economics’ debate of evaluating the contribution of an income factor by eliminating it versus making it an equally distributed component.

indexes that can be used. Another shortcoming of existing techniques is the unsatisfactory way in which they deal with decompositions of factor components *together* with population subgroups (Shorrocks, 1999). Driven by these concerns, Shorrocks (1999) developed a unified framework for distributional analysis called the Shapley decomposition. To explain the logic behind the Shapley decomposition, let us define  $f(\cdot)$  as a general function determining income distribution:

$$I(\mathbf{Y}) = f(\mathbf{X}_1, \dots, \mathbf{X}_m, \dots, \mathbf{X}_M) \quad (6)$$

Where  $\{\mathbf{X}, \dots, \mathbf{X}_M\}$  is the set of explanatory variables determining income distribution. Notice that  $f(\cdot)$  is a generally defined function that is able to incorporate an income-generating model like equation 3 and even more sophisticated ones like the one developed in Bourguignon, Fournier and Gurgand (2001). The amount of total income inequality that is accounted for by  $\mathbf{X}_m$  can be estimated via a counterfactual analysis. The counterfactual capturing the contribution of  $\mathbf{X}_m$  in total income inequality is the following: *what would income distribution look like had factor  $m$  been eliminated?* Define income inequality under this counterfactual as  $I(\mathbf{Y}|\mathbf{X}_m \in \emptyset)$ . The contribution made by factor  $m$  to total income inequality is defined as (Shorrocks 1999):

$$C_k = I(\mathbf{Y}) - I(\mathbf{Y}|\mathbf{X}_m \in \emptyset) \quad (7)$$

The problem with this type of decomposition is that normally  $\sum_{m=1}^M C_m \neq I(\mathbf{Y})$ , i.e. the sum of the contributions won't necessarily give an exact decomposition. A second possibility is to eliminate the  $M$  factors in sequences however the contribution of factor  $m$ ,  $C_m$ , depends on the remaining factors in the counterfactual therefore there is a 'path dependency' problem. In other words,  $C_m$  might be different if element  $m'$  had been eliminated or not. Shorrocks (1999) deals with this 'path dependency' problem by eliminating  $\mathbf{X}_m$  in all possible sequences, computing the contribution made by factor  $m$  in each round and finally averaging over all rounds. Define the set of all possible scenarios given by the eliminating sequences as  $L = \{1, \dots, M!\}$ . All possible contributions of factor  $m$  are then defined as  $\{C_m^1, \dots, C_m^{M!}\}$ . The Shapley value contribution is computed as:

$$C_m^* = \frac{1}{M!} \sum_{m=1}^{M!} C_m \quad (8)$$

Equation 8 can be interpreted as the expected contribution of factor  $m$  in total inequality over all possible elimination paths (Shorrocks 1999). The decomposition rule

specified by equation 8 is exact (the sum of the contributions is equal to total inequality) and treats factors symmetrically.<sup>13</sup>

The Shapley decomposition can handle satisfactorily decompositions by income factors and population subgroups given that the underlying model,  $f(\cdot)$ , had been appropriately specified. In the context of income factor decompositions, equation 8 can be further simplified. Taking the square of the coefficient of variation ( $CV^2$ ) as our preferred inequality index and defining the underlying model,  $f(\cdot)$ , simply as the sum of inequality shares attributed to the different income sources (similar to the model defined by equation 1), then Shorrocks (1999) shows that the Shapley contribution of income factor  $k$  is equal to:

$$C_k^* = \frac{cov(\mathbf{Y}, \mathbf{X}_k)}{\mu^2} \quad (9)$$

And the *proportion* of total inequality explained by factor  $k$  is equal to:

$$s_k = \frac{C_k^*}{CV^2} = \frac{cov(\mathbf{Y}, \mathbf{Y}_k)}{\sigma^2(\mathbf{Y})} \quad (10)$$

Equation 10 is exactly the unique decomposition rule proposed in Shorrocks (1982). This is quite a powerful result because it allows us to interpret the unique decomposition results (equation 2) as the expected value of the factor's contribution over all possible scenarios, given that  $CV^2$  is chosen as the income inequality index and equation 1 is the underlying income distribution model.

## 4 Decomposition Results

### 4.1 Non-Parametric Approach

Given the advantages and disadvantages of the different decomposition methods, in this paper we will undertake conventional non-parametric decompositions and complement them with the semi-parametric approach just described.<sup>14</sup>

---

<sup>13</sup>A point that it is not clear from the Shapley decomposition rule is the way in which interactions ought to be treated. The basic idea is to compute the average marginal impact of factor  $k$  over all possible eliminating sequences. However, the elimination of, say,  $X_m$  will capture part of the contribution made by, say,  $X_{m'}$  if model 6 includes interaction term  $X_m X_{m'}$ .

<sup>14</sup>The method developed by Bourguignon, Fournier and Gurgand (2001) goes beyond the scope of this paper. For two recent applications of this kind using Mexican data see: De Hoyos (2005b) and

To capture the distributional effects of the sectoral asymmetries described in section 2, let us define household income as the sum of incomes derived from manufacturing earnings ( $\mathbf{Y}_m$ ), other earnings ( $\mathbf{Y}_s$ ), incomes from informal activities ( $\mathbf{Y}_i$ ), agricultural incomes ( $\mathbf{Y}_a$ ) and other incomes ( $\mathbf{Y}_o$ ).<sup>15</sup> An alternative approach would have been to classify heads of households in the different sectors as forming different population subgroups and then undertaking between and within subgroups inequality type of analysis. Although useful, this approach cannot be combined with the semi-parametric techniques described in the last section. Therefore in this subsection we treat incomes derived from the different sectors as different household income sources.<sup>16</sup>

In Table 2 we show the proportion of inequality explained by each element of household per capita incomes using Shorrocks' (1982) decomposition rule (equation 2). Or else, if we are willing to take  $CV^2$  as our preferred inequality measure, we can interpret the results in Table 2 as the Shapley contribution of each factor to total income inequality within a more general and integrated framework.

Throughout the period, the element explaining the highest proportion of inequality is  $\mathbf{Y}_o$ , accounting for more than 50 per cent of total household income inequality. This is not surprising given the way in which we define household income components.  $\mathbf{Y}_o$  is basically composed of incomes derived from entrepreneurial activities and the returns to financial assets, where inequality is concentrated. These two elements are highly correlated with household incomes, specially at the upper tail of the income distribution. On the other hand, incomes from agricultural and informal activities,  $\mathbf{Y}_a$  and  $\mathbf{Y}_i$ , jointly account for at most 4 percent of total income inequality. This is due to the poor correlation between these income components and total household income. Between 1989 and 1994, within a growing economy context, the proportion of inequality attributable to  $\mathbf{Y}_o$  decreased and this came together with an increase in the earning components,  $\mathbf{Y}_m$  and  $\mathbf{Y}_s$ . These changes are the result of the macroeconomic stability observed during those years, with earnings increasing its covariance with total household income. This trend was reversed during the post-crisis year of 1996 but then resumed after 1998. It is important to notice that incomes from agricultural activities made a negative contribution to total household income inequality during years 1994 and 1996. This shows how households dependent upon incomes from agricultural

---

Legovini, A., Bouillón, C. and Lustig, N. (2005)

<sup>15</sup>We classify workers in the informal sector when they are non-professional self-employed workers. We exclude family workers that get no monetary remuneration (See Maloney 1999).

<sup>16</sup>This approach is based on a definition of total household income as the sum of earnings, income from self employment activities and an exogenous income. We further divide earnings and self-employed incomes into earnings in the tradeable sector (manufacturing) and non-tradeable sector (services) and agricultural and informal incomes, respectively.



activities were falling down in overall income distribution before the peso crisis.

Table 2: Proportion of Inequality Explained by Factor Components

	1989	1992	1994	1996	1998	2000
$s_m$	2.43	4.77	7.35	7.38	4.66	8.58
$s_s$	10.57	16.84	36.84	24.44	25.02	31.76
$s_i$	3.15	0.92	3.89	2.74	1.38	3.81
$s_a$	0.33	0.14	-0.13	-0.04	1.21	0.17
$s_o$	83.49	77.31	52.04	65.49	67.72	55.65

Source: Own calculations with data from ENIGH

The results in Table 2 can be used to address our main question, i.e. *what proportion of total household income inequality can be accounted for by incomes derived in the different sectors?* Incomes derived from service, manufacturing, agricultural and informal sectors can account for almost half of total household income inequality.<sup>17</sup> The proportion of inequality that is attributable to incomes generated in these sectors, increases in periods of stabilization. Despite the significant sectoral redistribution documented in section 2, the proportion of inequality that is explained by earnings in the manufacturing and service sectors, reduced between 1994 and 2000.

The results from Table 2 made apparent how little we actually know about the causes behind the high levels of Mexican income inequality. Even using complex enough models able to fully parameterize and account for total variation within income components  $\mathbf{Y}_m$ ,  $\mathbf{Y}_s$ ,  $\mathbf{Y}_i$  and  $\mathbf{Y}_a$ , we still wouldn't be able to account for more than half of total household income inequality in Mexico during the 1990s.<sup>18</sup> Our results show that much of the income variation is actually concentrated in the highest income cohorts, where very little is known about their income generating processes. To the best of our knowledge, there are no models trying to estimate structural relationships capturing the income generating process occurring at the very top of the distribution. This leaves more than 50 percent of total income disparities without any structural interpretation, opening a mandatory line for future research on the causes behind Mexican high levels of inequality.

<sup>17</sup>Taking  $CV^2$  as the inequality index, we can assert that this proportion of total inequality explained by sectoral differences is indeed the expected value of their contribution (the Shapley value).

<sup>18</sup>For example, De Hoyos (2005b) estimates a model that parameterizes income components  $\mathbf{Y}_m$ ,  $\mathbf{Y}_s$ ,  $\mathbf{Y}_i$  incorporating behavior and labor market restrictions. This model would be able to account for less than half of total income inequality between 1994 and 2000.

## 4.2 Semi-parametric approach

The trade versus relative wages literature and some of the broader decomposition studies had emphasized the role played by the distribution of skills (education) in reshaping income distribution. To quantify how much inequality can be accounted for by the distribution of education (and other personal and household characteristics) we will use the regression-based decomposition approach explained in Section 3. Taking  $CV^2$  as our inequality measure and hence using equation 5 as our regression-based decomposition method, let us define the following human capital regression model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (11)$$

Where  $\mathbf{X}$  is a  $(N \times K)$  matrix containing  $(K - 1)$  personal and household characteristics plus a constant for  $N$  heads of household;  $\boldsymbol{\beta}$  is a  $(K \times 1)$  vector with the ‘prices’ of those characteristics, and  $\boldsymbol{\varepsilon}$  is a vector of random components assumed to be normally distributed with zero mean.<sup>19</sup> The elements forming  $\mathbf{X}$  include three personal characteristics: years of schooling, years of schooling interacting with a dummy variable for higher education, experience, experience squared and gender; and three household characteristics: household size and the ratio of dependent to total household members and a regional dummy variable for households located in the north of Mexico.<sup>20</sup> Equation 11 is a rather rigid specification imposing a constant ‘price’ of characteristics  $\mathbf{X}$  and the same functional form across all sectors of the economy. It also assumes no labor supply effects on  $\mathbf{Y}$ . Bearing these constraints in mind, the regression results are shown in the top part of Table 3. All the variables included in our regression are highly significant and show the expected sign with the unexpected exception being the dummy for the gender of the head of household, which turned out to be not-significant. Heads of household with more education, more than high school level and/or more experience tend to have higher incomes. At some points in time, particularly during the post-1994 recovery years, households located in the north of Mexico had, on average, higher per capita household incomes. On the other hand, larger households and especially those with a high dependency ratio have significantly less per capita incomes compared with smaller households.

---

<sup>19</sup>Notice that since we are relating total household income inequality with personal characteristics such as education, we are constraint to use *only* the heads of the household. Nevertheless, the original sampling weights are multiply by the household size to preserve a representation of the entire population.

<sup>20</sup>In this context, each element  $X_k\beta_k$  and  $\boldsymbol{\varepsilon}$  can be interpreted as an income factor  $Y_k$  within Shorrocks’ (1982) framework.

Table 3: Regression-Based Decomposition Results

	1989	1992	1994	1996	1998	2000
Schooling	158.0**	263.0**	305.0**	168.0**	187.0**	237.0**
Schooling*(H)	100.0**	153.0**	149.0**	114.0**	130.0**	154.0**
Experience	73.0**	174.0**	131.0**	49.0**	78.0**	128.0**
Experience Sq.	-0.7**	-2.0**	-1.2**	-0.3	-0.7*	-1.2*
Gender	1.6	462.5	116.9	206.4	98.3	-86.1
HH size	-157.0**	-255.0**	-245.0**	-214.0**	-247.0**	-324.0**
Dep. Ratio	-2289.0**	-2922.0**	-3179.0**	-1893.0**	-2096.0**	-2548.0**
Region	256.0*	77.0	185.0	170.0	301.0**	598.0**
Intercept	1630.0**	103.0	850.0*	1396.0**	1159.0**	830.0
R-squared	0.102	0.231	0.306	0.280	0.295	0.291
N	10,005	9,181	11,123	12,220	9,430	8,593
<b>% Contribution to total inequality</b>						
Schooling	7.2	14.8	24.5	21.1	18.2	21.7
Experience	-1.2	-2	-2.9	-2.2	-2	-1.9
Gender	0	0.2	0	0.1	0	0
HH charact.	2.9	7.2	7.8	8.3	6.4	7.4
Region	0.1	0	0	0.1	0.2	0.5
Residual	91	79.8	70.7	72.6	77	72.3

Dependent variable: per capita household incomes

\*,\*\* significant at the 5% and 1% level respectively

In the bottom part of Table 3 we show the percentage contribution to total income inequality (using equation 5 as the decomposition rule) of each element included in the income equation 11. A story similar to the one told by the Shorrocks (1982) orthodox decomposition, or the more general Shapley procedure with inequality index  $CV^2$ , is shown by the regression-based method. The amount of inequality explained by our regression model increases during the stable period 1989-1994, mainly driven by the proportion of inequality accounted for by education and household characteristics, and then it marginally reduces during the crisis years.<sup>21</sup> The parallelism between the non-parametric and semi-parametric results suggest that at least part of the covariance

<sup>21</sup>The contribution of elements that enter more than one time in the regression equation is simply the sum of their individual contribution as computed by equation 5.

between elements  $\mathbf{Y}_m$ ,  $\mathbf{Y}_s$ ,  $\mathbf{Y}_i$ ,  $\mathbf{Y}_a$  and per capita household income  $\mathbf{Y}$  is given via  $\mathbf{X}\boldsymbol{\beta}$ . The distribution of education and household characteristics together with their respective ‘prices’ account for as much as 25 percent and 8 percent of total household income inequality, respectively. Experience, on the other hand, has a negative, though small contribution to inequality. Therefore higher rewards to experience help ameliorating income disparities. This result is driven by the fact that experience, measured as age minus years of schooling, is an endowment that is evenly distributed among the population. As we would have expected from the regression results, given their lack of significance, the gender and regional dummies do not help to explain much of the differences in incomes. Indeed, the significance of the elements included in  $\mathbf{X}$  is related to their regression-based decomposition rule (equation 5) in a way such that total inequality explained by elements in equation 11—excluding the residual—is equal to the proportion of variance of  $\mathbf{Y}$  explained by the variance of  $\mathbf{X}$ , i.e.  $R^2$  in Table 3 (Fields, 2003).

From equation 5 we know that the contributions shown in the bottom part of Table 3 are the outcome of the distribution of characteristics  $\mathbf{X}$  (more precisely, their standard deviation and correlation with total household income) and their respective market rewards,  $\hat{\boldsymbol{\beta}}$ . Therefore the documented pattern followed by the contribution to inequality made by schooling and household characteristics between 1989 and 2000, could be the outcome of shifts in market returns to those characteristics. From the upper part of Table 3 we can corroborate that, regarding education endowments, this is indeed the case. Market returns to schooling (and the premium for higher education) showed a positive trend between 1989 and 1994, decreased during the crisis and then recovered between 1996 and 2000. This was exactly the same pattern followed by the proportion of inequality accounted for by education endowments. Therefore much of the changes in the proportion of inequality that is attributable to education is, indeed, explained by changes in returns to schooling. These results show that increases in returns to schooling are inequality-increasing, implying that skill (education) endowments are a unequally distributed asset.

So far we have shown that differences in education, household characteristics and their respective market ‘prices’ can account for at most 25 percent of total income inequality during the 1990s. This is a significant proportion of total income disparities, however the importance of these characteristics in determining within  $\mathbf{Y}_k$  could be different across  $k$ . For instance, it might be the case that the distribution of education endowments plays a much more important role in explaining earnings disparities than do income differences in the informal sector. More important, given the tradeable/non-tradeable asymmetries occurring during the period, returns to education could have

had an unequal effect within the different sectors. To quantify the proportion of total within-sector inequality that is attributable to differences in education endowments and other characteristics, let us define income components  $\mathbf{Y}_k$ , where  $k = \{m, s, i, a, o\}$ , as a function of  $\mathbf{X}_k$ , in the following way:

$$\mathbf{Y}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_k \quad (12)$$

Equation 12 allows for full flexibility in the functional form used in each sector  $k$ , i.e. explanatory variables as well as the value of the parameters can differ across income factor components. The estimation of equation 12 still involves some restrictive assumptions though. The conventional human capital equation regresses the log of **hourly** wages (or self-employed incomes) against vector  $\mathbf{X}$ ; in our specification, the dependent variables are the incomes from factor  $k$  in **levels**,  $\mathbf{Y}_k$ . Hence each  $\mathbf{Y}_k$  includes a labor supply component.<sup>22</sup> Furthermore, while regressing incomes from agricultural and informal activities ( $\mathbf{Y}_a, \mathbf{Y}_i$ ) we are implicitly assuming that their labor markets are complete and they are *free* of other production inputs (i.e. we are assuming *separability* in their production function). In the case of urban informal activities, the assumption is justified by empirical evidence supporting the existence of complete labor markets in the Mexican informal sector (see Maloney 1999). Bearing these assumptions in mind, we regress income factor components  $\mathbf{Y}_k$  against matrix  $\mathbf{X}_k$  where  $\mathbf{X}_k$  is defined as in our previous regression estimates (equation 11).

---

<sup>22</sup>In a fully parameterized income generating model,  $\ln(\mathbf{Y}_k) = (\boldsymbol{\beta}_k \mathbf{X}_k + \boldsymbol{\varepsilon}_k) * L_k |_{L_k > 0}$ , where  $L_k |_{L_k > 0}$  is a labor supply function and  $(\boldsymbol{\beta}_k \mathbf{X}_k + \boldsymbol{\varepsilon}_k)$  is a function estimating the log of hourly wages. See De Hoyos (2005b) for an estimation of this kind.

Table 4: % Contribution ( $s_k$ ) to Total Within  $\mathbf{Y}_k$  Inequality

	1989	1992	1994	1996	1998	2000
<b><math>\mathbf{Y}_m</math></b>						
Education	25.6	25.2	26.1	27.8	22.4	31.4
Experience	-4.2	-3.1	-4.1	-1.0	-2.4	-3.3
Gender	0.0	0.1	0.0	0.1	0.2	0.0
HH charact	11.3	9.2	8.4	7.3	9.3	7.6
Region	0.0	0.0	0.3	0.1	0.2	0.0
Residual	67.2	68.6	69.3	65.7	70.3	64.2
<b><math>\mathbf{Y}_s</math></b>						
Education	17.1	19.6	23.6	19.5	19.8	14.1
Experience	-2.8	-2.1	-2.8	-1.9	-1.7	-1.0
Gender	0.0	0.6	0.0	0.0	0.1	0.0
HH charact	14.8	13.9	9.9	12.1	9.7	8.2
Region	0.0	0.0	0.0	0.1	0.3	0.1
Residual	70.8	68.0	69.4	70.1	71.9	78.7
<b><math>\mathbf{Y}_i</math></b>						
Education	2.5	11.3	10.3	7.1	9.0	7.1
Experience	0.8	-1.5	-1.1	-0.2	-0.8	-0.5
Gender	0.2	0.0	0.0	0.2	0.0	0.4
HH charact	7.1	12.6	11.4	11.7	15.9	11.3
Region	0.2	0.0	0.1	0.3	0.8	0.2
Residual	89.3	77.6	79.4	80.9	75.1	81.5
<b><math>\mathbf{Y}_a</math></b>						
Education	3.1	1.8	3.2	6.8	27.9	1.6
Experience	-0.3	-0.4	-0.5	-0.3	-0.9	0.0
Gender	1.1	0.1	0.1	0.0	2.2	0.0
HH charact	7.9	7.4	12.1	9.0	3.8	6.8
Region	1.9	2.5	2.5	5.1	0.4	3.5
Residual	86.3	88.6	82.6	79.4	66.6	88.1
<b><math>\mathbf{Y}_o</math></b>						
Education	2.3	5.1	7.5	6.6	6.0	9.1
Experience	-0.4	-0.5	-0.9	-0.7	-0.4	-0.2
Gender	0.0	0.1	0.4	0.3	0.0	0.0
HH charact	0.7	2.5	3.5	3.9	2.7	3.5
Region	0.1	0.0	0.0	0.0	0.0	0.5
Residual	97.3	92.8	89.6	89.9	91.6	87.1

The regression results are presented in Tables 1 to 5 in Appendix A. Although there are some interesting results in the within sector income regressions, given the objective of this paper, we will center our discussion in their contribution to within-sector inequality. These results are shown in Table 4. As the human capital theory would predict, differences in years of formal education are much more closely related to differences in earnings (both the manufacturing and non-manufacturing earnings sectors,  $\mathbf{Y}_m$  and  $\mathbf{Y}_s$  respectively) than they are to differences in other sources of income. Distribution of education endowments account for around 20 to 25 percent of total earnings inequality, whereas the same factor accounts for only around 6 percent of total inequality within other sectors ( $\mathbf{Y}_o$ ). On the other hand, notice how the distribution of household characteristics are relatively more important to determine income differences in the informal and agricultural sectors ( $\mathbf{Y}_i$  and  $\mathbf{Y}_a$ ). This result suggests that the variables included as household characteristics represent important factors of production in the urban informal and in the agricultural sectors. A second interesting result is found in the regional impact to distribution. Despite the fact that regional differences account for less than 1 percent of total within-sector inequality in all urban sectors, regional differences can account for up to 5 percent of total income dispersion within the agricultural sector. This is not a surprising result given the huge differences between rural areas in the north of Mexico (basically large cattle fields) compared with the south of the country (small parcels of *ejidos* or communal agricultural production).

*Are the contributions made by the different characteristics to within sector inequalities explained by their market returns?* From Table 4 we can see that the proportion of within sector inequality explained by education endowments increased between 1989 and 1994 for all sectors. These results are, at least partly, explained by an increase in the returns to schooling during the same period (see Tables 1 to 5 in Appendix A). Hence an increase in the returns to schooling have an adverse within-sector distributive impact in all sectors of the economy. Hence, as it is the case for the whole population, skill endowments are unevenly distributed assets within sectors.

In 1996, as a result of the economic crisis and NAFTA, the returns to schooling were lower in all non-tradable sectors, i.e. formal services ( $\mathbf{Y}_s$ ), informal sectors ( $\mathbf{Y}_i$ ) and other urban ( $\mathbf{Y}_o$ ), hence educational endowments accounted for less inequality in these sectors during that year. However, quite a different pattern is shown in the tradable sectors: manufacturing ( $\mathbf{Y}_m$ ) and agricultural ( $\mathbf{Y}_a$ ). Although average returns to schooling also decreased between 1994 and 1996, the premium for higher education in these two sectors showed a important increase (see Tables 1 and 4 in Appendix A). Given that higher education endowments (Schooling\*H) are very unequally distributed among the population, a rise in its market premium was enough to compensate for

the reduction in returns to schooling and end up increasing the contribution made by education to within-tradeable sectors' inequality. The same pattern is somehow shown in the years following NAFTA and the crisis, with the proportion of inequality attributable to educational endowments decreasing in the non-tradable sectors and increasing the tradable ones. This evidence is consistent with the results found by the trade-relative wages literature and also by other studies undertaking decomposition analysis.<sup>23</sup> Our results show that the sectoral redistribution caused by the combination of the peso devaluation and the enactment of NAFTA had as a consequence an increase in the proportion of inequality accounted for by skill endowments in the tradable sector (manufacturing and agricultural).

## 5 Summary and Conclusions

The present study analyzes the underlying causes behind the high *levels* of Mexican household income inequality during the 1990s. In order to capture the sectoral disparities observed during this period, total household income was divided into income derived from activities taking place in different sectors of the economy. Our discussion focused in the importance played by sectoral disparities, skill endowments, and their market returns to account for total income inequality.

The results from conventional income factor decomposition à la Shorrocks (1982) shows how little we actually know about the causes behind the high levels of income inequality in Mexico. More than 50 percent of total income dispersion is accounted for by income derived from entrepreneurial activities and returns to financial assets. These are income sources with few structural interpretations and that nevertheless account for a large income proportion of household situated at the very top of the income distribution.

Trying to add more economic structure to the conventional non-parametric decompositions, we decompose income inequality using the semi-parametric approach developed by Morduch and Sicular (2002) and Fields (2003). We find that skill endowments accounted for, at most, 25 percent of total household income distribution in Mexico during the 1990s. Our results show that the proportion of inequality attributable to skill endowments increased during periods of macroeconomic stability and economic growth and reduced during the crisis. This pattern is, to a great extent, explained by

---

<sup>23</sup>A literature review on the trade versus relative wages studies can be found in Hanson (2003). Lopez-Acevedo (2000), De Hoyos (2005b) and Legovini (2005) undertake decompose household income distribution to analyse the importance of returns to schooling in the trade-inequality debate.



shifts in returns to schooling rather than changes in the distribution of skills. Positive shifts in the market returns to schooling are associated with increases in the proportion of inequality accounted for by skill endowments, hence indicating that skills are an unevenly distributed asset.

Applying the same semi-parametric model to decompose within-sector income inequality, we show that skill endowments are also unevenly distributed within sectors, hence increases in the market results to schooling, have an adverse within-sector distributional impact. Difference in educational endowments and their market returns account for as much as 1/4 of total income inequality in the *earnings* sectors but as little as 5 percent in the *non-earnings* sectors. The currency crisis of December 1994 together with the enactment of NAFTA caused a sectoral redistribution favoring the tradable sectors. This redistribution was associated with lower returns to schooling in the non-tradable sectors and a higher premium for well-educated workers in the tradable sector. Given the positive relationship between returns to schooling and within-sector income inequality, after 1994, skill endowments had a lower contribution to income inequality in the non-tradable sectors and a higher in the tradable sectors. These results corroborate the results found by the trade versus relative wages literature, supporting the view of a skill-biased shift in labor demand brought about by the Mexican liberalizing reforms of the 1990s.

## References

- [1] Bourguignon, F., Fournier, M. and Gurgand, M. (2001) 'Fast development with stable income distribution: Taiwan, 1979-1994', *Review of Income and Wealth*, vol. 47 (June), pp. 1-25.
- [2] Cowel, F.A. (2000) '*Measuring inequality*', Oxford University Press.
- [3] Cowel, 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*, vol. 105. pp. 421-430.
- [4] De Hoyos, R.E. (2005a) 'Constructing inequality and poverty indexes using Mexican ENIGH', available from: <http://www.econ.cam.ac.uk/phd/red29/>
- [5] De Hoyos, R.E. (2005b) 'The microeconomics of inequality, poverty and market liberalizing reforms', research paper 2005/63 *UNU-WIDER*, Helsinki.

- [6] Fields, G.S. and Yoo, G. (2000) 'Falling labor income inequality in Korea's economic growth: Patterns and underlying causes', *Review of Income and Wealth*, vol. 46, pp. 139-159.
- [7] Fields, G.S. (2003) 'Accounting for income inequality and its changes: A new method with application to the distribution of earnings in the United States', *Research in Labor Economics*, vol. 22, pp. 1-38
- [8] Feenstra, R.C. and Hanson, G. (1997) 'Foreign direct investment and relative wages: Evidence from Mexico's maquiladoras', *Journal of Development Economics*, vol. 42, pp. 371-393.
- [9] Gunatilaka, R. and Chotikapanich, D. (2005) 'Inequality trends and determinants in Sri Lanka 1980-2002: A Shapley approach to decomposition', Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia, mimeo.
- [10] Hanson, G.H. (2003) 'What has happened to wages in Mexico since NAFTA? Implications for Hemispheric Free Trade', *NBER Working Papers*, 9563, National Bureau of Economic Research.
- [11] Harrison, A. and Hanson, G. (1999) 'Who gains from trade reform? Some remaining puzzles', *Journal of Development Economics*, vol. 59, pp. 125-154.
- [12] Jenkins, S.P. (1991) 'The measurement of income inequality', in Lars Osberg (ed) '*Economic inequality and poverty: International perspectives*', M.E. Sharpe, Armonk N.Y. and London.
- [13] Lopez-Acevedo, G. (2000) 'Earnings inequality after Mexico's economic reform', mimeograph, *The World Bank*.
- [14] Lopez-Acevedo, G. and Salinas, A. (2000) 'How Mexico's financial crisis affected income distribution', *Policy Research Working Paper*, WPS2406, The World Bank.
- [15] Legovini, A., Bouillón, C. and Lustig, N. (2005) 'Can education explain changes in income inequality in Mexico?', in Bourguignon, F., Ferreira, F. and Lustig, N., editors, *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*, Oxford University Press.
- [16] Maloney, W. (1999) 'Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico', *The World Bank Economic Review*, vol. 13, 2.

- [17] Morduch, J. and Sicular, T. (2002) 'Rethinking inequality decomposition, with evidence from rural China', *The Economic Journal*, vol. 112, pp. 93-106.
- [18] Redmond, G. and Kattuman, P. (2001) 'Employment polarisation and inequality in the UK and Hungary', *Cambridge Journal of Economics*, vol. 25, pp. 467-480.
- [19] Revenga, A. (1997) 'Employment and the wage effect of trade liberalization: The case of Mexican manufacturing', *Journal of Labor Economics*, vol. 15, pp. S20-S43.
- [20] Shorrocks, A.F. (1982) 'Inequality decomposition by factor components', *Econometrica* 50, pp. 193-211.
- [21] Shorrocks, A.F. (1983) 'The impact of income components on the distribution of family incomes', *The Quarterly Journal of Economics*, vol. 98, pp. 311-26.
- [22] Shorrocks, A.F. (1999) 'Decomposition procedures for distributional analysis: A unified framework base on the Shapley value', Department of Economics, University of Essex, mimeo.
- [23] Székely, M. (1995) 'Apectos de la desigualdad en Mexico', *El Trimestre Economico*, Fondo de Cultura Economica.
- [24] Wood, A., (1997) 'Openness and wage inequality in developing countries: The Latin American challenge to East Asian conventional wisdom', *The World Bank Economic Review* 11, pp. 33-57.

# Appendix

## A Regression Results for Household Income Components $Y_k$

Table 1: Earnings from Manufacturing,  $Y_m$

	1989	1992	1994	1996	1998	2000
Schooling	103.6**	158.3**	203.8**	92.5**	86.5**	159.8**
Schooling*(H)	57.6**	99.8**	80.8**	91.7**	88.8**	124.8**
Experience	32.4*	51.2*	57.2	-19.1	15.7	64.4*
Experience Sq.	-0.3	-0.4	-0.4	0.5	-0.0	-0.5
Gender	-3.5	-246.3	-61.7	-90.5	-165.0	97.0
HH size	-121.2**	-200.5**	-122.5**	-111.5**	-157.1**	-207.9**
Dep. Ratio	-1213.0**	-899.4**	-1894.4**	-735.9*	-871.6**	-855.4
Region	147.8	0.1	-238.0	43.1	120.4	-28.4
Intercept	1214.1**	1014.6*	739.8	1464.1**	1441.8**	146.2
R-squared	0.324	0.309	0.305	0.342	0.291	0.347
N	1,779	1,525	1,831	2,134	1,681	1,524

\*,\*\* significant at the 5% and 1% level respectively

Table 2: Earnings from Other Sectors,  $Y_s$ 

	1989	1992	1994	1996	1998	2000
Schooling	94.2**	142.7**	171.2**	113.8**	106.3**	123.5**
Schooling*(H)	60.4**	74.6**	132.5**	66.8**	91.4**	91.5**
Experience	34.6**	56.2**	83.2**	12.9	33.9*	11.5
Experience Sq.	-0.3	-0.7*	-0.8**	0.0	-0.3	0.1
Gender	-50.8	440.9*	-235.0	14.2	90.5	-66.7
HH size	-131.2**	-114.7**	-196.7**	-128.7**	-174.2**	-177.6**
Dep. Ratio	-2711.9**	-2537.4**	-3298.2**	-2269.2**	-1872.1**	-2868.9**
Region	-39.8	-61.0	133.7	132.0	210.5*	209.7
Intercept	2537.0**	1099.9**	2049.3**	1906.8**	1514.2**	2575.2**
R-squared	0.288	0.313	0.300	0.294	0.281	0.216
N	5,167	4,704	5,602	6,257	4,786	4,462

\*,\*\* significant at the 5% and 1% level respectively

Table 3: Income from Informal Activities,  $Y_i$ 

	1989	1992	1994	1996	1998	2000
Schooling	65.1**	104.5**	122.1**	57.3**	63.4**	121.5**
Schooling*(H)	-21.4	-4.5	30.5	43.0	3.7	9.6
Experience	0.3	15.9	39.8	60.1**	14.8	69.8*
Experience Sq.	-0.1	-0.1	-0.3	-0.8**	-0.1	-0.7*
Gender	-201.1	-106.7	5.0	89.6	1.3	-448.7**
HH size	-102.9**	-127.7**	-185.1**	-118.6**	-109.1**	-261.6**
Dep. Ratio	-1324.0	-596.7**	-906.8**	-1027.8**	-733.7**	-950.6**
Region	143.0	28.9	95.8	129.2	156.3**	157.9
Intercept	2460.9*	1127.1**	934.6	440.5	962.7**	1086.0*
R-squared	0.104	0.217	0.202	0.186	0.245	0.173
N	2,034	1,973	3,028	3,190	2,467	1,859

\*,\*\* significant at the 5% and 1% level respectively

Table 4: Income from Agricultural Activities,  $Y_a$ 

	1989	1992	1994	1996	1998	2000
Schooling	50.4**	52.8**	76.2*	17.1*	107.1*	31.7*
Schooling*(H)	38.8	33.0	-1.1	73.8*	313.9**	42.6*
Experience	13.1	20.1	42.0	12.7*	33.7	22.6
Experience Sq.	-0.2	-0.3	-0.6	-0.1	-0.3	-0.3
Gender	451.9	-196.5	-242.5*	0.3	817.7	-19.5
HH size	-46.8**	-69.0**	-67.1**	-43.7**	-51.2**	-74.6**
Dep. Ratio	-844.5**	-610.3**	-1372.6*	-352.5**	-575.0*	-651.1**
Region	263.1**	333.6**	381.1*	276.7**	106.9	470.2*
Intercept	496.5	1009.8**	1154.6**	488.3**	-803.6	684.0*
R-squared	0.137	0.114	0.174	0.206	0.334	0.119
N	1,940	1,703	2,094	2,138	1,552	1,569

\*,\*\* significant at the 5% and 1% level respectively

Table 5: Other Income,  $Y_o$ 

	1989	1992	1994	1996	1998	2000
Schooling	80.6**	121.7**	123.9**	73.3**	79.0**	107.8**
Schooling*(H)	87.7**	104.0**	57.3**	46.5**	70.8**	91.9**
Experience	61.5*	122.0**	50.5**	31.0**	53.0*	102.2**
Experience Sq.	-0.6	-1.4**	-0.3	-0.2	-0.4	-1.0*
Gender	19.4	278.6	313.2*	203.2*	29.7	-9.4
HH size	-145.3**	-202.9**	-181.6**	-164.0**	-193.4**	-257.9**
Dep. Ratio	609.4	-745.2*	-60.9	385.5*	-227.8	199.4
Region	266.6	35.0	25.2	17.7	72.1	450.8
Intercept	-843.7	-824.6	-479.6	-259.8	11.4	-1093.6*
R-squared	0.030	0.086	0.117	0.102	0.117	0.140
N	6,841	8,774	10,575	11,744	8,955	8,229

\*,\*\* significant at the 5% and 1% level respectively