# A cohort analysis of the income distribution in Chile*

*Un análisis de cohorte de la distribución del ingreso en Chile*

CLAUDIO SAPELLI**

## Abstract

*In this paper we look at the income distribution by cohort in Chile. We construct a synthetic panel from cross section surveys and estimate the income distribution for cohorts born between 1902 and 1978. We then decompose the evolution of these distributions into age, year and cohort effects. The cohort effects show a period where inequality increases, to then decrease. We attempt to explain this evolution. The rise can be explained by variables associated with education, while the fall appears to be the consequence of a flattening of the income-age profile and hence a reduction in the returns to experience.*

Key words: *Synthetic cohorts, Income distribution, Rates of return, Education.*

## Resumen

*En este artículo analizamos la distribución del ingreso en Chile por cohortes. Construimos un panel sintético a partir de encuestas de corte transversal y estimamos la distribución de ingresos para los cohortes nacidos entre 1902 y 1978. Luego, descomponemos la evolución de estas distribuciones en efectos de edad, año y cohorte. El efecto de cohorte muestra un período donde la desigualdad aumenta, para luego decrecer. Intentamos explicar esta evolución. El incremento puede se explicado por variables asociadas a la educación, mientras la caída parece ser consecuencia de un aplanamiento del perfil de ingreso-edad y por lo tanto a una reducción de los retornos a la experiencia.*

Palabras clave: *Cohortes sintéticas, Distribución del ingreso, Tasas de retorno, Educación.*

JEL Classification: *I20, J24, J31.*

---

* I would like to thank Alejandra Candia, Rafael Sanchez, and in particular Will Mullins for research assistance in the production of this paper. Comments by Harald Beyer, Dante Contreras, Kevin Cowan, Eduardo Engel, Walter Sosa, Bernardita Vial, and the participants at the SECHI 2005/2006 Meetings and the LACEA 2005/2006 meetings are gratefully acknowledged. All remaining errors are my sole responsibility.

** Instituto de Economía, Pontificia Universidad Católica de Chile.

## 1. Introduction

The central contribution of this paper is its use of synthetic cohorts to analyze the issue of income distribution, providing a new perspective on the topic. The paper studies the evolution of income distribution for cohorts born between 1902 and 1978, using synthetic cohorts constructed from successive cross section surveys (with data taken from the Employment surveys of the University of Chile - 1957 to 2004). It decomposes this evolution into cohort, age and year effects and then analyzes the pattern of cohort effects. In particular it looks into whether these patterns can be explained by trends in the mean and dispersion of both years of education and returns to education in the cohort. To do this we estimate the rates of return to schooling in Chile for those cohorts (see Sapelli and Mullins (2006) and Sapelli (2009)).

A cohort analysis of the issue of income distribution opens up a new perspective on the topic. The evolution of cohort income inequality during the period 1902-1978 shows an interesting dynamics: inequality first increases and later decreases[1]. However, the public policy discussion in Chile regarding income distribution is made on the basis of the cross section income distribution (based on one cross section survey, be it the U. de Chile survey or the CASEN survey) which show inequality basically flat for many decades now.

However, from the point of view of designing public policy this appears to be a mistake, since it is very hard for public policy to act on the "stock" (the sum of all cohorts) income distribution, since it would imply addressing a multiplicity of causes. But public policy can act on the income distribution of recent cohorts (or "marginal" cohorts) with a much wider range of policies (by improving the quality of education, decreasing desertion, increasing tuition credit). Since these policies act at the margin, they should not be judged according to what happens to the stock. Hence this paper argues that the evolution of income distribution that is described here should be considered more indicative of what is currently happening than estimates based on cross section data for one year.

Since both methods tell different stories, the analysis of inequality using data from one cross section survey (the methodology that is most commonly used), may lead to incorrect public policy decisions. Panel data would probably be more precise, but in Chile, as is the case in many developing countries they do not exist yet: cross section surveys are all that is available.

We need to look at whether policy is having an effect on the inequality of recent cohorts. Changes in the "marginal" distribution that are sustained over time will eventually affect the "stock" distribution. Moreover, by looking at these changes we could predict what the "stock" income distribution will look like in the future (however, this is not done here). Another angle from

---

1 Inequality is measured by the Gini index. The method of estimation of these Ginis is described bellow.

which this dichotomy can be viewed is that of the difference between intra-generational inequality and inter-generational inequality. We will not look into inter-generational inequality in this paper[2], but we will look at intra-generational inequality and its evolution.

## 2. The Data

We use the Universidad de Chile "Encuesta de Ocupación y Desocupación en el Gran Santiago" surveys of Gran Santiago called for the period June 1957-2004. The sample is composed of men and women in the greater Santiago area who report non-zero work income, are aged between 18 and 65 inclusive, and who were born between 1904 (on occasion 1902) and 1978 inclusive[3].

The individuals are grouped into "cells" of individuals that were all born in the same year and are all observed at the same age e.g. men and women born in 1940 observed at age 20 (in 1960). The tables below detail the means, standard deviations, maximums and minimums of the data thusly organized. Hence we construct synthetic panels of the population born between 1904 and 1978.

| Variable | Obs. | Mean. | Std. Dev. | Min. | Max. |
|---|---|---|---|---|---|
| Age | 161,827 | 36.732 | 11.55452 | 18 | 65 |
| Work Income | 161,827 | 131326.1 | 238897 | 84.5 | 1.22e+07 |
| Education in years | 161,827 | 9.389971 | 4.399738 | 0 | 20 |
| Gini Education x cell | 161,827 | 0.2445167 | 0.0667459 | 0 | 0.5460317 |
| Gini Education x cohort | 161,827 | 0.2545859 | 0.0623036 | 0.1191496 | 0.4082756 |
| Income Gini x cell | 161,827 | 0.4652665 | 0.0911754 | 0.1521739 | 0.8339984 |
| Mean Education x cohort | 161,827 | 9.389971 | 1.449989 | 6.605911 | 12.14029 |
| Mean Education x cell | 161,827 | 9.389971 | 1.787015 | 4.153846 | 13.57282 |

2 Inter-generational inequality is marked by the existence of older generations with a low average number of years of education. Since they are present in the "stock" distribution along with generations with more education, and since over time the number of college graduates has increased, and university rates of return have also risen, it is unsurprising that the tendency of the stock income distribution has been to deteriorate or at best to stay the same. However this group does not exist in the most recent cohorts.

3 Some individuals were dropped from the sample, They were individuals that had either: 1) Zero work income; 2) Missing/wrongly coded work income data; 3) Missing/wrongly coded educational data (This excludes all data from the 1958, 1963 and 1964 surveys: they did not collect educational data); 4) Younger than 18 or Older than 65; 5) Born before 1904 (in some cases 1902) or after 1978; 6) Belong in a cohort-age cell (i.e. born in year 19wx aged yz) composed of less than 15 individuals. This group is composed of approximately 630 individuals.

## 3. Income Distribution by Cohort: What Does the Data Say?[4]

To analyze the data and estimate the cohort effects we first estimate the Gini coefficients by cohort and year. For example, we estimate the Gini coefficient for the cohort born in 1939 in every year we observe them within our sample period of 1957 to 2004. This is done for all cohorts born from 1902 to 1978 inclusive, and these Gini coefficients, calculated from an average of 80 individuals' incomes, are used as the dependent variable. We set the cut-off points at the 1902 and 1978 cohorts because working with individuals between 18 and 65 years of age implies that we observe these cohorts at least in nine different years, allowing a sufficiently trustworthy estimate of the "cohort effect."

With the data so generated we run a regression to separate cohort, age and year effects. Because of the well-known estimation difficulties that arise from the perfect co-linearity between cohort birth year, age, and year of survey, we use the method developed by Deaton (1997) to sort between these three effects. The idea behind this methodology is to run a regression between these Gini coefficients and dummies by cohort, by year and by age (regression results available from the author). The assumption used by the Deaton method to identify these three effects is that the year effects have no trend and therefore add to zero. Another assumption is that there are no interactions between the three effects.

The Deaton method, among other things, permits us to abstract from the fact that our cohorts have different age compositions. This is important since the earlier cohorts we observe only for their later years and the later cohorts only for their earlier years.

Figure 1 shows the results of applying this method[5]. We graph only the cohort effects, i.e. the coefficients of the cohort dummies. In this paper we abstract from the results of the year effect (that show that inequality increases at times of high growth and decreases with recessions) and from the results of the age effect (that show the standard result that inequality grows with age, at a declining rate). The cohort effects we show in Figure 1 are what later we call the effects estimated with a "pure Deaton" regression (i.e. only with the dummies for the three effects), to distinguish them from regressions where we attempt to explain the evolution of the cohort effects described here.

We obtain that the cohort effects explain changes of 9 points in the Gini index. Most recently cohort effects decrease systematically for cohorts born from the fifties onwards (and in particular since 1959). We can see in Figure 1 that the cohort effects describe an inverted U shaped curve from the cohort born in 1929 to that born in 1978. Starting with cohorts born in 1929 (which entered the labor market in the late forties/early fifties) there is an increase in cohort

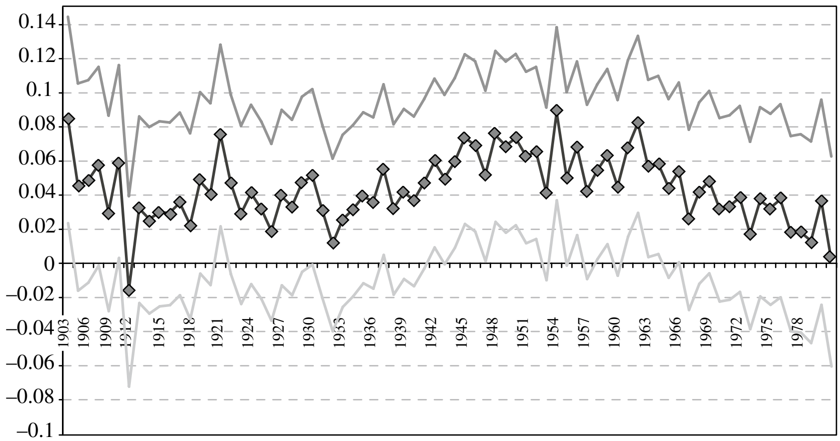
---

4 There are several changes between the data used here and that used for previous versions of this paper that are worth noting. First, the surveys of 1959, 1963, and 1964 are included in some regressions. They were excluded in the past because they lacked data on education. Data from 2003 and 2004 (which we did not have) were also added to the sample. All cohorts from 1902 to 1978 are now studied; previously we worked only with 1945-1978. Some changes relevant to the rate of return calculations were made to the inflation deflator (see Sapelli and Mullins 2006).

5 Detailed results for all regressions cited in this paper can be examined in Sapelli (2007).

income inequality that culminates in 1951. From 1951 on, and in particular from 1959 on, there is a downward trend.

FIGURE 1
COHORT EFFECTS FROM A "PURE DEATON" REGRESSION
(With a 95% confidence interval)




In Figure 1 we can appreciate that standard errors[6] are large and cohort effects are not significantly different from zero for most of the cohorts. However, there is a period (for cohorts born from 1939 to 1961) where most (20 out of 23) of the cohort effects are significantly different from zero. Thereafter they return to being statistically zero. Hence one could say that while there are no cohort effects at the beginning and the end, they exist for the mid period and hence income inequality is significantly larger for cohorts born from 1939 to 1961. However, given the low mass in each cell, standard errors are large, and most cohort effects are not significantly different from 0.04. To diminish the problem of large standard errors, we use a standard methodology in cohort analysis to increase mass in each cell and achieve more precision in our estimates[7]: we work with moving averages of several cohorts instead of with each cohort individually. We therefore work with moving averages of 3 (MA3) and 5 cohorts

---

6 We report here results with normal standard errors. Robust standard errors were also used to correct for some indications of heteroskedasticity caused by the fact that our dependent variable (Ginis by cohort and year) are being calculated from different numbers of individuals' incomes e.g. some are calculated from 200 individuals' incomes, while others were based on 15 individuals (the average number of individuals used to calculate the Ginis is 80). However, results with normal SE or robust SE do not differ much.

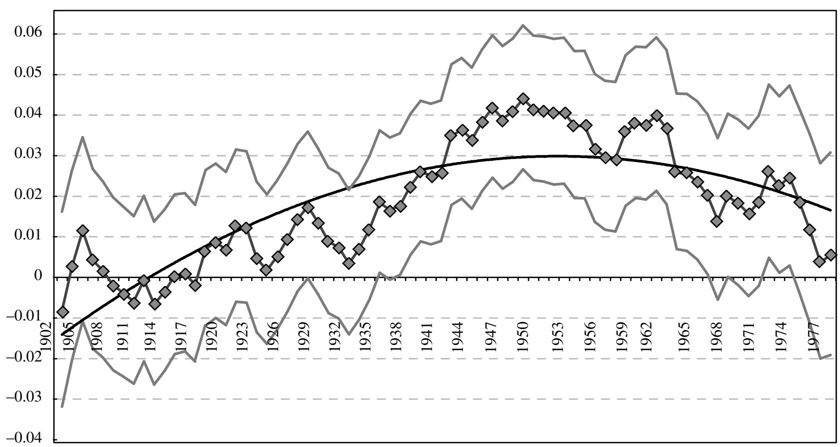
7 This is standard procedure in the literature (see, for example, Eckstein and Nagypal "The Evolution of U.S. Earnings Inequality: 1961-2002," Federal Reserve Bank of Minneapolis Quarterly Review 28 (2), 2004).

(MA5)[8]. In each moving average we include all the data of 3 or 5 generations to characterize the cohort at the centre of the interval. We only report here the results for the MA5.

The MA5 estimates are shown in Figure 2. They show basically the same evolution seen in Figure 1, however the inverted U now covers the whole period: first a rise in inequality and then a fall. The statistical significance rises and the period where cohort effects are significantly different from zero is now more extensive and ALL cohort effects are significant within this period. The period with significant cohort effects is now for the generations born between 1935 and 1965. Most importantly, the cohort effects show a statistically significant pattern, in particular regarding the rise in inequality which is significant at the 5% level. Regarding the fall towards the end, it is not significant at the 5% level, but it is significant at the 10% level.

Regarding the magnitude of the changes in inequality, in the case of the estimation without moving averages, the cohort effects rise 7 points between 1929 and 1951, then are stable from 1951 to 1959, and then drop 8 points. In the case of the MA5 estimates, the rise occurs from 1911 to 1947 (and is of 5 points) and then there is a drop of about the same magnitude (4 points)[9]. As is usually the case, trends are softened when working with moving averages.

FIGURE 2
COHORT EFFECTS FOR A "PURE DEATON" REGRESSION WITH COHORTS DEFINED AS A MOVING AVERAGE OF 5 COHORTS
(Shown together with a 95% confidence interval)




8 We do this only for the "pure Deaton" regression but not for other regressions in this paper since the generation of variables by cohort in this definition is cumbersome (especially for the rates of return) and interpretation of results is not straightforward.

9 Quantitatively these falls can be appreciated in the following manner. If we take into account a Gini of between 0.5 and 0.55 (which is the result from cross section estimates) a decrease of 8-9 point would imply a fall of 16-18% in the index. One of 4-5 points would imply a fall of about 8-10 percent.

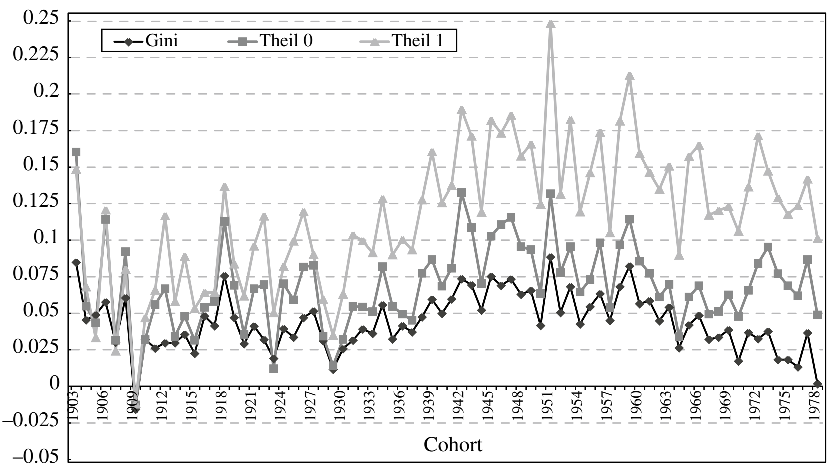
It is important to note that the evolution of inequality we have described (i.e. the inverted U shape) is robust to working with other inequality indexes. In particular we have verified this using both the Theil 0 and Theil 1 indexes and several percentile indices of dispersion, such as the interquartile range, as the dependent variable. The evolution of the different percentiles tells an interesting story that is discussed in the following section.

## 4. Robustness of the Evolution of Inequality

In this section we report the cohort effects estimated with the Theil 1 and Theil 0 indexes and compare them with the Gini index. We also analyze the evolution of different percentile measures.

In Figure 3 we can see that the evolution of the Theil 1, Theil 0 and Gini indexes are very similar. The three show similar behavior throughout. In all there is a period of increase in income inequality, followed by a period of decreasing inequality. In all the period of increase includes the 1929-1951 cohorts, though in the case of the Theil 1 it is possible to argue that the trend to increase started beforehand. The three show a drop, but the duration of the drop varies: it is longer for the Gini index (up until the end of the series) but shorter in the other two series (up to the 1964 cohort).

FIGURE 3
COMPARISON OF THE EVOLUTION OF THE THEIL 1, THEIL 0 AND GINI INDEXES OF INCOME DISTRIBUTION




Since we observe a somewhat different behavior between the three measures we look deeper into the evolution of the different parts of the distribution to try to assess why they differ. Possibly the most notable similarity is in the peaks and troughs of the three series. However, there is a clear difference in the magnitude of the fall in the later years. Excluding both the 1909 cohort and the last cohort, we find that in the case of the Gini index, the fall is larger in magnitude than the rise. However in the case of both Theil indexes the fall is smaller than the rise (two thirds of the rise in the case of the Theil 1 and three quarters in the case of the Theil 0).

It is useful to examine the evolution of the P90-P50 difference (the upper half of the distribution), of the interquartile range (the middle of the distribution) and the P50-P10 difference (the lower half of the distribution). The interquartile range behaves pretty much like the P90-P50 difference but the upper half of the distribution and the lower half behave very differently. While there is a systematic decrease of the inequality in the bottom half of the distribution (up until the 1972 cohort), inequality in the upper half rose between the 1929 cohort and the 1941 cohort, then fell between the 1941 and 1962 cohorts and then rose again.

The rise in inequality in the upper half of the distribution explains most of the raise in overall inequality and then the fall occurs when both halves of the distribution are following the same trend and are decreasing in inequality. However, the periods when the percentiles fall and when the overall indexes fall do not coincide. Hence we take a more detailed look at the different parts of the distribution, looking at the differences between the following percentiles: $95^{th}$ and $80^{th}$, $80^{th}$ and $65^{th}$, $65^{th}$ and $50^{th}$, $50^{th}$ and $35^{th}$, $35^{th}$ and $20^{th}$, and finally $20^{th}$ and $5^{th}$. We omit their presentation not to clutter the paper with too many numbers. However, the conclusions from their analysis follow.

The best way to analyze these data is to separate the period into three sub-periods (according to the evolution of the Theil and Gini indexes), and see which parts of the distribution contribute to explain the evolution of the overall indexes in each of the sub-periods. We separate the period into the 1903-29, 1929-1951 and 1951-78 sub-periods, which correspond to periods of: trend stability (but with great variation), rise, and fall of the indexes, respectively. We analyze the evolution of these percentiles by looking at the simple correlations between the different measures.

In the first period it is mostly the upper part of the distribution that leads the evolution of the overall indexes. The lowest part (P20-P5) has a life of its own, since the correlation with most other percentile differences is negative.

In the second period the same is true again. The overall indexes are lead by the upper part, while the lower part has the opposite behavior (there is an improvement in inequality in the lower part while there is deterioration in the upper part). Again the P20-P5 is an outlier, this time having the same behavior as the upper part of the distribution (deterioration). Hence the period when inequality worsens is a period where this evolution is mostly explained by an increase in inequality in the upper half of the distribution. The P95-P80 measure has a steep increase in the period 1923-1941 and then stays stable. These cohorts entered the labor market in the forties and fifties.

If we take into account that most of the present cross section inequality is explained by inequality in the upper quintile (see Torche (2005)), explaining this deterioration that is lead by the upper part of the distribution is key in explaining the high level of inequality in Chile today.

In the last period (the period where the decline occurs) we observe that the overall indexes are highly correlated with the bottom of the distribution. Hence the reason behind the increase is deterioration at the top, but the reason behind the decrease is an improvement at the bottom. That leaves the distribution exactly where Torche describes it: with very good distribution at the bottom, but unequal at the top.

In the section that follows we will attempt to explain the evolution of the overall indexes through the evolution of human capital variables.

## 5. How Education Evolved Over the Period

What we have found shows that there is an interesting dynamic in the behavior of income distribution. This runs counter to the common observation that income distribution has not changed in Chile for many years. This stagnation is frustrating, specially given the fact that during the 20th century the population systematically increased its average education level, and the dispersion of the education level also fell steadily, as can be seen in Figures 4 and 5. These trends make clear that education cannot explain both the increase and the decrease in inequality, since these variables increase or decrease continuously over the period. We will address this issue later.

Figure 4 shows that mean education, in particular starting in 1939 has grown steadily. It climbed at a higher rate in the period 1939-1958, which roughly coincides with the period in which the income distribution deteriorated.

Figure 5 shows how the dispersion of education within cohorts (as measured by the Gini index of education) has decreased steadily. In this case acceleration in the rate of decrease is observed for the cohorts born between 1947 and 1961 (those going through the education system starting in the mid fifties to the late sixties). However, this does not coincide with the period where income inequality decreases.

FIGURE 4
MEAN EDUCATION BY COHORT (1902-1978)

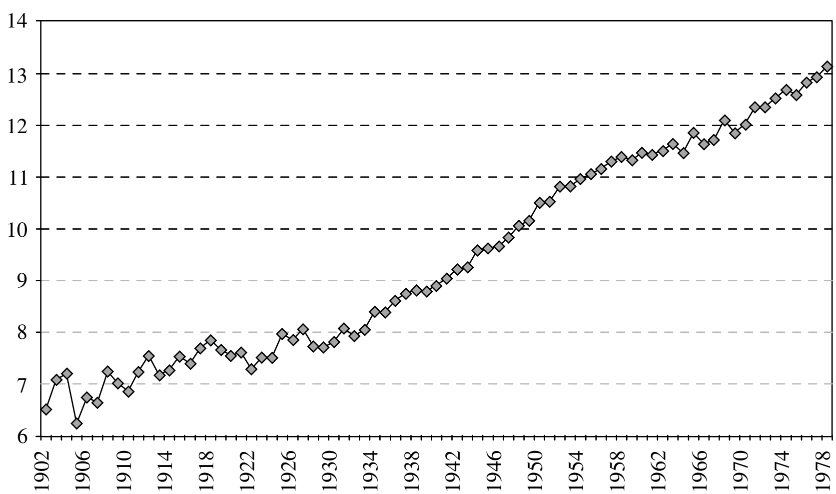
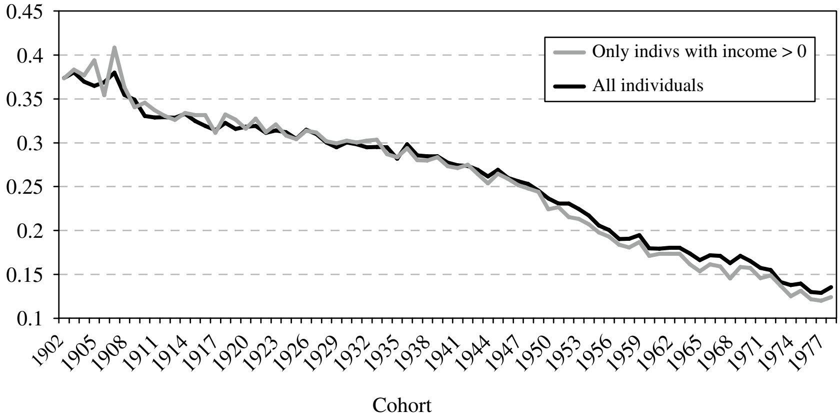


FIGURE 5
DISPERSION OF EDUCATION BY COHORT
(Cohorts born 1902-1978)




To link the evolution of education to that of income inequality we use the human capital model and in particular what is known as the "Mincer equation". Hence our theoretical framework is one in which the evolution of income is determined by changes in human capital and in the rate of return to human capital. As a result, the evolution of the income distribution is determined by the means and standard deviations of the distribution of education and of the rates of return to education for each cohort.

The traditional Mincer equation (where we ignore experience) is:

$$\ln Y = A + RS \tag{1}$$

Where Y is income, S is education, and A is a constant and R can be interpreted as the rate of return to education. If we take the variance of equation (1), assuming S and R are independent[10] we obtain:

$$V(\ln Y) = [\text{mean }(R)]^2 V(S) + [\text{mean}(S)]^2 V(R) + V(S)V(R). \tag{2}$$

Where V(S) and V(R) are the variance of education and the variance of the rate of return, respectively.

Hence, all else equal, the partial derivatives of the variance of log income with respect to mean education, variance of education, mean returns to education and variance of rates of return, are all positive. This provides an explanation for the fact that income inequality increased at the time when mean education rose more rapidly. From the point of view of the human capital Mincerian framework, during the 20th century, the evolution of the distribution of education

[10] This is a reasonable assumption if the returns to education are mainly driven by demand shocks rather than by supply shocks.

provided two countervailing forces: one force increasing income inequality (the increase in the mean) and the other force decreasing inequality (the decrease in dispersion)[11].

Since the inverted U shape of income distribution cannot be explained by variables that follow basically a linear trend, it is possible that the returns to education play a part in this story – the topic of the following sections.

## 6. How Rates of Return Evolved

Here we will only report findings and discuss how they may explain the evolution of income inequality. Further on we will run a more formal test of the influence of education and rates of return on income inequality[12]. Alternative estimates of rates of return are reported and discussed in Sapelli and Mullins (2006). The results are shown in Figure 6. There we include the "returns" to complete university education (i.e. the number of percentile points that persons climb in the income distribution if they have complete university education as opposed to having complete secondary education) and complete secondary education (the same, comparing complete secondary with complete primary). We omitted primary education (returns for the seventh and eighth years here have been steady at about five percentile points).

As can be seen in the Figure 6, returns for secondary school were higher than those of university up until cohorts born in 1927, when both lines cross for the first time. Both returns continue at similar levels until 1943. From cohorts born in 1944 and after returns diverge sharply, with returns to complete university education implying changes in the income distribution of up to 45 percentile points, and with returns to complete secondary education falling to about 5 percentile points. The maximum for one series and the minimum for the other coincide in

[11] This process will end with the mean reaching a plateau, as it has in developed countries, at which point the dispersion of education will be the only force driving income inequality. Nonetheless, we are not there yet, and we will probably have to wait for most people to complete secondary education before this occurs.

[12] What we have here are "rates of return" estimated as how many percentile points in the income distribution a person ascends within his own cohort with additional education. These "rates of return" we can estimate for the full sample of cohorts. What we do is basically a Mincer regression in which we do not use income as the dependant variable but the percentile where the person is in the income distribution of his cohort. The coefficient of education gives us a measure of how many percentiles an additional year of education will lift a person in the income distribution of his own cohort. We work with a spline and interactions by cohort to estimate returns to different levels of education and sheepskin effects for different cohorts. See Sapelli and Mullins (2006) for a more detailed discussion of the methodology. This measure of the return to education is not sensitive to either the correction used for inflation (very relevant for periods of high inflation and dubious price statistics e.g. Chile 1970-78), or highly determined by the first five or so income flows, both problems faced by traditional rates of return. Moreover, the latter approach requires data on a cohort from age 12 onwards, making it impossible to estimate rates of return for cohorts born before 1945 (our earliest data is the 1957 survey), a problem sidestepped by the regression structure of the percentile approach. For alternative measures of the rates of return to education see Sapelli (2009).

time, for cohorts born in the early seventies. Then there is a sharp reversal in the trend and both returns tend to equality at about 30 percentile points.

As we will see, the beginning of the period, when returns diverge, coincides with the period when income distribution deteriorates. However returns converge too late to explain the decline in inequality that starts with the cohorts born in the fifties.

FIGURE 6
"RATES OF RETURN" TO UNIVERSITY EDUCATION (LIGHT) AND SECONDARY EDUCATION (DARK)

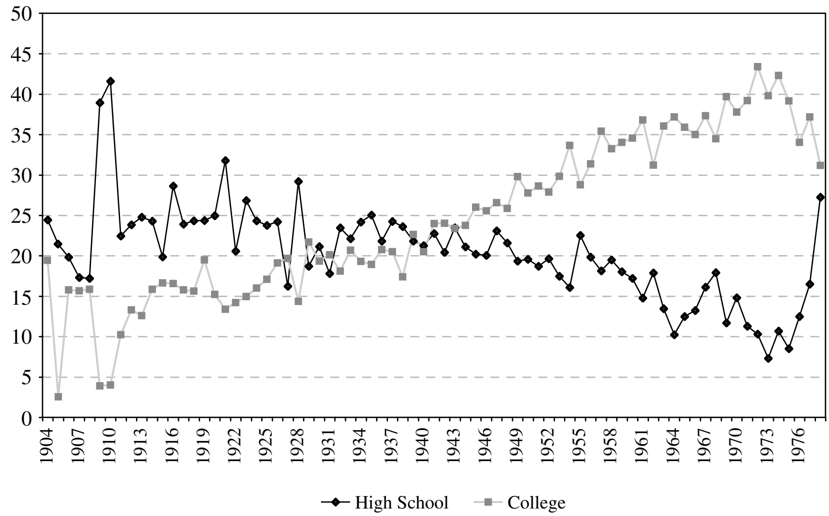



Figure 7 shows the relative returns to different levels of education. This is estimated as the ratio of the number of percentiles you climb by completing one level over the number of percentiles you climb by completing the other. In the case of primary school we use the number of percentiles a person that completes the last two years of primary ($7^{th}$ and $8^{th}$ grades) climbs in the income distribution.

The dark line tells us the same story as Figure 6: relative returns of university compared to secondary schooling are approximately flat (and similar) for most of the period, and start rising (at an increasing pace) with the cohorts born in the mid forties. They then fall sharply towards the end of the period.

The ratios of university and secondary to primary tell a new story. Both ratios rise until the late forties and fall after that, but the ratio of secondary to primary falls more sharply. Since these ratios are an indication of how dispersed the rates of return are, and that the dispersion of returns increase income inequality, they may be part of the explanation of the trends in cohort income inequality. This is more probable since these broad trends do coincide with those in income inequality, an issue we will examine more formally later.

FIGURE 7
RATIO OF RETURNS: UNIVERSITY/SECONDARY (DIAMONDS), UNIVERSITY/PRIMARY (SQUARES) AND SECONDARY/PRIMARY (TRIANGLES)

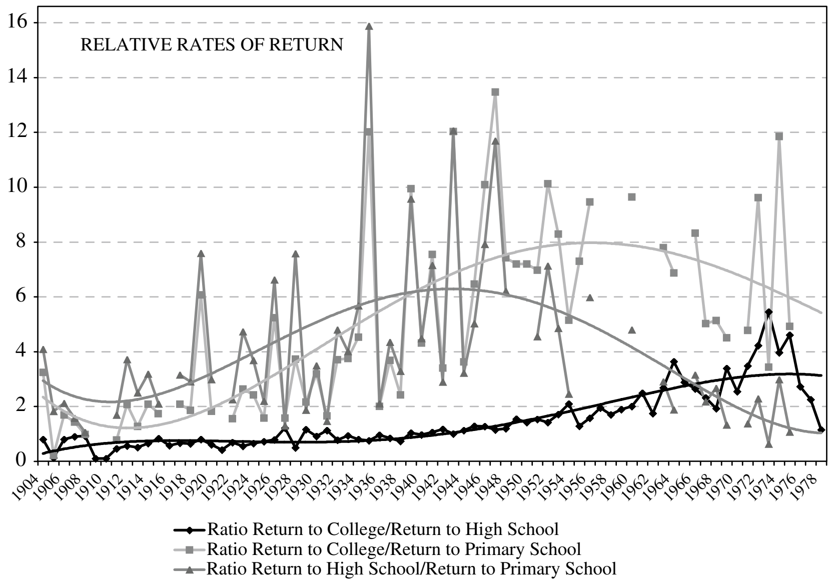
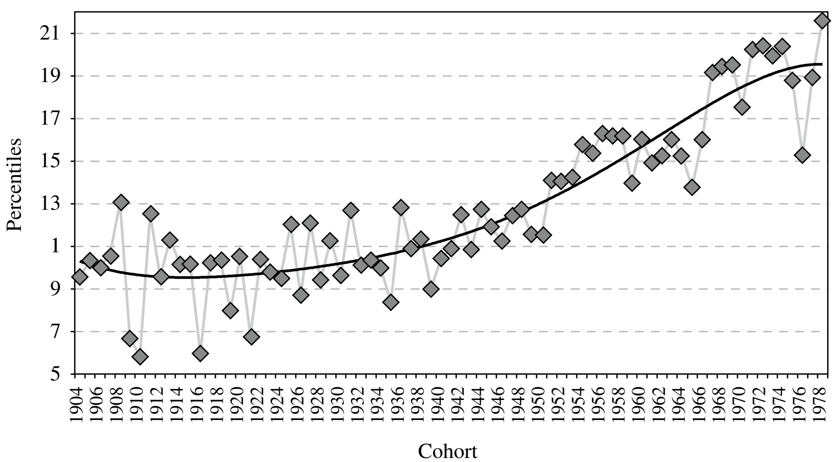



FIGURE 8
MEAN 'RETURN' TO EDUCATION MEASURED IN PERCENTILES
(Weighted average of 'returns' to the three levels)




Equation (2) above tells us that average returns (given dispersion) will increase income inequality. The data shows average returns increasing for most of the century and stabilizing towards the end.

In sum, most of the indicators we have looked at could contribute to explain the increase in inequality, but the only one that appears to contribute to the explanation of the most recent downward trend is the dispersion of returns. To take a more detailed look at these, we look at "sheepskin effects" (the returns to finishing a level, be it primary, secondary or university, the actual values are not reported here), the existence of which will contribute to increasing the dispersion of rates of return per year of education.

The estimates show positive sheepskin effects for most of the period. They are of a similar magnitude for about a decade starting with cohorts born in 1918. They then diverge slowly, and starting in the late thirties they diverge strongly, to converge again towards the end of the period (starting in the mid sixties). This confirms that there may be an important contribution of the evolution of the dispersion of rates of return in the reduction in income inequality.

## 7. Do Education Variables Explain the Trend in Cohort Income Inequality?: A Formal Test

In this section we add variables to the "pure" Deaton regression to see if the trend in cohort income inequality is explained by the evolution of variables related to the level of education of the cohort, its dispersion, the rates of return of the cohort and its dispersion. If when we add them, the cohort effects disappear, then these variables explain the cohort effects.

TABLE 1
SUMMARY OF REGRESSIONS (EXCLUDING COEFFICIENTS OF DUMMY VARIABLES –COHORT, AGE AND YEAR EFFECTS)

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| | Coefficient | Coefficient | Coefficient | Coefficient |
| Rate of return (mean) | –0.0067596 | na | –0.0068581 | na |
| College/High School | 0.0091804 | na | 0.0122137 | na |
| College/Primary | 0.0008937 | na | 0.0009814 | na |
| mean education | 0.0241826 | 0.0241826 | 0.0035959 | 0.0062458 |
| Gini education | 0.5091556 | 0.5091556 | dropped | dropped |
| Number of obs | 2025 | | 2025 | |
| R-squared | 0.6719 | 0.6719 | 0.6479 | 0.6479 |
| Root MSE | 0.05804 | 0.05804 | 0.0601 | 0.0601 |

All coefficients significant at 1% level.

| | |
|---|---|
| Column 1 | DEATON PLUS EDUCATION PLUS RETURNS (CELL DEFINITION) |
| Column 2 | DEATON PLUS EDUCATION (CELL DEFINITION) |
| Column 3 | DEATON PLUS EDUCATION PLUS RETURNS (COHORT DEFINITION) |
| Column 4 | DEATON PLUS MEAN EDUCATION AND GINI EDUCATION (BY COHORT) |

We estimate the additional variables both by cohort and by age-cohort cell. In the first case we look into variables that could explain the cohort effect as estimated by a pure Deaton regression. In the second case we use variables that could explain part of both the cohort effect and the age effect. Hence we implicitly introduce an interaction term that implies modifying slightly the traditional Deaton methodology. We do this because, as we have seen, trends in education are continuous during the century (in particular for the Gini of years of education and the mean level of education achieved), and hence cohorts are systematically different in these respects. The relatively crude separation technique offered by Deaton may confound what in reality are the effects of the changes in education (and hence cohort effects) with age effects. Since older persons in every year have a higher degree of education inequality, the method may attribute this to pure age effects when in reality they are cohort effects.

The point is that since there is a correlation between what happens by cohort and what happens by age, it is possible that the Deaton methodology is not correctly separating cohort and age effects. More to the point, it is possible that the reduction in inequality due to cohort effects is even larger than the one showed by the pure Deaton regression, since part of the effect of the reduction in the Gini index of education may incorrectly be attributed to age (being younger) rather than cohort. What one would expect if this is true is that the age effects would be appreciably reduced when using the age-cohort characterization of education variables (only the mean and dispersion of the years of education is changed in the age-cohort specification, not the variables associated with the returns to education which are necessarily estimated by cohort).

Regressions results with variables defined by cohort are not discussed here though they are reported in Table 2. Variables defined by cohort do not help us explain the evolution of cohort inequality.

TABLE 2

COHORT EFFECTS FROM TROUGH TO PEAK AND FROM PEAK TO TROUGH FOR REGRESSIONS INCLUDING DIFFERENT VARIABLES

(Changes in points of the Gini index)

| Regression | Variables defined by cohort | | Variables def. by cohort-age cell | |
|---|---|---|---|---|
| | Inequality increases | Inequality falls | Inequality increases | Inequality falls |
| Pure Deaton | 6 (1936-1951) | 9 (1952-1978) | 6 (1936-1951) | 9 (1952-1978) |
| Plus education | 5 (1937-1951) | 11 (1952-1978) | 3 (1909-1951) | 12 (1952-1978) |
| Plus education and returns | 5 (1936-1951) | 8 (1952-1978)<br>12 (1952-1977) | 0 (1909-1951) | 8 (1952-1978)<br>13 (1952-1977) |

## 8. Regressions with Variables Defined by Cohort-Age Cell

Figures 9 and 10 show the cohort effects after we control only for mean years of education and Gini of education (Figure 9) and for those variables plus variables that characterize the distribution of returns (Figure 10). Regressions are summarized in Table 1.

FIGURE 9
COHORT EFFECTS, WITH MEAN AND GINI EDUCATION ADDED TO PURE DEATON REGRESSION, VARIABLES DEFINED BY CELL

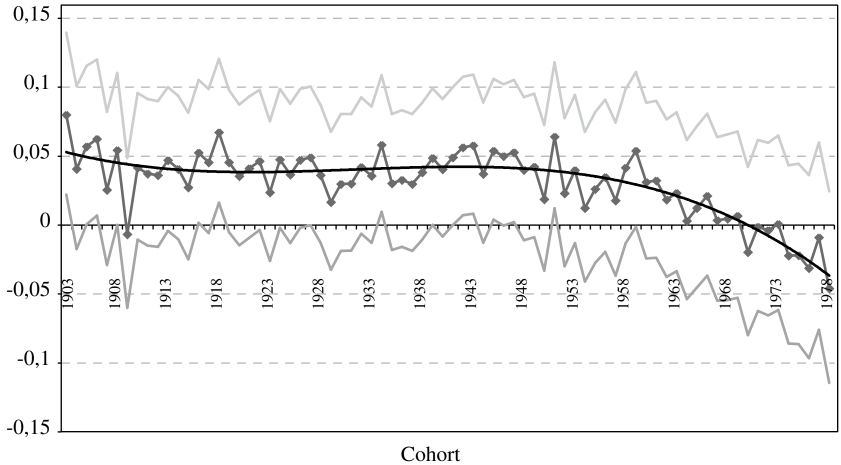
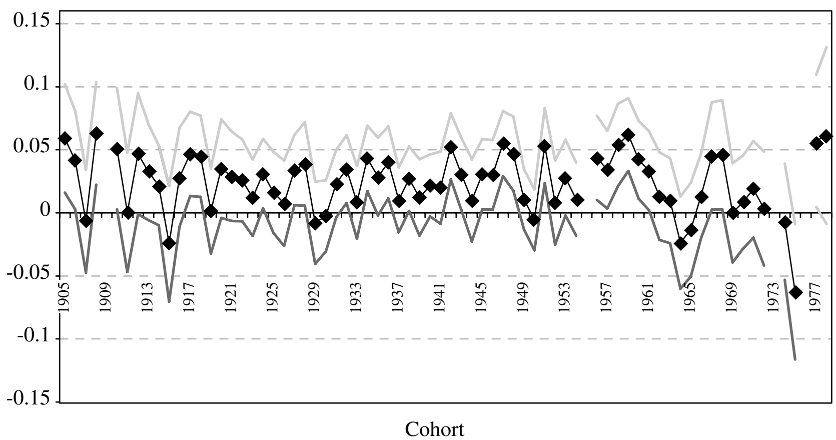



FIGURE 10
COHORT EFFECTS, REGRESSION INCLUDING EDUCATION AND RETURNS TO PURE DEATON, EDUCATION VARIABLES DEFINED BY CELL




The key conclusions from this exercise are as follows. When controlling for the number of years of education (mean and dispersion) the increase in inequality disappears and the drop in inequality is accentuated, showing that education variables fully explain the increase in inequality but do not explain the reduction. Actually, the evolution of the education variables would have justified a further increase in inequality that is not observed. Hence the unexplained fall once education variables are incorporated is larger. When we also control for returns (mean and dispersion, see Figure 10) then we observe that returns do explain part of the decline in inequality.

What is interesting is that the evolution of the number of years of education and of returns completely explains the increase in inequality from 1936 to 1951. However, they would have lead to a further increase in inequality during the period when cohort income inequality falls, hence the drop is accentuated when one includes education variables. This is so with the exception of the last few years (in particular the last-1978) where returns do appear to explain the drop in inequality.

It is also interesting to note that when we use the data by cell, the education variables explain 10 points of the Gini index that in the other specifications are explained by the age effects. Age effects go from zero to 35 points as a cohort ages when education variables specified by cell are not included, but go from zero to 25 points when they are. This proves that there is a possible absorption of cohort effects by the age effects in specifications such as the "pure Deaton" regression.

We can conclude that all the specifications we have tried, even though they do a fair job of explaining the rise, fail to explain the fall. Up to now we have constrained the regressions to have the same coefficients for the three periods we are attempting to explain. We will now relax this assumption.

## 9. What Explains the Drop in Income Inequality?

If education variables do not explain it, then what is it that explains the drop in cohort income inequality? One candidate that we have left out of our human capital framework is experience. A drop in the returns to experience would lead to a decrease in cohort inequality. And there is evidence that this has occurred in the data.

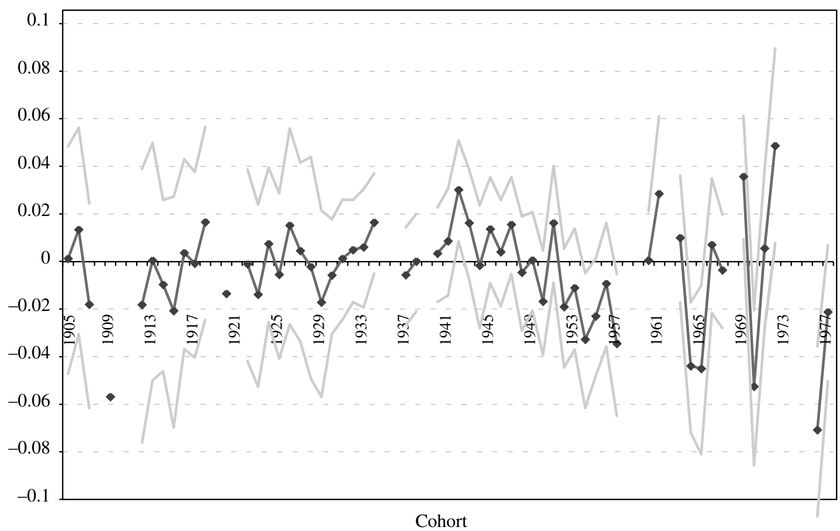
People born between 1953 and 1957 entered the labor market in the early to mid seventies. At that time other studies (Haindl (1987), Sapelli (1990), Gill, Haindl, Montenegro and Sapelli (2002), and Lima and Paredes (2004)) show an important change in the behavior of the labor market with a significant increase in turnover. A higher rate of turnover implies less accumulation of specific human capital and hence less increase of income with age.

This is very good evidence to proceed to the next step, that is, to split the sample. What we are saying here is that the age effects may change as cohorts become younger. But in our methodology we are forcing the age effects to be the same across all generations. We split the sample into 3: the first period when cohort effects are constant (1902-1929), the period when they are rising (1929-1959) and the period when they are falling (1959-1978)[13]. Hence we run the same regressions we were running beforehand (with number of years of

[13] We use the "pure Deaton" cohort effects for the regression without moving averages to define these periods. We tried with a split in 1951 instead of 1959 but the latter seams to approximate better the structural break.

education and returns) for the three subperiods (i.e. three separate regressions). These are our "final" regressions.

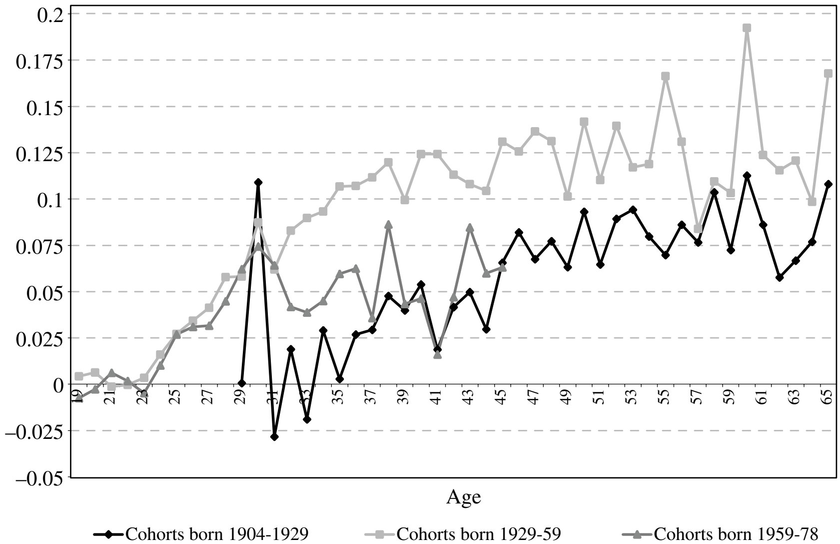
FIGURE 11
COHORT EFFECTS FROM SPLIT SAMPLE




As can be seen in Figure 11, this specification manages to eliminate most of the cohort effects, and the ones that continue to be there do not show any pattern. This specification, therefore, has achieved to explain both the rise and the drop: the rise through education variables and the drop through the returns to experience. However it should be noted that splitting the sample diminishes the power of our tests. This can be seen by the wide standard errors for the first subperiod or the wild change from significant negative to significant positive cohort effects in the case of our third subperiod. We proceed to look at the differences in age effects for the three periods in Figure 12.

Figure 12 shows that the age inequality profile was much steeper during the intermediate period (when inequality was rising) than in the other two periods. In the intermediate period the inequality due to age effects added up to 15 points to the Gini index. However in the other two periods they did not add more than 10 points. These five points are key to the explanation in the fall of inequality. Although the higher education of the latter sample would justify a pattern of much steeper increase in income by age, we observe a less steep increase than that observed in the mid period. This shows that returns to experience must have decreased. Since then the fanning out of income with age is lessened, cohorts are more equal. The cause of this decrease in the returns to experience could be the increase in turnover (but this is a hypothesis that requires testing, we leave this for future work).

FIGURE 12
AGE EFFECTS FOR 3 PERIODS IN SPLIT SAMPLE




## 10. Conclusion

For cohorts born from 1902 to 1978 one observes three periods in the evolution of cohort effects in income distribution. At first it is flat. Starting with cohorts born in the early thirties there is a deterioration that continues up to cohorts born in the fifties. The income distribution then undergoes an improvement, starting with cohorts born in the early fifties and especially after the cohort born in 1959[14].

The improvement is statistically significant and numerically important: of about 9 points in the Gini index (or about 20% considering a Gini index of 0.50).

We attempt to explain this evolution controlling for variables that are important in the human capital framework: mean and deviation of education levels and rates of return. This evolution does help understand the evolution of inequality, and in particular it explains fully the increase from the thirties to the fifties. Once one includes these variables, the cohort effects in those decades do not show deterioration in income inequality. However, these variables do not explain the recent improvement. In some regressions the drop even increases, implying that education variables in the period would have contributed to a further deterioration. Hence, once one controls for education variables, the gap to explain is even larger. The only variable that helps in explaining the decrease

[14] About half of the working population was born before this cohort. Hence this improvement at the margin is not yet discernable in the stock. But the improvement is already present, and visible if one looks at the marginal instead of the stock distribution.

is the evolution of rates of return, specially the drop in means and dispersion that has occurred recently (though this does not explain what happened when the decline in inequality started).

What appears to be driving the improvement is a reduction in the return to experience and hence a decrease in the "fanning out" that occurs after persons with different educational levels leave school. There has been a reduction in the increase of inequality with age.

## References

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy,* Baltimore, Johns Hopkins University Press.

Eckstein, Z. and E. Nagypal (2004). "The Evolution of U.S. Earnings Inequality: 1961-2002", Federal Reserve Bank of Minneapolis Quarterly Review, Vol. 28, Nº 2, December 2004, pp. 10-29.

Gill, I. S., E. Haindl, C. E. Montenegro and C. Sapelli (2000). Chile: Has Employment Become more Precarious?", in I. S. Gill and C. E. Montenegro (eds.), *Readdressing Latin America's "Forgotten Reform:" Quantifying Policy Changes in Argentina, Brazil, and Chile,* World Bank.

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

Paredes, R. & V. Lima (2004). "Labor Market Regimes and Mobility through a Markov Chain in Chile", *Econometric Society 2004 Latin American Meetings* 317, Econometric Society.

Sapelli, C. (2007). "A Cohort Analysis of the Income Distribution in Chile". *Working Paper* 290, Department of Economics, Pontificia Universidad Católica de Chile.

Sapelli, C. (2009). "Los Retornos a la Educación en Chile: Estimaciones por Corte Transversal y por Cohortes". *Working Paper* 349, Department of Economics, Pontificia Universidad Católica de Chile.

Sapelli, C. and W. Mullins (2006), "Returns to Schooling by Cohort in the Twentieth Century in Chile. An Analysis Based On Synthetic Panel Data". Mimeo, 2006.

Torche, F. (2005). "Unequal but Fluid: Social Mobility in Chile in Comparative Perspective". *American Sociological Review* 70 (3): 422-450.