UNITED NATIONS
UNIVERSITY
UNU-WIDER

DRAFT

This paper is a draft submission to

Inequality—Measurement, trends, impacts, and policies

5–6 September 2014 Helsinki, Finland

This is a draft version of a conference paper submitted for presentation at UNU-WIDER's conference, held in Helsinki on 5–6 September 2014. This is not a formal publication of UNU-WIDER and may reflect work-in-progress.

THIS DRAFT IS NOT TO BE CITED, QUOTED OR ATTRIBUTED WITHOUT PERMISSION FROM AUTHOR(S).

Analyzing Income Distribution Changes:

Anonymous versus Panel Income Approaches

Gary S. Fields
Robert Duval-Hernández
George H. Jakubson*

August 19, 2014

Abstract

For decades, the study of changes in income inequality among anonymous individuals has been the leading way of gauging who benefits and who is hurt as a result of the pattern of economic growth or decline. A different, less common approach is to utilize data on a panel of people and assess their pattern of panel income changes. This paper summarizes our theoretical findings on how the answers provided by these two methods can be reconciled, and it empirically illustrates this reconciliation using earnings data from urban Mexico. Finally, it examines how our view of inequality is altered if instead of looking at inequality at a point in time we focus on the inequality of average earnings. We look at the trends of short-run inequality and of inequality in average earnings. We also explore what factors account for their evolution. In general, earnings changes are convergent, irrespectively of whether inequality rises or falls. This is caused by a small fraction of individuals experiencing large and convergent earnings changes. The equalization that earnings changes bring over a year is mainly driven by changes in the employment status of workers.

JEL Codes: J31, D63
Keywords: Income Inequality; Economic Mobility

*Fields is at the Department of Economics, Cornell University, IZA and WIDER. Jakubson is at the Department of Economics, Cornell University. Duval-Hernández is at the University of Cyprus and CIDE.

I. Motivation

The economic literature analyzing income inequality has devoted most of its attention to comparing the dispersion of income distributions over two or more points in time. By looking at how the shape of this distribution has changed over time, this literature has compared anonymous individuals at different points in time. The "anonymity" in this comparison arises because it looks at the income of whichever individual is in the p'th position in each distribution, regardless of whether that is the same person in one distribution as in another. Analysts compare income distributions in this way, either because they do not know which individual is which in the two distributions, or if they do know, they choose to ignore the specific identities of the different individuals and talk about "the poorest," "the richest", and so on.

An alternative approach for analyzing distributional changes is to follow identified individuals over time using panel data and see how their incomes evolve. By tracking individuals over several periods, this alternative approach removes the aforementioned "anonymity" from the analysis of income distributions. More specifically, panel data can be used to analyze changes in the shape of the income distribution, but it can do more by also displaying the evolution of income for each individual that appeared in the initial survey (leaving aside issues of attrition).

To the extent that people move around in the income distribution, the answers obtained by looking at anonymous individuals in a given income quantile might or might not coincide with the ones derived by identifying those individuals who started in a given income quantile and tracking those individuals over time. For instance, the answer to whether the people in the bottom 10% of the income distribution became poorer might change depending on whether we look at the income of the anonymous bottom 10%, or whether we track with panel data the incomes of those who initially were in the bottom 10%. In other words, the standard inequality analysis follows the evolution of incomes of whoever is in the bottom 10%, irrespective of whether they are the same people or not, but the panel approach tracks the income change of those who started in the bottom 10%, but who might or might not have moved to other points in the income distribution.

In this paper, we have three goals. First, we summarize in an accessible manner our theoretical findings on how the answers provided by the anonymous and panel methods can be reconciled. Second, we empirically illustrate this reconciliation using earnings data from Mexico for several decades, including periods of economic growth and decline and of rising and falling inequality. Finally, we examine how our measures of inequality are altered if instead of looking at earnings inequality at a point in time, we focus on the inequality of average earnings. More specifically, we present trends in single-period and multiperiod earnings inequality, and we explore what observable factors at the individual and aggregate level account for their evolution.

II. Reconciling Anonymous and Panel Income Changes

There is a large literature on how to measure relative inequality and its changes. Standard methods include comparison of Lorenz curves and the change in inequality indices like the Gini, the Theil, and the variance of log-incomes, among others. A rise in inequality as gauged by these measures mean that the gaps between the *anonymous* persons in different parts of the income distribution have increased.

To gauge convergence or divergence in incomes the more traditional approach is to estimate a linear model like

$$\Delta y = \gamma_y + \delta_y y_0 + u_y, \qquad (1)$$

where y is a measure of income which can be dollars, log-dollars, shares of mean (or of total income), etc., Δy is the change in that income variable, and y_0 is the initial value of y. If δ_y is positive, then incomes will be said to be *divergent* and the income gap between the *initially* rich and the *initially* poor will grow. If δ_y is negative, the changes will be said to be *convergent* and the gap will diminish. Equivalently, much of the literature estimates

$$y_1 = \alpha_y + \beta_y y_0 + u_y,$$

in which case income changes are said to be divergent or convergent as $\beta_y \gtrless 1$.[1]

[1] These two equations are equivalent in that one can recover γ_y and δ_y from α_y and β_y and vice versa. However, each regression leads to different coefficients of determination.

The main question then is whether it is possible for all four combinations – i) rising inequality and divergent mobility, ii) rising inequality and convergent mobility, iii) falling inequality and divergent mobility, and iv) falling inequality and divergent mobility – to arise. These four possibilities are shown in Table 1. In this section, we present a non-technical summary of the theoretical findings in our work (Duval-Hernandez, Fields and Jakubson 2014), where we show that it is possible to have all of them. Furthermore, not only do we show that these possibilities can all be reconciled, but we explain what underlying conditions need to occur for the reconciliation to take place.

Out of the four cells illustrated in Table 1, most practitioners tend to accept the validity of cells (1,1) and (2,2). That is, people tend to associate rising inequality with panel divergence in incomes, and falling inequality with panel convergence in incomes. When someone talks about "the poor getting poorer, and the rich getting richer" they usually don't qualify whether they are referring to the *initially* poor or to the *anonymous* poor, presumably because they tend to believe both are the same people.

In the next two subsections we outline how can cells (1,2) and (2,1) can be obtained. Namely, how can rising inequality can be reconciled with convergent income changes, and how can divergent income changes can be reconciled with falling inequality.

A. Reconciling Rising Inequality and Convergent Income Changes

Having rising inequality means that the incomes of the *anonymous* rich are moving farther away from the incomes of the *anonymous* poor. Having convergent income changes (as gauged by regressions like (1)) means that the *initially* poor are experiencing larger income changes than the *initially* rich, and hence their incomes are closer to one another after a certain amount of time.

The only possible way for these two circumstances to occur simultaneously is if the *anonymous* rich are not the same people as the *initially* rich, and likewise for the anonymous poor and the initially poor. To illustrate with a simple example how can this occur consider the simple 5-person income vector in the initial period

$$y_0 = [20, 41, 45, 49, 70]$$

which becomes after some time

$$y_1 = [100, 41, 45, 49, 10].$$

In this example, inequality rose, judging by the Lorenz-dominance criterion, as can be seen in the left panel of Figure 1. Yet, the coefficient δ_d of regression (1), when expressed in dollars (d),

$$\Delta d = \gamma_d + \delta_d d_0 + u_d,$$

is negative ($\delta_d = -2.73$), indicating convergence in incomes. The negative slope is apparent from the vectors themselves, since in this case the poorest and richest individuals swapped positions, at the same time when the income gap between the anonymous poor and rich grew. The scatterplot and prediction line of this regression are displayed in the right panel of Figure 1.

In our paper Duval-Hernandez, Fields and Jakubson 2014 we reconcile rising inequality as judged by the Lorenz-criterion or by a Lorenz-consistent index, with convergence in regressions like (1), for incomes appearing in dollars, as shares of mean income, in log-dollars (to approximate proportional income changes), or in a regression with exact proportional changes

$$\frac{d_1 - d_0}{d_0} = \phi + \theta d_0 + u_{pch}.$$

These reconciliations are made for economies with an arbitrary number of individuals, both in periods of economic growth and recession.

In all cases, the key ingredient for the reconciliation of rising inequality with convergent mobility is to have earnings changes large enough, so that some individuals change positions as they go from one period to the next.

It is instructive to illustrate one such reconciliation with our 5-person example. In particular, if we denote by r_l the correlation coefficient between initial and final dollars, i.e.,

$$r_l = \frac{cov(d_0, d_1)}{\sqrt{V(d_1)}\sqrt{V(d_0)}},$$

if we let $CV(d_t)$ be the Coefficient of Variation of incomes in dollars in period t and g be the economy-wide income growth rate, then we can show that dollar changes will be convergent (i.e., $\delta_d < 0$) if and only if

$$r_l \frac{CV(d_1)}{CV(d_0)}(1+g) < 1. \qquad (2)$$

In other words, Equation (2) shows that dollar changes can be convergent, even when inequality is rising (i.e,. if $CV(d_1) > CV(d_0)$) if the correlation coefficient r_l is small enough.[2]

In our previous 5-person example, we find: $\frac{CV(d_1)}{CV(d_0)} = 1.66$, indicating rising inequality; g=0.08, indicating income growth; and $r_l = -0.96$. Since the product of these terms is smaller than one (in fact, is negative), then there is convergence in dollar changes. In this case, the convergence arises because of the strong *negative* correlation between initial and final incomes.

In section III we illustrate this reconciliation with an empirical exploration of earnings data for Mexican labor markets. In particular, we illustrate in more detail the nature of these large changes. For instance, we explore whether in order to obtain the aforementioned reconciliation, it is enough to have few individuals experiencing large earnings changes, or whether we need a large number of crossings among panel people.

B. Reconciling Divergent Income Changes and Falling Inequality

Another point that often confuses practitioners is whether it is possible to have divergent income changes at the same time that inequality falls.

From an intuitive point of view, it seems contradictory to have the incomes of the initially rich and the initially poor drifting apart, while inequality falls concurrently. Furthermore, the literature offers conditions when such reconciliation is literally impossible. For instance, Furceri

[2] Normally, in empirical applications, r_l would be positive. If it is not too positive, the expression in (2) could be less than one. Of course, if r_l is negative, the expression in (2) would surely be less than one.

(2005) and Wodon and Yitzhaki (2006) show that it is impossible to have divergent log-income changes as gauged by a regression of log-income change on initial log-income

$$\Delta \ln y = \gamma_{log} + \delta_{log} \ln y_0 + u_{log},$$

together with a fall in the variance of log-incomes.

In our companion paper (Duval-Hernandez, Fields and Jakubson 2014), we show that for *specific* types of divergence and specific measures of inequality, it is indeed impossible to reconcile divergent mobility with falling inequality. For instance, in addition to the impossibility result by Furceri and Wodon/Yitzhaki, it is impossible to have share-divergence and a fall in inequality as judged by the Lorenz criterion (i.e. a Lorenz-improvement). Also, it is impossible to have divergent income changes (for income measured in dollars) with Lorenz-improvements in times of economic decline.

However, it is perfectly possible to have divergence in dollars and Lorenz-improvements in times of economic growth, as the example [5, 20]→[7,23] shows. It is also possible to have divergent log-incomes and Lorenz-improvements, as can be witnessed by the example [1,1,1,1,1,1,1,1,6,9] → [1,1,1,1,1,1,1,1,7,8]. In fact, it is possible to have many different types of divergence with falling inequality, as long as we allow for crossings in Lorenz-curves and we judge inequality by using some specific Lorenz-consistent measure of inequality.

In summary, the impossibility of having divergent incomes and falling inequality only arises when restricting ourselves to specific income change regressions paired with specific inequality measures.

III. Empirical Reconciliation for Mexico

In the previous section we explained the mechanisms that need to operate in order to reconcile rising inequality with convergent mobility. In this section we illustrate how the aforementioned reconciliation occurs in a real life example analyzing the evolution of inequality and mobility of labor-market earnings in urban Mexico from 1987 to 2013. Over this period the Mexican economy experienced moderate growth and several episodes of recession.

A. Data

The data used is the Encuesta Nacional de Empleo Urbano, ENEU, and its successor, the Encuesta Nacional de Ocupación y Empleo (ENOE).These labor market surveys are rotating panels following the same individuals in the Mexican workforce for 5 quarters. They are suited to provide answers both cross-sectionally as well as dynamically. While the time coverage of any given panel is short, by having many of these short-lived panels we are able to track the evolution of our indicators across different macroeconomic environments.

Over the years the geographical coverage of the survey has changed, including first a few urban centers, and later covering rural areas. We limit our sample to the urban areas that consistently appear in all the surveys. Furthermore, we limit our sample to labor force participants (either employed or unemployed) aged between 18 and 65 years of age at the end of the panel.

Our variable of interest will be monthly earnings measured in 2010 Mexican Pesos. We assign an earnings level of 0 to unemployed individuals, except in the case when dealing with log-earnings. In that case, we assign 1 Mx Peso to the unemployed individuals so that their log-earnings become 0. This imputation is innocuous to the extent that the open unemployment levels are rather low in urban Mexico.[3] All the analysis is performed using the survey sampling weights of the last period when earnings are measured. The basic descriptive statistics of the sample used are presented in Table A-1 in the Appendix.

B. Inequality Changes and Convergent Earnings Reconciled

In the left panel of Figure 2 we present the evolution of earnings inequality over the period 1987-2013. There we observe that inequality rose during the years of economic liberalization from 1987 to 1994. At the end of that year a sharp economic downturn took place as a consequence of the infamous "Tequila crisis". This crisis triggered a reduction in inequality that lasted until the beginning of the new century, after which inequality either leveled or started rising, depending on which measure is used to gauge it.

---

[3] Further evidence that this imputation doesn't alter the conclusions in mobility analyses similar to the one presented next can be found in the online appendix to Fields et. al. (2014).

In contrast, in Figure 3 we present the δ_y coefficients from regression (1), for yearly changes in earnings, i.e. from one initial quarter to the same quarter one year after. In this figure it is apparent that in spite of the ups and downs in inequality displayed in Figure 2, earnings changes either in pesos or in log-pesos are always convergent, and nearly always significantly so. As indicated in the previous section, this means that there must be enough individuals experiencing large enough earnings changes, leading to substantial losses for some initially high income workers, as well as substantial gains for some initially low-earners. All this occurring even as the gap between the highest earnings and the lowest earnings is widening.

These crossings mentioned in the previous paragraph are illustrated in Figure 4. The graphs included in this figure display the initial and final period log-earnings of 27 illustrative individuals (chosen as described below) in the panel from the 3rd quarter of 1987 to the corresponding quarter one year later in 1988. This panel was selected based on the fact that it had one of the largest increases in relative inequality.

To select the 27 individuals in Figure 4, we split the population according to the quintiles of the initial period earnings distribution, and then for each quintile we randomly select an individual located at the 5th, 25th, 50th, 75th, and 95th percentile of a given quintile.[4] We also select two individuals non-randomly, namely, the one individual with the highest initial earnings and the one with the highest final earnings. We plot the location of initial period log-earnings (top line) and final period log-earnings (bottom line), looking at the distributions anonymously (left column) and tracking individuals over time (right panel).[5]

It is clear from these pictures that in spite of having a widening earnings distribution, some individuals experience large earnings changes, both in the positive and in the negative direction, leading to the aforementioned crossings.

---

[4] In other words, we have randomly selected individuals located at the following percentiles of the initial earnings distribution: 1, 5, 10, 15, 19, 21, 25, 30, 35, 39, 41, 45, 50, 55, 59, 61, 65, 70, 75, 79, 81, 85, 90, 95, and 99 percentile.

[5] If the individuals in the 1st percentile of the distribution had earnings equal to zero we added 1 Peso to their earnings, so their log-earnings would be depicted as 0 in the graph.

While illustrative, the previous figure has the disadvantage of being based on the income trajectories of a few selected individuals. To reach a similar conclusion using data from all the sample workers in the panel q3-87 to q3-88, we present in a transition matrix between fixed income categories. This matrix shows that while most individuals have small income changes over the course of a year, there are a few of them who experience large changes that bring the initially rich closer to the initially poor. To wit, while most workers earn between 3 and 4 thousand pesos a month, 10 percent of the labor force experience earnings *changes* larger than 3 thousand pesos.

This small fraction of large convergent changes translates into a low coefficient of determination between initial and final earnings. In fact, applying the reconciliation formula (2) we have that in this period the Coefficient of Variation rose ($\frac{CV(d_1)}{CV(d_0)} = 1.36$), and earnings grew by almost 13%. However, the correlation between initial and final earnings is only 0.54 due to the aforementioned large changes.

One last point to emphasize is that, even while movements in and out of unemployment play a role in explaining the large convergent earnings changes observed in the data, they are by no means the only source of churning in the labor market. This can be better appreciated by looking at Figure 5, which displays the density of final log-earnings and of log-earnings changes for *employed* workers classified according to their *initial* earnings quartile.

Several interesting facts are seen in this figure. First, the distribution of *final*-period log-earnings shifts to the right as we move from poorer to richer *initial*-earnings quartiles, indicating that initially richer individuals tend on average to stay richer one year later. Second, the distribution of log-earnings changes shifts to the *left* as we move from poorer to richer initial-earnings quartiles, illustrating convergence between initial high and low earners. Third, there is a fair degree of overlap between the distributions of *final* log-earnings of individuals who initially belonged to different quartiles. These overlaps are an indication of the moderate to large changes among some members of the employed population.

This section has presented several findings. First, the fact that inequality rises does not necessarily means that the *initially* rich are becoming richer than the *initially* poor. In fact, the data shows the opposite, namely the convergent earnings changes denote that the initial low-earners experience larger gains both in Pesos and proportionally than the high earners. Second, in spite of there being convergence in all periods, this convergence is not strong enough to make the bulk of the initial high-earners poorer than the initial low-earners one year later. Instead, while the majority of the population experience moderate earnings changes, there is a small fraction of the population that has large convergent earnings changes.

Research presented in Fields et. al. (2014) indicates that to an important extent these changes are of transitory in nature. If so, then it remains to assess how these earnings change influence a more permanent measure of inequality, as for instance one based on the average of earnings across periods. This analysis is presented next.

IV. Inequality of Average Earnings and Equalizing Mobility

The previous sections illustrated, that the evolution of inequality among anonymous individuals does not capture the effects of earnings mobility. In this section we illustrate one way of incorporating mobility notions into the analysis of inequality.

In particular, we analyze the inequality of average earnings y_a, in this case defined as the average earnings of an individual over the five quarters for which we observe him/her in the Mexican panels.

Unlike earnings measured at a single point in time, average earnings over several periods capture the effects of economic mobility because they incorporate the ups and downs in earnings over several periods. Hence by focusing on the inequality of these average earnings, we can approximate the inequality of a more permanent measure of earnings.

Furthermore, we can analyze whether economic mobility equalizes or disequalizes these average earnings, in comparison to the earnings that would occur in a world without such mobility. In particular, for an income inequality measure $I(\cdot)$, we can measure the inequality in average earnings $I(y_a)$ and compare it to the inequality that would have prevailed had changes in income shares not taken place, i.e. to $I(y_0)$. This measure EqM (for equalization brought about by mobility)

$$EqM = I(y_0) - I(y_a) \tag{3}$$

would take positive values if earnings changes *equalized* average earnings relative to initial, and it would take negative values if it *disequalized* them.[6]

The right panel of Figure 2 plots the evolution of the inequality of average earnings over five quarters, as gauged by the Gini index and by the variance of log-earnings. A quick comparison of this plot with the one in the left panel of the same figure reveals that for the most part inequality of average earnings follows the same trend as the single-period inequality. However, the levels of inequality of earnings accumulated over five quarters are smaller than the single period ones.

This can also be appreciated in Figure 6, where we display the equalizing mobility gap (3). This figure shows that average earnings are more equally distributed than single-period ones (judging by the positive sign of the EqM measure). Also, the degree of equalization is more or less stable across years, with the exception of the period going from early 2000s to date, when there seems to be a greater degree of equalization brought by the mobility in earnings.

A. Accounting for Levels of Inequality

One interesting analysis is to explore what observable factors account for the levels of single-period and average earnings inequality, when this is measured by the variance of log-earnings. A simple way to do this is to apply the methodology developed by Fields (2003).

In particular, consider a regression of the logarithm of earnings ln y on a vector of observable characteristics W,

$$\ln y = W\gamma + u. \tag{4}$$

[6] This measure is just an algebraic transformation of Fields' (2010) index of mobility as an equalizer of longer-term incomes relative to initial.

Fields shows that the contribution of a regressor w_k to the variance of logarithms equals

$$\gamma_k cov(w_k, \ln y), \tag{5}$$

which can be expressed in absolute levels or as a share of the overall variance of log-earnings $V(\ln y)$.

Table 3 shows the result of applying this decomposition to the Mexican data. In particular, we pooled data from several panels into recessionary and non-recessionary periods. The recessionary periods include the year following the 1994-95 "Tequila Crisis", the early 2000s recession, and the Great Recession of 2008-09. The list of regressors included in equation (4) are a gender dummy, a 4th order age polynomial, a 2nd order polynomial in the years of schooling, an unemployment dummy, industry and occupation dummies, as well as dummies for whether the individual is an employee in the formal sector, an employee in the informal sector, or self-employed. In addition to that city, and period dummies were included as well. [7]

This table shows that both in recession and in non-recession years, being unemployed is by far the greatest contributor among observables to inequality of both initial and average log-earnings. The second most important observable factor contributing to inequality is the sector of employment of the worker (formal/informal/self-employed). After that, occupation, years of schooling, gender and age each contribute between 1 and 2% to the level of variance of log-earnings. Finally, around 40% of the variance of log-earnings remains unexplained by the observable characteristics.

Among the differences that we can observe between the correlates of inequality of initial and average log-earnings, are that the contribution of unemployment to inequality falls when looking at average log-earnings, while the relevance of the sector of employment rises. Similarly, occupation, education, and gender play a larger (though still relatively small) role in accounting for the inequality of average log-earnings. These results do not vary much depending on the macroeconomic environment.

---

[7] The underlying regressions that were used to generate this decomposition are available from the authors upon request.

B. Accounting for Equalizing Mobility

So far the previous decomposition accounted for the levels of both single-period and average log-earnings. However, we can also use this methodology to explore what factors account for our equalization measure EqM in equation (3).

In performing the accounting of the gap in (3) it is useful to distinguish between the contribution brought by *changes in observable characteristics* and the *changes in the coefficients* of these characteristics, much in the spirit of Oaxaca (1973) and Juhn-Murphy-Pierce (1993) decompositions.

In particular, we can construct a counterfactual predicted log-earnings, $\ln y_c$, using the observed average characteristics of the worker W_a and the coefficients estimated in the initial period 0, γ_0, i.e.,

$$\ln y_c = W_a \gamma_0. \tag{6}$$

Denote by σ^2_{w0} and σ^2_{wa} the portion of the variance of initial and average log-earnings, respectively, accounted for by observable factors. Furthermore, denote by σ^2_c the variance of the counterfactual log-earnings in (6). Finally, denote by σ^2_{r0} and σ^2_{ra} the residual variance of initial and average log-earnings, respectively. Then, we can decompose the gap $EqM = V(\ln y_0) - V(\ln y_a)$ as

$$EqM = (\sigma^2_{w0} - \sigma^2_c) + (\sigma^2_c - \sigma^2_{wa}) + (\sigma^2_{r0} - \sigma^2_{ra}). \tag{7}$$

The first term, $\sigma^2_{w0} - \sigma^2_c$, represents the equalization brought about by changes in the observed characteristics, when the coefficients are kept at their initial level γ_0. The second term, $(\sigma^2_c - \sigma^2_{wa})$, represents the equalization brought about by changes in coefficients, when the observable characteristics are kept at their average levels W_a. Finally, the last term is the contribution to equalization coming from the differences in residuals between both models. One advantage of this method is that we can readily obtain the detailed contribution of individual observable variables to the first two terms in (7).

This method is an application of the method proposed by Yun (2006), which in turn is an extension of the method by Fields (2003). The innovation of our paper is instead the application of such decomposition to analyze the equalization of average earnings relative to initial earnings due to mobility, rather than the changes in inequality between two anonymous distributions. [8] This decomposition is presented for the Mexican data in Table 4. Several interesting findings arise from this exercise.

First, the largest contribution to equalizing earnings over time comes from changes in the employment status of the workers. The fact that transitions in and out of unemployment *equalized* rather than disequalized earnings can be explained by the fact that in Mexico there is a higher incidence of unemployment among educated individuals (see for instance Duval Hernandez and Orraca, 2011), mainly because poor uneducated workers cannot afford being jobless for a long time. This implies that transitions into unemployment will usually involve high-earners losing a substantial amount of money, while transitions out of unemployment will usually involve high-earners moving from zero-earnings to a high income level. In practice, both movements get recorded as equalizing average earnings relative to initial.

Second, changes in the coefficients associated with the sector of employment (formal/informal/self-employed) and with occupation play a moderate disequalizing role, while transitions across occupations and industries counteract one another: namely, changes across occupations equalize average earnings relative to initial by about 1.3% of the total equalization, while changes across industries have a disequalizing effect of similar magnitude. Finally, other observable factors play a minor role in accounting for the equalization of average relative to initial earnings.[9]

[8] The full derivation of this decomposition is included in the Appendix of the paper.
[9] The fact that time-invariant factors like gender, years of schooling, and age (which is invariant in these one year panels) contribute to the impact of *changing* characteristics on the inequality gap, is due to the correlation that these factors have with time-varying characteristics. For more details, refer to the Appendix.

V. Conclusions

This paper showed how our view of who benefits and who is hurt when the economic environment changes is different if we look at the changes in income inequality among anonymous individuals, or if instead we track the individuals' incomes by means of panel data.

In section II of the paper we discussed how rising and falling inequality can be reconciled with convergent and divergent income changes. Our theoretical discussion of possibilities was empirically illustrated using a panel dataset that tracks the earnings of workers for one year in urban Mexico.

In the empirical analysis we observed that while earnings inequality sometime rises and sometimes falls, earnings changes in Mexico are never divergent. The reason for the convergence between initial high-earners and initial low-earners is that over the course of a year a small fraction of the initially rich experiences large losses, while another small fraction of the initially poor experiences large gains. On average, though, most people tend to experience small changes in earnings.

Since any single-period measure of inequality will capture a transitory component of earnings as well as a more permanent component, it then becomes relevant to: i) calculate the inequality of a more permanent measure of income than the one obtained from single-period earnings, , and ii) explore what factors account for the equalization/disequalization that occurs over time as a result of the changes in earnings. These two aspects were studied in section IV of the paper.

In that section, we showed that the average monthly earnings (over 5 quarters) are more equally distributed than monthly earnings in any single quarter. Also, both for single-period earnings, and for average earnings, the employment status and the sector of employment (formal/informal/self-employed) of the worker are the most important observable factors that account for the levels of inequality.

Regarding what factors account for the equalization of average earnings relative to initial earnings, we found that changes in the employment status of workers are by far the single most important equalizing factor.

The methods applied in the empirical part of the paper could be used to analyze the equalization/disequalization brought about by earnings changes, in different economies and in different economic contexts. In particular, it would be interesting to apply them to income data covering longer time horizons, and to explore whether with this data sociodemographic characteristics like schooling and gender play a larger role in equalizing/disequalizing average earnings relative to initial. Also, it would be interesting to assess the role that labor market policies play in such equalization/disequalization.

References

Duval-Hernández, Robert, Gary S. Fields, and George H. Jakubson (2014) “Changing Income Inequality and Panel Income Changes”, unpublished manuscript.

Duval Hernández, Robert, and Pedro Orraca Romano (2011) “Análisis por cohortes de la participación laboral en México” *El Trimestre Económico*, LXXVIII (2), 310, pp. 343-375.

Fields, Gary S. (2003) “Accounting for Income Inequality and Its Change: A New Method, With Application to the Distribution of Earnings in the United States”, *Research in Labor Economics*, Elsevier, vol. 22, 1-38.

Fields, Gary S. (2010) “Does income mobility equalize longer-term incomes? New measures of an old concept”, *Journal of Economic Inequality*, 8:409–427.

Fields, Gary S., Robert Duval-Hernández, Samuel Freije, and Maria Laura Sánchez-Puerta (2014) “Earnings Mobility, Inequality, and Economic Growth in Argentina, Mexico, and Venezuela”, *Journal of Economic Inequality*, forthcoming.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo (2011) “Decomposition Methods in Economics” in David Card and Orley Ashenfelter (eds.) *Handbook of Labor Economics*, Elsevier, vol. 4, 1-102.

Furceri, D. (2005) “β and σ-convergence: A mathematical relation of causality”, *Economics Letters*, 89: 212–215.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce (1993) “Wage Inequality and the Rise in Returns to Skill,” *Journal of Political Economy*, 101(3), 410–42.

Oaxaca, Ronald (1973) “Male-Female Wage Differentials in Urban Labor Markets”, *International Economic Review*, 14: 693-709.

Wodon, Q. and S. Yitzhaki (2006) “Convergence forward and backward?”, *Economics Letters*, 92: 47–51.

Yun, Myeong-Su (2006). “Earnings Inequality in USA, 1969–99: Comparing Inequality Using Earnings Equations”, *Review of Income and Wealth*, 52(1): 127-144.

Tables and Figures

Table 1. Possibilities for Rising/Falling Inequality and Convergent/Divergent Mobility.

	Falling Inequality	**Rising Inequality**
Convergent Mobility	√	√
Divergent Mobility	√	√

√ : This cell is possible.

Figure 1. 5-person example.

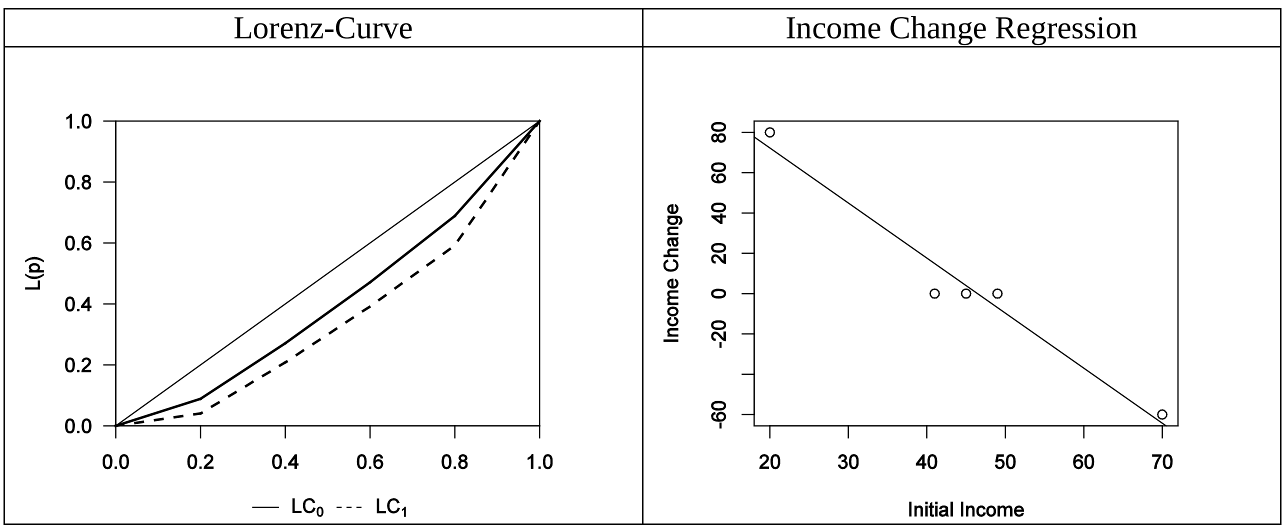
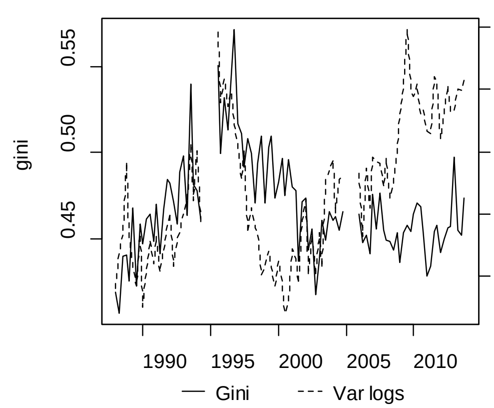


Figure 2. Evolution of Earnings Inequality.

Initial Earnings



Average Earnings

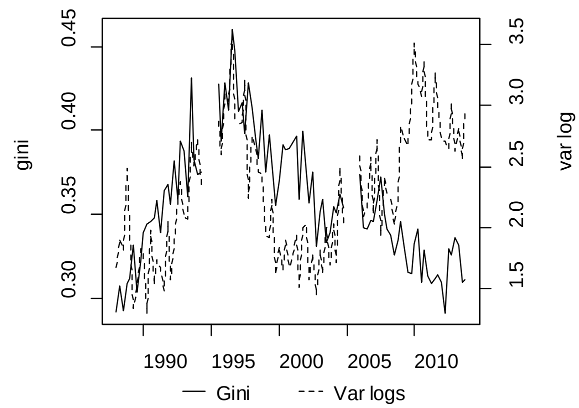
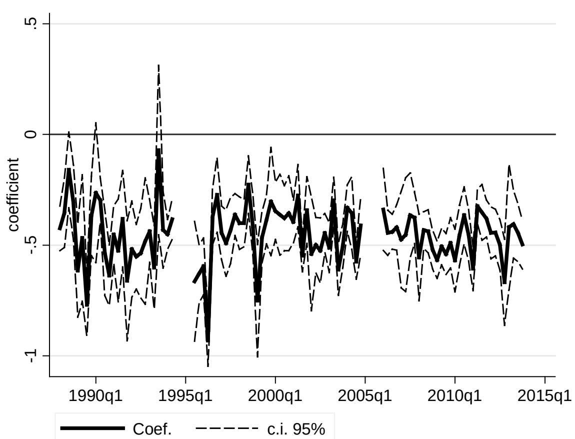


Figure 3. Convergence Coefficients from Linear Regression Model.

Earnings in Levels



Log-Earnings

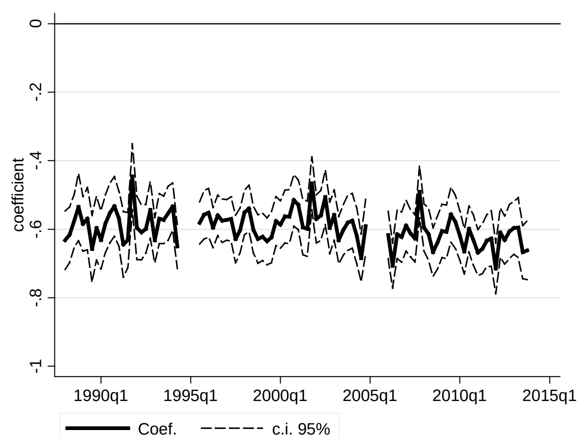


Figure 4. Earnings Distributions for 27 Illustrative Individuals in a Period of Rising Inequality.

Treated Anonymously	Treated as Panel

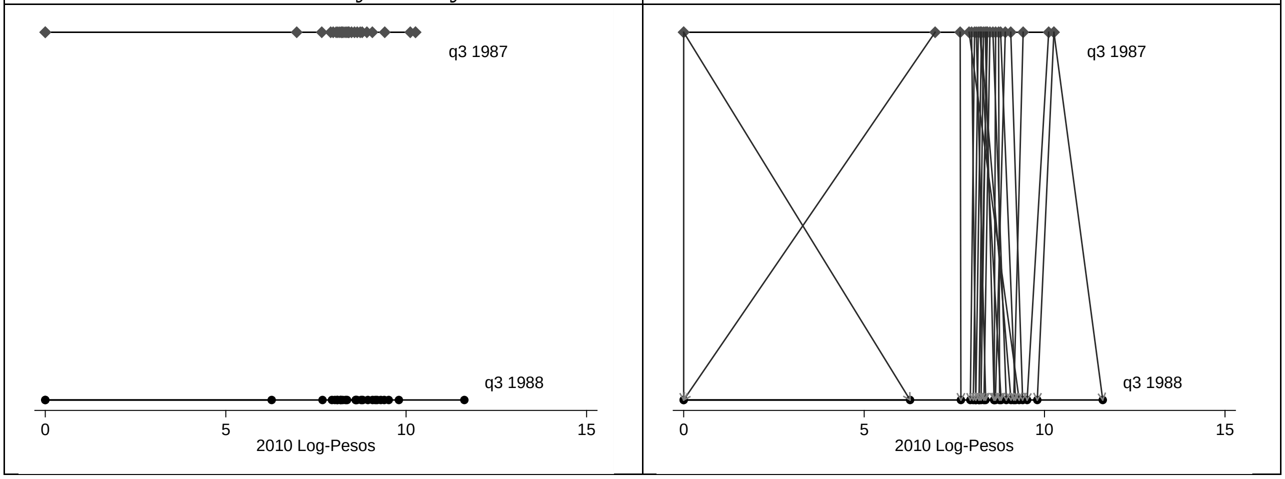
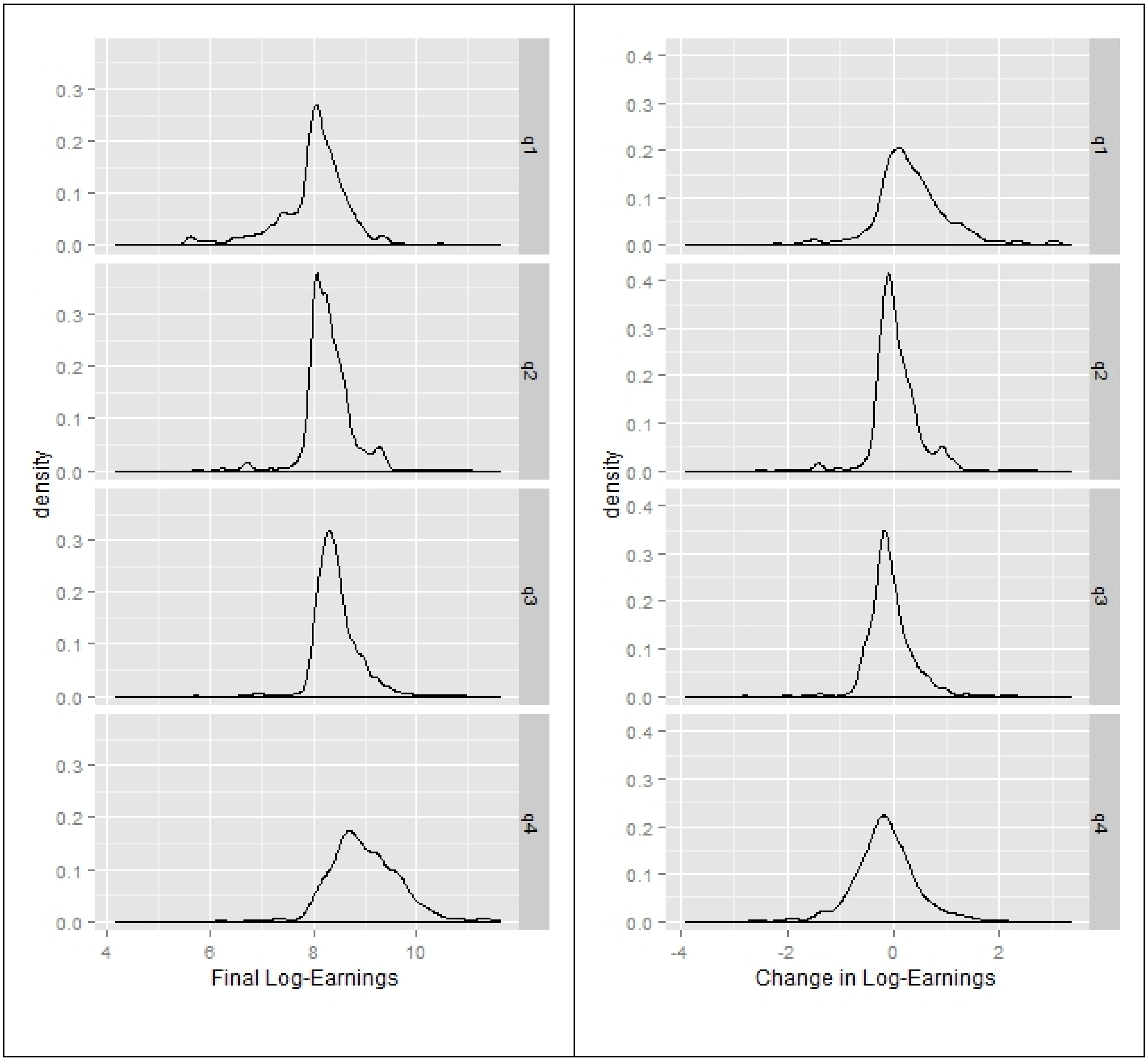


Table 2. Transition Matrix across Fixed Earnings Categories, in Thousands of 2010 Mexican Pesos.

	Final Earnings (000s)									
Initial Earnings (000s)	[0,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,)	Total
[0,1)	**3.7**	0.9	1.1	1.8	0.8	0.7	0.2	0.0	0.3	**9.7**
[1,2)	0.9	**1.3**	0.9	0.8	0.5	0.2	0.0	0.0	0.2	**4.7**
[2,3)	0.6	0.8	**1.5**	2.6	1.1	0.5	0.2	0.3	0.3	**7.9**
[3,4)	1.1	0.4	4.2	**10.5**	4.5	3.0	1.1	0.6	1.3	**26.9**
[4,5)	0.3	0.2	0.7	8.0	**5.1**	1.7	0.9	0.4	1.6	**18.9**
[5,6)	0.2	0.0	0.4	2.4	2.2	**1.4**	0.9	0.8	1.3	**9.6**
[6,7)	0.2	0.1	0.1	1.1	0.7	1.2	**0.8**	0.5	1.2	**5.9**
[7,8)	0.1	0.0	0.1	0.6	0.4	1.0	0.4	**0.6**	1.5	**4.7**
[8,)	0.2	0.1	0.1	0.5	0.6	0.9	1.1	0.9	**7.3**	**11.7**
Total	**7.4**	**3.8**	**9.1**	**28.4**	**15.9**	**10.6**	**5.7**	**4.0**	**15.1**	**100**

The cells are % of the sample population.
The data corresponds is from the panel ENEU q3-1987 to q3-1988.

Figure 5. Densities of Final Log-Earnings and Log-Earnings Changes by Quartile of the Initial Earnings. Employed Workers Only.



Based on data from panel q3-1987 to q3-1988.

Figure 6. Equalizing Mobility Gap.

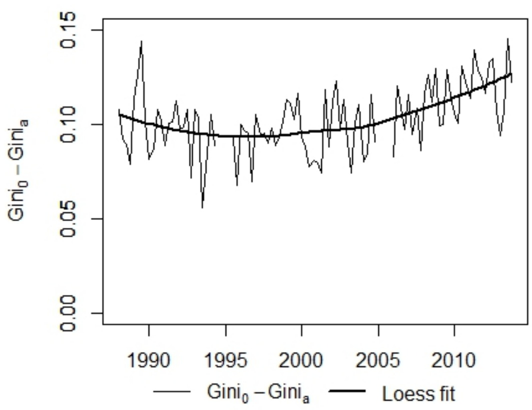
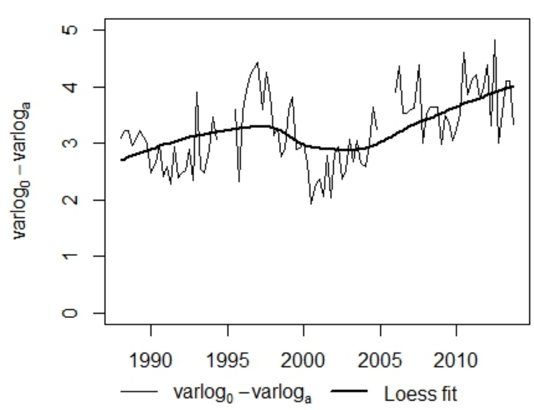
Gini	Variance of Log-Earnings
	

Table 3. Accounting for Levels of Single-Period and Average Log-Earnings Inequality.

	Non-recession		Recession	
	Initial Earnings	**Average Earnings**	**Initial Earnings**	**Average Earnings**
V(ln y)	4.51	2.35	3.90	2.17
	(100)	(100)	(100)	(100)
Gender	0.048	0.043	0.053	0.050
	(1.1)	(1.8)	(1.4)	(2.3)
Age	0.049	0.030	0.046	0.029
	(1.1)	(1.3)	(1.2)	(1.3)
Education	0.053	0.047	0.080	0.065
	(1.2)	(2.0)	(2.1)	(3.0)
Unemployment	2.350	0.845	1.875	0.692
	(52.1)	(35.9)	(48.0)	(31.8)
Informality	0.144	0.200	0.137	0.190
	(3.2)	(8.5)	(3.5)	(8.8)
Occupation	0.088	0.082	0.086	0.089
	(2.0)	(3.5)	(2.2)	(4.1)
Industry	0.002	0.033	-0.001	0.027
	(0.03)	(1.4)	-(0.03)	(1.2)
City dummies	0.015	0.013	0.018	0.016
	(0.3)	(0.5)	(0.5)	(0.7)
Period dummies	0.013	0.010	-0.002	-0.001
	(0.3)	(0.4)	-(0.05)	-(0.03)
Residuals	1.750	1.053	1.612	1.017
	(38.8)	(44.7)	(41.3)	(46.8)

Percentage Shares of V(ln y) are reported in parentheses.
All earnings measures are in natural logarithms.
Recession periods include: q3-94 to q4-95, q4-00 to q1-02, and q3-08 to q2-09.

Table 4. Accounting for Earnings Equalizing Mobility.

	Non-recession		Recession	
V(ln y_0) - V(ln y_a)	2.16	(100)	1.73	(100)
	Chars	**Coeff**	**Chars**	**Coeff**
Gender	0.002	0.003	0.005	-0.001
	(0.1)	(0.2)	(0.3)	-(0.1)
Age	0.005	0.014	0.004	0.014
	(0.3)	(0.7)	(0.2)	(0.8)
Education	-0.001	0.008	-0.001	0.016
	-(0.1)	(0.4)	-(0.1)	(0.9)
Unemployment	1.503	0.002	1.164	0.020
	(69.7)	(0.1)	(67.2)	(1.1)
Informality	-0.005	-0.051	0.008	-0.061
	-(0.2)	-(2.4)	(0.4)	-(3.5)
Occupation	0.024	-0.018	0.024	-0.027
	(1.1)	-(0.8)	(1.4)	-(1.5)
Industry	-0.027	-0.005	-0.019	-0.009
	-(1.3)	-(0.2)	-(1.1)	-(0.5)
City dummies	1.88E-05	0.003	-2.91E-05	0.002
	(0.001)	(0.1)	-(0.002)	(0.1)
Period dummies	0.002	0.001	0.002	-0.003
	(0.1)	(0.03)	(0.1)	-(0.2)
Residuals	0.696		0.596	
	(32.3)		(34.4)	

ln y_0 denotes initial log-earnings, ln y_a denotes average log earnings.

Percentage Shares of V(ln y_0) - V(ln y_a) are reported in parentheses.
Char and Coeff are, respectively, characteristics and coefficients effects.
Recession periods include: q3-94 to q4-95, q4-00 to q1-02, and q3-08 to q2-09.

Appendix

Table A-1. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Initial Earnings	295,439	5,853.88	6,493.27	0	355,878
Earnings Change	295,439	22.75	7,966.65	-350,188	2,522,127
Male	295,439	0.69	0.46	0	1
Age	295,439	35.82	11.29	18	65
Yrs. Schooling	295,439	9.33	4.37	0	25
Unemployed	295,439	0.03	0.16	0	1
Sector					
Formal Employee	295,439	0.56	0.50	0	1
Informal Employee	295,439	0.23	0.42	0	1
Self-Employed (omitted)					
Occupation					
Professional/Manager	295,439	0.15	0.36	0	1
Supervisor/Operator	295,439	0.13	0.34	0	1
Production/Craft	295,439	0.24	0.43	0	1
Transport/Mechanic	295,439	0.06	0.24	0	1
Clerical/Sales	295,439	0.25	0.43	0	1
Services	295,439	0.12	0.32	0	1
Security (omitted)					
Industry					
Agricultural	295,439	0.01	0.09	0	1
Extractive/Electricity	295,439	0.01	0.10	0	1
Construction	295,439	0.07	0.25	0	1
Manufacturing	295,439	0.25	0.44	0	1
Trade	295,439	0.18	0.39	0	1
Services	295,439	0.40	0.49	0	1
Public Admin (omitted)					

Method to Account for Income Equalizing Mobility

This section provides the details on how we obtain the contribution of observable and unobservable factors to the gap between the relative inequality of initial versus average earnings. In particular, we will focus on the variance of log-incomes as our measure of relative inequality. Hence, the goal is to account for the gap

$$V(\ln y_0) - V(\ln y_a),$$

where y_a and y_0 are average and initial earnings respectively.

The method here presented is similar to that of Yun (2006) which is an extension of the one presented in Fields (2003).

Consider a logarithmic regression for initial earnings,

$$\ln y_0 = Z\alpha_0 + X_0\beta_0 + u_0, \qquad \text{(A-1)}$$

where Z and X_0 are vectors of time-invariant and time-variant observable characteristics, respectively, with coefficients α_0 and β_0, and u_0 is the error term. Equation (4) in the text provides a compact way of expressing this equation as

$$\ln y_0 = W_0\gamma_0 + u_0 \qquad \text{(A-2)}$$

for $W_0 = [Z, X_0]$, $\gamma_0 = [\alpha_0, \beta_0]$, and u_0.

Furthermore, assume the error term is uncorrelated with the regressors, i.e.,

$$cov(u_0, w_k) = 0 \quad \forall k.$$

This assumption means that the coefficients γ_0 are not to be interpreted as the structural impacts of the independent variables on the conditional expectation of log-earnings, but merely as the coefficients of the linear projection of the dependent variable on the observable characteristics.

We can define a similar model for the log of average earnings as

$$\begin{aligned} \ln y_a = Z\alpha_a + X_a\beta_a + u_a \Leftrightarrow \\ \ln y_a = W_a\gamma_a + u_a, \end{aligned} \qquad \text{(A-3)}$$

where X_a denotes the average time-varying observable characteristics. We again maintain the non-correlation between errors and regressors.

As previously mentioned in the text, Fields (2003) shows that the contribution of the variance of log-earnings attributable to each observable factor w_k can be estimated as

$$\gamma_{k0}cov(w_{k0}, \ln y_0), \qquad \text{(A-4)}$$

for the initial period equation, and as

$$\gamma_{ka}cov(w_{ka}, \ln y_a). \qquad \text{(A-5)}$$

for the equation of log average earnings. In addition to these contributions attributable to observable factors, there is a contribution of the residuals to the variance of logs. These contributions from the residuals will be denoted by σ^2_{r0} and σ^2_{ra}, for the initial and the average earnings equations, respectively.

In summary, if we define the contribution of all observable factors to the log variance as

$$\sigma^2_{ws} = \Sigma^K_{k=1}\gamma_{ks}cov(w_{ks}, \ln y_s) \quad for\ s \in \{0, a\},$$

we can then express the log-variance of earnings as

$$V(\ln y_s) = \sigma_{ws}^2 + \sigma_{rs}^2 \quad for\ s \in \{0, a\}. \tag{A-6}$$

Since the variance σ_{ws}^2 can be decomposed as a sum of individual terms, one for each regressor, this forms a "detailed decomposition", following the terminology adopted in the literature (see for instance Fortin, Lemieux, and Firpo, 2011).

Now, define the counterfactual log-earnings based on observables, $\ln y_c$, as the log-earnings that would arise if we predict using the observed average characteristics of the worker and the coefficients estimated in the initial period 0. More precisely, let

$$\ln y_c = W_a \gamma_0. \tag{A-7}$$

Using these counterfactual earnings based on observables helps us to further decompose the contribution of each factor into changes in the observable characteristics (evaluated at fixed coefficients), and changes in coefficients (holding constant observable characteristics), in the same spirit of a Oaxaca decomposition.

More specifically denote the counterfactual log-variance σ_c^2 based on observables as

$$\sigma_c^2 = \Sigma_{k=1}^{K} \gamma_{k0} cov(w_{ka}, \ln y_c). \tag{A-8}$$

Using equations (A-6) and (A-8) we can rewrite the total equalization gap as

$$V(\ln y_0) - V(\ln y_a) = (\sigma_{w0}^2 - \sigma_c^2) + (\sigma_c^2 - \sigma_{wa}^2) + (\sigma_{r0}^2 - \sigma_{ra}^2). \tag{A-9}$$

The first term, $\sigma_{w0}^2 - \sigma_c^2$, represents the equalization brought about by changes in the regressors,[10] when the coefficients are kept at their initial level γ_0. The second term, $(\sigma_c^2 - \sigma_{wa}^2)$, represents the equalization brought about by changes in coefficients when the observable characteristics are kept at their average levels W_a. Finally, the last term is the contribution to equalization coming from the residuals in the two models.

Since the method outlined here provides a detailed decomposition for each regressor, we can express these "coefficient effects" and "characteristics effects" for a typical regressor w_l. In particular, the "characteristics effect" for the factor w_l can be expressed as

$$\begin{aligned} &\gamma_{l0} cov(w_{l0}, \ln y_0) - \gamma_{l0} cov(w_{la}, \ln y_c) \\ &= \gamma_{l0} cov(w_{l0}, w_{10}\gamma_{10} + \cdots + w_{K0}\gamma_{K0}) - \gamma_{l0} cov(w_{la}, w_{1a}\gamma_{10} + \cdots + w_{Ka}\gamma_{K0}) \\ &= \gamma_{l0} \Sigma_{k=1}^{K} \gamma_{k0} [cov(w_{l0}, w_{k0}) - cov(w_{la}, w_{ka})], \end{aligned} \tag{A-10}$$

while the "coefficient effect" can be written as

$$\begin{aligned} &\gamma_{l0} cov(w_{la}, \ln y_c) - \gamma_{la} cov(w_{la}, \ln y_a) \\ &= \gamma_{l0} cov(w_{la}, w_{1a}\gamma_{10} + \cdots + w_{Ka}\gamma_{K0}) - \gamma_{la} cov(w_{la}, w_{1a}\gamma_{1a} + \cdots + w_{Ka}\gamma_{Ka}) \\ &= \Sigma_{k=1}^{K} (\gamma_{l0}\gamma_{k0} - \gamma_{la}\gamma_{ka}) cov(w_{la}, w_{ka}) \end{aligned} \tag{A-11}$$

Expression (A-10) serves to illustrate that even time-invariant factors like z_l can have a non-zero contribution to the aforementioned "characteristics effect". This occurs because the covariance differences $cov(z_l, x_{k0}) - cov(z_l, x_{ka})$ are not necessarily equal to zero.

[10] If the term is positive it means that log average earnings were more equally distributed than initial ones.