WIDER Working Paper 2015/026

Analysing income distribution changes

Anonymous versus panel income approaches

Robert Duval-Hernández,[1] Gary S. Fields,[2] and
George H. Jakubson[3]

February 2015

Abstract: We reconcile, both theoretically and empirically, changes in inequality with panel income changes over periods of economic growth and decline. We also explore what factors account for the trends of short-run inequality and of inequality in individual average earnings. Finally, we explore what factors account for the equalization brought about by economic mobility. Using panel earnings data from Mexico we find that earnings changes are convergent, irrespective 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 and sector of workers.

Keywords: Income inequality, economic mobility
JEL classification: J31, D63

Acknowledgements: The authors would like to thank the participants at the UNU-WIDER conference on 'Inequality – Measurements, Trends, Impacts, and Policies', held 5-6 September 2014 in Helsinki, Finland for their valuable comments

---

[1]University of Cyprus, CIDE; [2]Department of Economics, Cornell University, IZA, UNU-WIDER, gsf2@cornell.edu; [3]Department of Economics, Cornell University, gj10@cornell.edu; corresponding author: rduval@ucy.ac.cy

This paper has been presented at the UNU-WIDER 'Inequality—Measurement, Trends, Impacts, and Policies' conference, held 5–6 September 2014 in Helsinki, Finland.

ISSN 1798-7237 ISBN 978-92-9230-911-4 https://doi.org/10.35188/UNU-WIDER/2015/911-4

Typescript prepared by Judy Hartley for UNU-WIDER.

UNU-WIDER gratefully acknowledges the financial contributions to the research programme from the governments of Denmark, Finland, Sweden, and the United Kingdom.

The World Institute for Development Economics Research (WIDER) was established by the United Nations University (UNU) as its first research and training centre and started work in Helsinki, Finland in 1985. The Institute undertakes applied research and policy analysis on structural changes affecting the developing and transitional economies, provides a forum for the advocacy of policies leading to robust, equitable and environmentally sustainable growth, and promotes capacity strengthening and training in the field of economic and social policy-making. Work is carried out by staff researchers and visiting scholars in Helsinki and through networks of collaborating scholars and institutions around the world.

UNU-WIDER, Katajanokanlaituri 6 B, 00160 Helsinki, Finland, wider.unu.edu

1 Motivation

The literature analysing inequality has devoted a lot of 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, this literature has compared anonymous individuals at different periods. 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 rather talk about 'the poorest', 'the richest', and so on.

An alternative approach for analysing 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 and replaces it by what is sometimes called 'two-period anonymity' (or 'T-period anonymity' if the panel is T-periods long). More specifically, panel data can be used to analyse changes in the shape of the income distribution, but they can do more by also displaying the evolution of income for each individual who 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 per cent of the income distribution became poorer might change depending on whether we look at the incomes of the anonymous bottom 10 per cent, or whether we track with panel data the incomes of those who initially were in the bottom 10 per cent. In other words, the standard inequality analysis follows the evolution of incomes of whoever is in the bottom 10 per cent, 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 per cent, but who might or might not have moved to other points in the income distribution.

In this paper, we have two goals. First, we summarize in an accessible manner our recent theoretical findings on how the answers provided by the anonymous and panel methods can be reconciled. We illustrate this reconciliation empirically using one-year panel data from Mexico over several decades, including periods of economic growth and decline and of rising and falling inequality. Second, we examine how our view of inequality is altered if instead of looking at earnings inequality at a point in time, we focus on the inequality of average earnings. Taking the average of earnings over time for each individual gives us a measure of earnings that is less affected by single-period shocks. More specifically, we compare trends in single- and multi-period earnings inequality, and we explore what individual and aggregate observable factors account for their levels and for the equalization brought by economic mobility.

2 Reconciling anonymous and panel income changes

There is a large literature on how to measure relative inequality and its changes. Standard methods include comparisons of Lorenz curves and calculations of changes 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 means 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

$$\Delta y = \gamma_y + \delta_y y_0 + u_y \tag{1}$$

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

$$y_1 = \alpha_y + \beta_y y_0 + u_y \tag{2}$$

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

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 convergent 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-Hernández, Fields, and Jakubson 2014), where we show that it is possible to have all four combinations. 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 displayed in Table 1, most practitioners tend to accept the validity of cells along the diagonal of the matrix—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 do not 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 sub-sections we outline how cells (1,2) and (2,1) can be obtained. Namely, how rising inequality can be reconciled with convergent income changes, and how divergent income changes can be reconciled with falling inequality.

[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, the two regressions lead to different coefficients of determination.

Table 1: Possibilities for rising/falling inequality and convergent/divergent mobility

	Falling inequality	Rising inequality
Convergent mobility	√	√
Divergent mobility	√	√

Note: √: This cell is possible.

Source: Authors' calculation.

2.1 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 on average 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 this can occur, consider the simple five-person income vector in the initial period

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

which becomes after some time

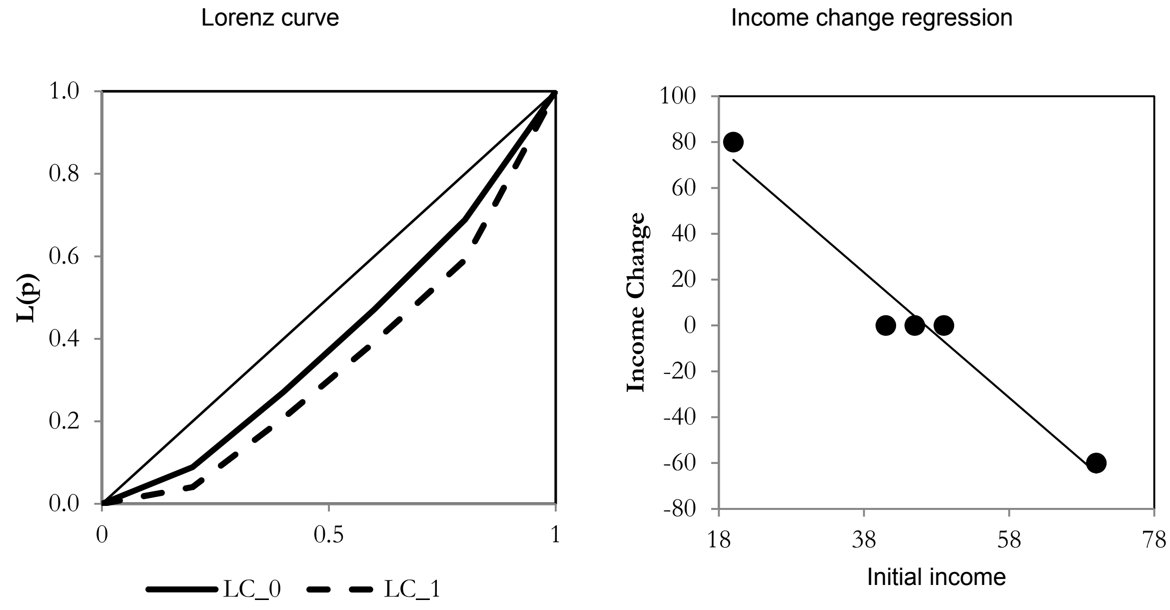
$$y_1 = [100, 41, 45, 49, 10]$$

(Throughout this paper, we follow the convention in the income mobility literature of ordering each vector in ascending order of *initial* incomes.) 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, while at the same time 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.

Figure 1: Five-person example



Source: Authors' illustration.

In our paper Duval-Hernández, 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 equation (1), for incomes measured 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 five-person example. In particular, if we denote by r_l the correlation coefficient between initial and final dollars in regression (2), 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 have shown in Duval-Hernández, Fields, and Jakubson (2014) that dollar changes will be convergent (i.e., $\beta_d < 1$, or equivalently $\delta_d < 0$) if and only if

$$r_l \frac{CV(d_1)}{CV(d_0)}(1+g) < 1. \tag{3}$$

In other words, equation (3) 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 between initial and final incomes r_l is small enough.[2]

In our previous five-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 3 we further illustrate this reconciliation with an empirical exploration of earnings data for Mexican labour markets. In particular, we illustrate in more detail the nature of these large changes.

2.2 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 (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-Hernández, 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 in 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

[2] Normally, in empirical applications, r_l would be positive. If it is positive but not too large, the expression in equation (3) could be less than one. Of course, if r_l is negative, the expression in equation (3) would surely be less than one.

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.

3 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, analysing the evolution of inequality and mobility of labour market earnings in urban Mexico from 1987 to 2013. Over this period the Mexican economy experienced moderate growth and several episodes of recession.

3.1 Data

The data used are the Encuesta Nacional de Empleo Urbano (ENEU) and its successor, the Encuesta Nacional de Ocupación y Empleo (ENOE).These labour market surveys are rotating panels following the same individuals in the Mexican workforce for five 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 centres, then adding more urban areas, 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 labour force participants (either employed or unemployed) aged between 18 and 65 years of age at the end of the panel.

Our variable of interest is 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 Mexican 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 interview quarter.

3.2 Inequality changes and convergent earnings reconciled

In the top 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 levelled off or started rising, depending on which measure is used to gauge it.

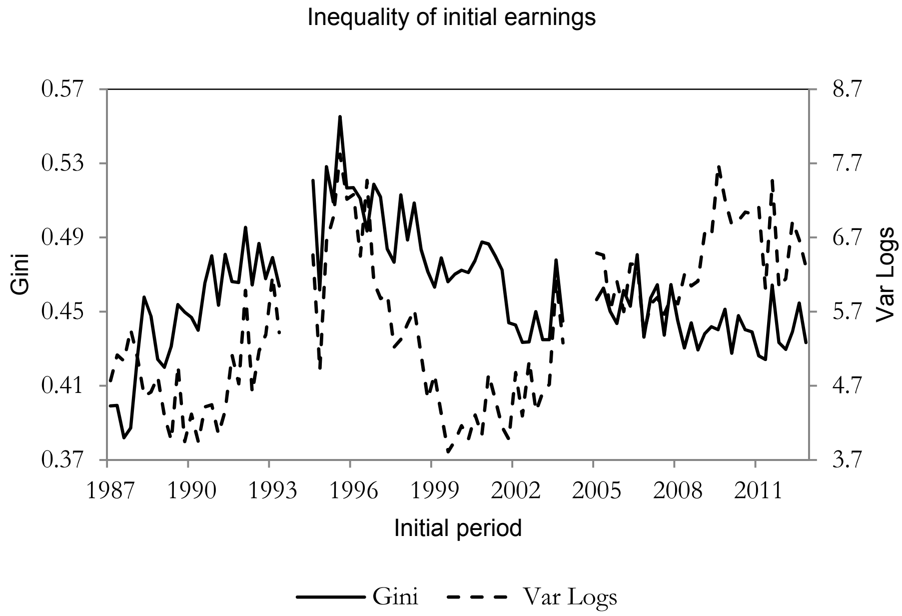
In contrast, in Figure 2 we present the δ_y coefficients from regression (1), for yearly changes in earnings – that is, from one initial quarter to the same quarter one year later – and test for

[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. (forthcoming).

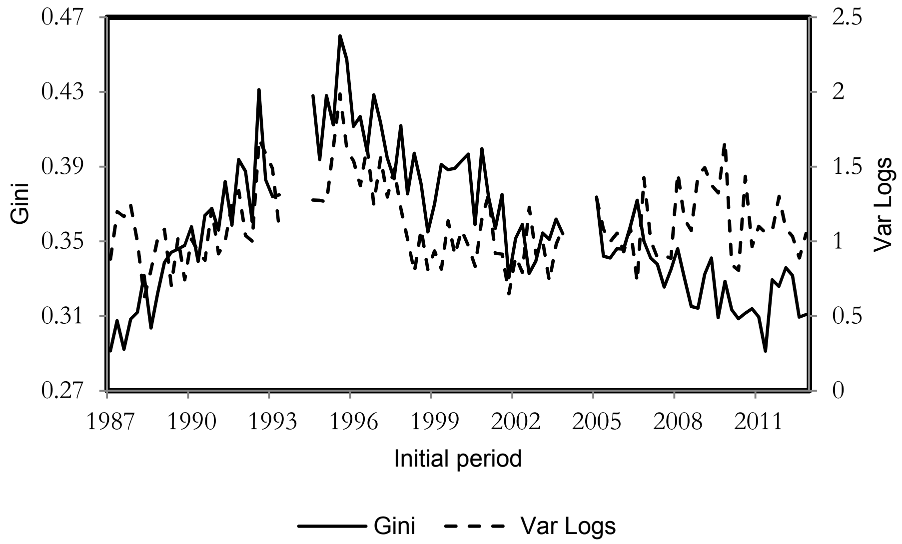
divergence/convergence according to $\delta_y \gtrless 0$. In this figure it is apparent that in spite of the ups and downs in inequality displayed in Figure 2, earnings changes either in pesos or in log-pesos are always convergent, and nearly always significantly so. As indicated in the previous section, this finding of convergence 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 is occurring even as the gap between the highest earnings and the lowest earnings is widening.

The 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.

Figure 2: Evolution of earnings inequality



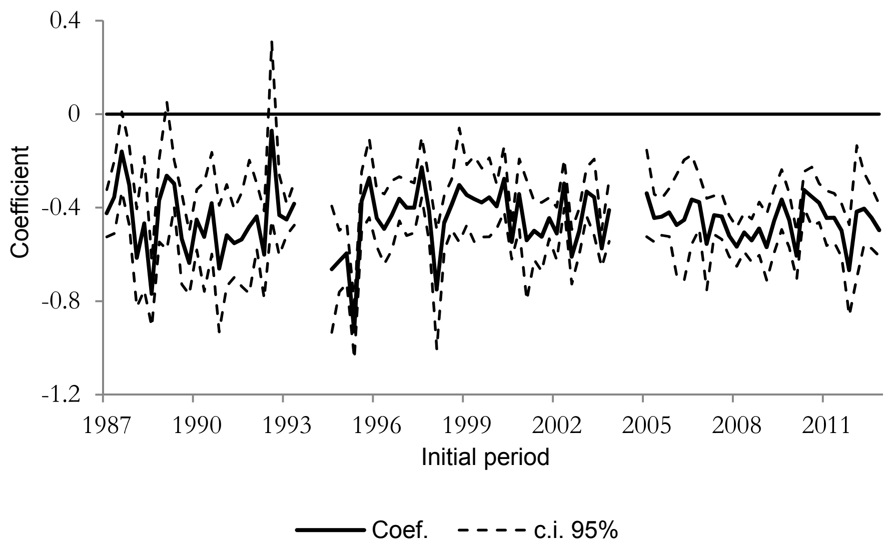
Inequality of average earnings



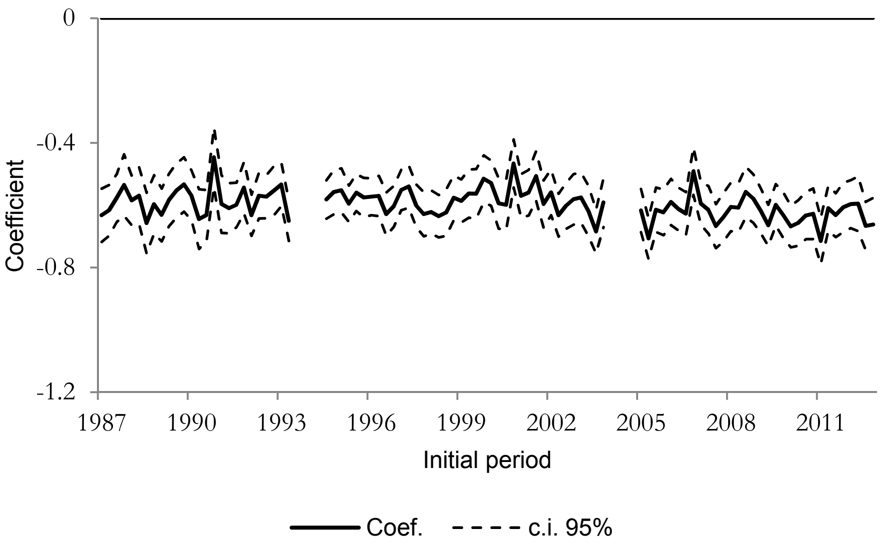
Source: Authors' illustration based on ENEU/ENOE data.

Figure 3: Convergence coefficients from linear regression model

Coefficient from regression with earnings in levels



Coefficient from regression with log-earnings



Source: Authors' illustration based on ENEU/ENOE data.

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 group we randomly select an individual located at the 5th, 25th, 50th, 75th, and 95th per centiles of a given quintile group.[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 (top panel) and tracking individuals over time (bottom panel).[5]

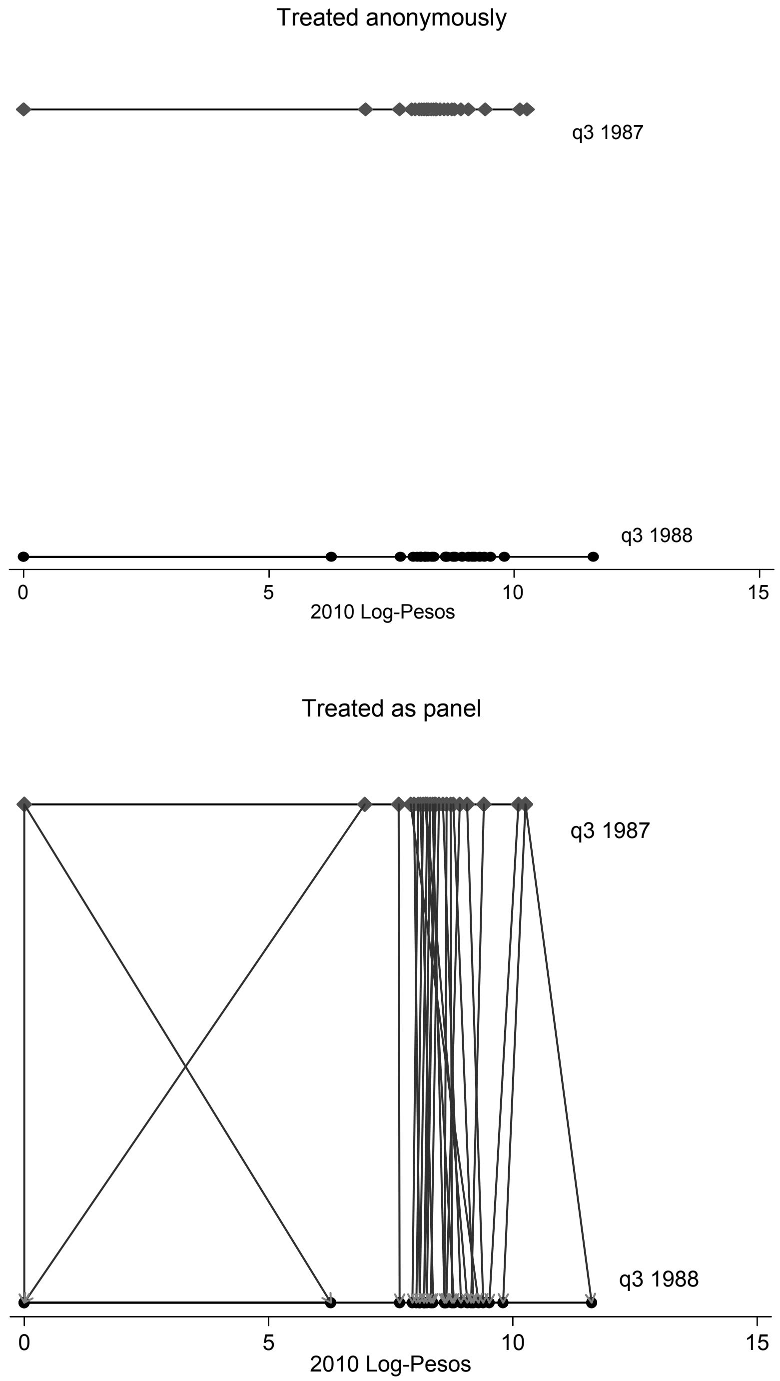
It is clear from these pictures that in spite of having a widening earnings distribution, some individuals experience large earnings changes, both in the positive and in the negative direction, leading to the aforementioned crossings.

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 the full sample of workers in the panel q3-87 to q3-88, we present in Table 2 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 3,000 and 4,000 pesos a month, 10 per cent of the labour force experience earnings *changes* larger than 3,000 pesos.

[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 percentiles.

[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.

Figure 4: Earnings distributions for 27 illustrative individuals in a period of rising inequality



Source: Authors' illustration based on ENEU panel q3-1987 to q3-1988.

Table 2: Transition matrix across fixed earnings categories, in thousands of 2010 Mexican pesos

Initial earnings (000s)	Final 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**

Note: The cells are per cent of the sample population.

Source: Authors' calculation based on ENEU panel q3-1987 to q3-1988.

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 (3) 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 per cent. However, the correlation between initial and final earnings is only 0.54 due to the aforementioned large changes, so earnings changes are convergent, as 1.36x1.13x0.54 = 0.83<1.

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 labour 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 with positive earnings, classified according to their *initial*-earnings quartile group.

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 quartile groups, indicating that initially richer individuals tend on average to stay richer one year later (left panel). Second, the distribution of log-earnings changes shifts to the *left* as we move from poorer to richer initial-earnings quartile groups, illustrating convergence between initial high and low earners (right panel). Third, there is a fair degree of overlap between the distributions of *final* log-earnings of individuals who initially belonged to different quartile groups (left panel). These overlaps are an indication of the moderate to large earnings changes among some members of the employed population. Finally, the distribution of log-earnings changes is more dispersed among the poorest and richest quartile groups than among the middle quartiles (right panel).

This section has presented several findings. First, the fact that inequality rises does not necessarily means that on average the *initially* rich are becoming richer at a faster rate than the *initially* poor. In fact, the data show 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, despite 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. (forthcoming) indicates that to an important extent these earnings changes are transitory in nature. If so, then it remains to assess how these changes influence a more permanent measure of inequality. One such measure can be obtained by looking at the inequality of the individual average earnings (where the average is taken across periods for each person). This analysis is presented next.

4 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 analyse 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 time. Hence by focusing on the inequality of these average earnings, we can obtain a measure of inequality less affected by transitory shocks.

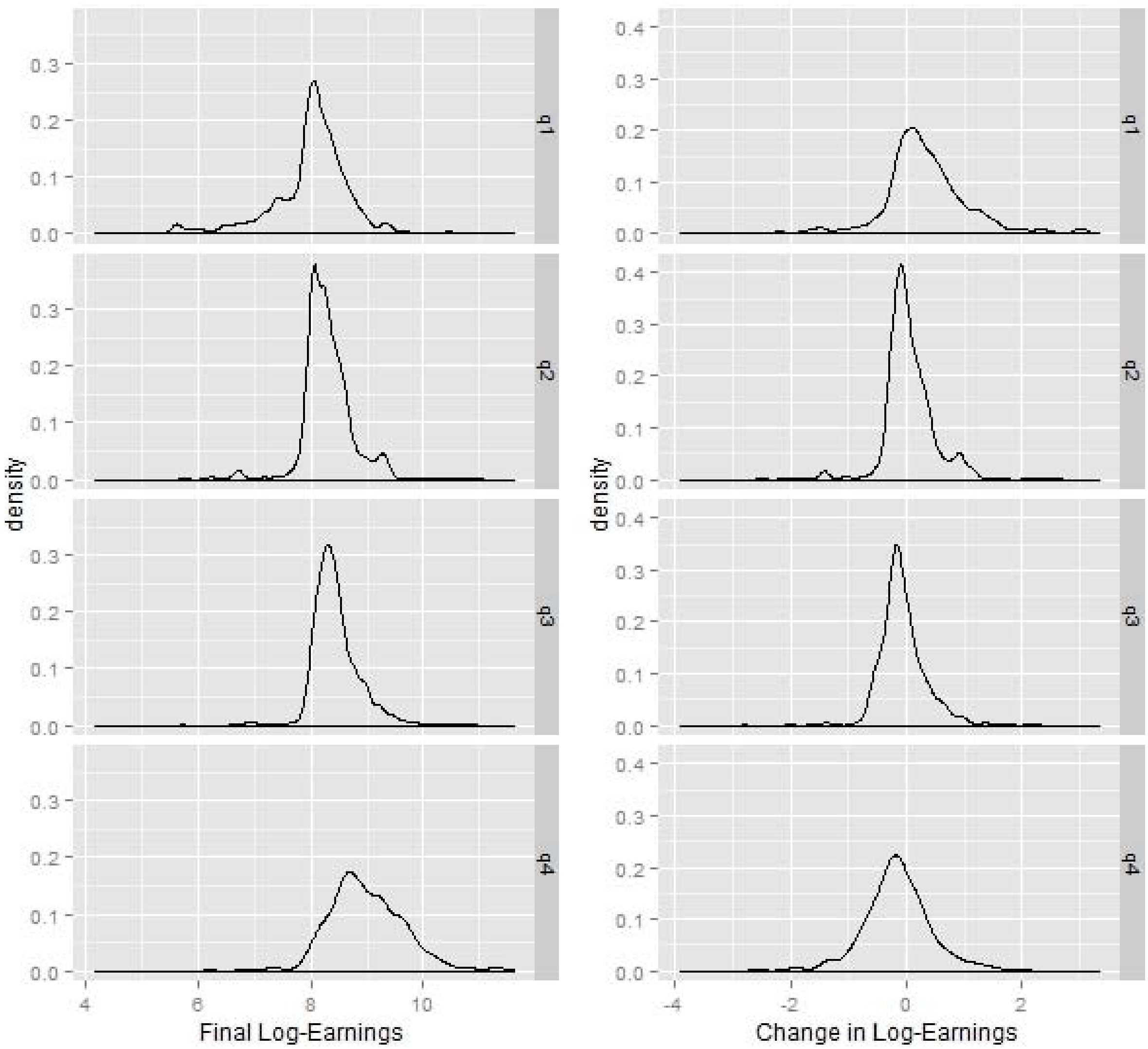
Furthermore, we can analyse whether economic mobility equalizes or disequalizes these average earnings, in comparison to the earnings that would occur in a world without such mobility. In particular, for an income inequality measure $I(\cdot)$, we can measure the inequality in average earnings $I(y_a)$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

$$EqM = I(y_0) - I(y_a) \tag{4}$$

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

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

Figure 5: Densities of final log-earnings and log-earnings changes by quartile group of the initial earnings, employed workers only

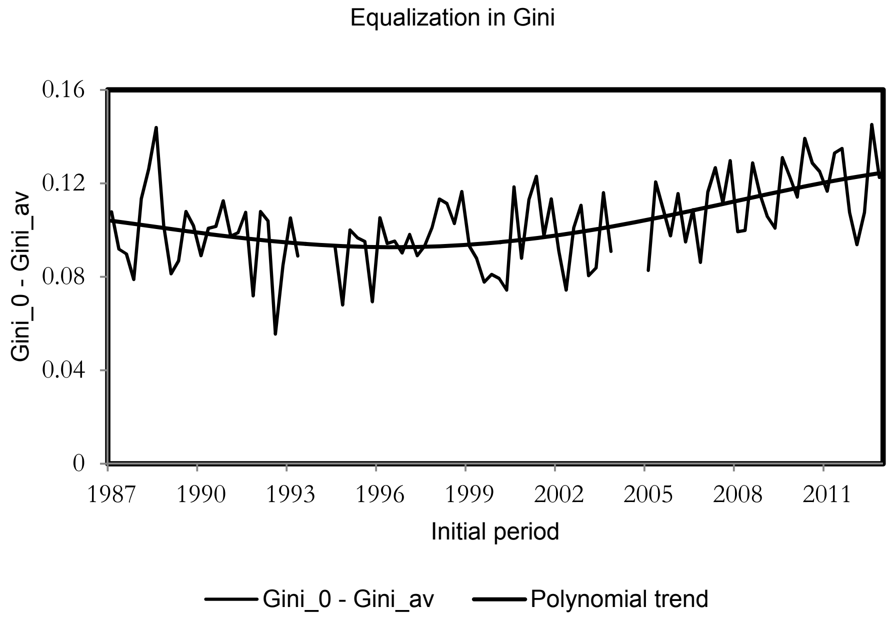


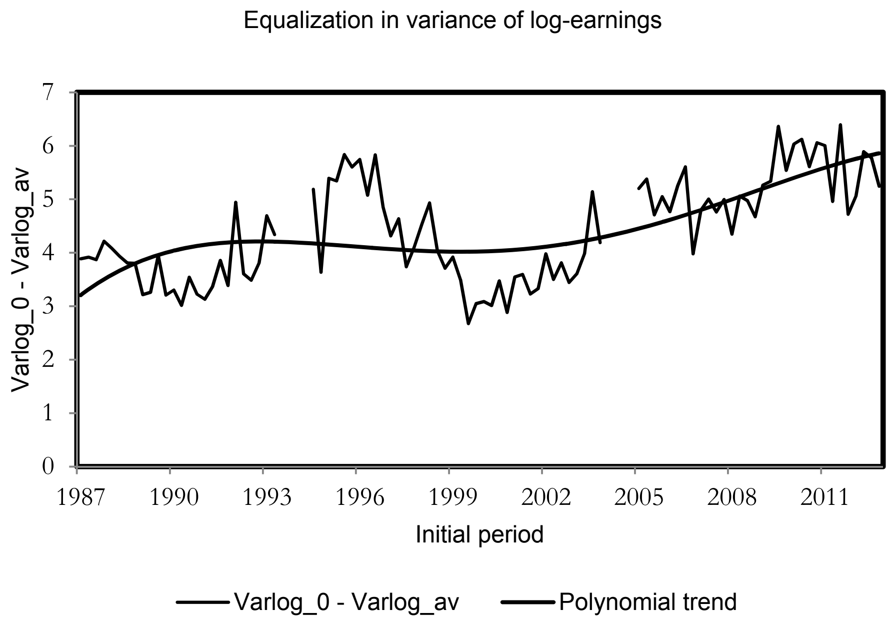
Source: Authors' illustration based on data from ENEU panel q3-1987 to q3-1988.

The bottom panel of Figure 2 plots the evolution of the inequality of individual average earnings, as gauged by the Gini index and by the variance of log-earnings. A quick comparison of this plot with the one on top 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 average earnings are smaller than the single-period ones.

This can also be appreciated in Figure 6, where we display the equalization brought about by mobility (4). 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 was more or less stable up to 2000 and increasing in the 2000s.

Figure 6: Equalization brought about by mobility





Source: Authors' illustration based on ENEU/ENOE data.

4.1 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

$$\ln y = W\gamma + u \tag{5}$$

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

$$\gamma_k cov(w_k, \ln y) \tag{6}$$

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 our Mexican data. In particular, we pooled data from several panels into two samples, one including all workers participating in the labour force (irrespective of whether they are employed or not), and another one including only those individuals who were employed over the five surveyed quarters.[7] The basic descriptive statistics of the pooled sample used are presented in Appendix Table A1.

The list of regressors included in equation (5) are a gender dummy, a 4th order age polynomial, a 2nd order polynomial in 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 those, city and period dummies were included as well. Since the employment dummy variables (unemployment, sector, industry and occupation) are time-varying, we include their average (across periods) in the regression of average earnings. This means that in such regressions we use as independent variables the fraction of time spent at each state by the workers. For brevity, the results pertaining to a group of variables, like the occupational dummies, or the age polynomials, are grouped together under a single heading in the tables reporting the decomposition results.[8]

The results for the full sample of labour force participants are included in the first two columns of Table 3. There, we observe that being unemployed is by far the greatest contributor to inequality of initial earnings (column 1). In fact, more than half of the dispersion of initial (log-) earnings is accounted for by the employment status of the worker (employed or unemployed). The second most important observable factor contributing to inequality is the sector of employment (formal/informal/self-employed), but it accounts for just 3 per cent of inequality. After that, occupation, years of schooling, gender and age each contribute between 1 and 2 per cent to the level of variance of log-earnings. Finally, around 40 per cent of the variance of log-earnings remains unexplained by the observable characteristics.

[7] In the case of the sample with all workers, we only consider individuals who were in the labour force during the five consecutive quarters.

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

Table 3: Accounting for levels of single-period and average log-earnings inequality. Percentage shares of V(ln y) are reported in square brackets

	All workers		Employed full year	
	Initial earnings	Average earnings	Initial earnings	Average earnings
V(ln y)	4.43	1.13	2.34	1.04
	[100]	[100]	[100]	[100]
Gender	0.049	0.035	0.064	0.039
	[1.1]	[3.1]	[2.7]	[3.8]
Age	0.049	0.013	0.033	0.013
	[1.1]	[1.2]	[1.4]	[1.3]
Education	0.057	0.058	0.072	0.069
	[1.3]	[5.1]	[3.1]	[6.7]
Unemployment	2.282	0.090		
	[51.6]	[7.9]		
Sector of employment	0.143	0.085	0.233	0.100
	[3.2]	[7.5]	[10.0]	[9.6]
Occupation	0.088	0.050	0.060	0.045
	[2.0]	[4.5]	[2.6]	[4.3]
Industry	0.001	0.018	0.065	0.024
	[0.02]	[1.6]	[2.8]	[2.3]
City	0.016	0.011	0.016	0.012
	[0.4]	[1.0]	[0.7]	[1.2]
Period	0.012	0.010	0.012	0.010
	[0.3]	[0.9]	[0.53]	[0.94]
Residuals	1.731	0.760	1.788	0.726
	[39.1]	[67.2]	[76.3]	[70.0]

Note: All earnings measures are in natural logarithms.

Source: Authors' calculation based on ENEU/ENOE data.

A very different picture arises when we look at the decomposition of the inequality of average earnings (column 2). Here, the largest fraction of the dispersion remains unaccounted for by observables, which only explain little more than 30 per cent of the variance of log-average earnings. Among the observables, unemployment status and sector of employment still account for the greatest share of variation, followed by schooling and occupational status, which account for about 5 per cent of the variation each. The significant drop in the explanatory power of unemployment in the average earnings equation reflects that transitions in and out of employment equalize earnings over time, something that will be better captured in the next section.

The last two columns of Table 3 present a similar estimation for the sample of workers who were employed for the full duration of the panel. There we see that among the observable characteristics, the sector of employment accounts for the largest fraction of the log-variance of both initial and average earnings (accounting for about 10 per cent in each case). Education accounts for between 3 and 7 per cent of inequality, with a larger impact on the dispersion of average earnings. In both cases, more than 70 per cent of the observed variation is unaccounted for by our observable variables.

4.2 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 (4).

In performing the accounting of the gap in equation (4) it is useful to distinguish between the contribution brought about by *changes in observable characteristics* and the *changes in the coefficients* of these characteristics, much in the spirit of Oaxaca (1973) and Juhn, Murphy, and 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

$$\ln y_c = W_a \gamma_0 \tag{7}$$

Denote by σ_{w0}^2 and σ_{wa}^2 the portion of the variance of initial and average log-earnings, respectively, accounted for by observable factors. Furthermore, denote by σ_c^2 the variance of the counterfactual log-earnings in (7). Finally, denote by σ_{r0}^2 and σ_{ra}^2 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

$$EqM = (\sigma_{w0}^2 - \sigma_c^2) + (\sigma_c^2 - \sigma_{wa}^2) + (\sigma_{r0}^2 - \sigma_{ra}^2) \tag{8}$$

The first term, $\sigma_{w0}^2 - \sigma_c^2$, 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, $(\sigma_c^2 - \sigma_{wa}^2)$, 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 differences in residuals between both models. For any of these terms a negative value would mean a disequalization of average earnings relative to initial earnings. One advantage of this method is that we can readily obtain the detailed contribution of individual observable variables to the first two terms in equation (8).

This decomposition is an application of the method proposed by Yun (2006), which in turn is an extension of the Fields (2003) method. The innovation of our paper is the application of this decomposition to analyse the equalization of average earnings relative to initial earnings due to mobility, rather than the changes in inequality between two anonymous distributions.[9] This decomposition is presented for the Mexican data in Table 4. Several interesting findings arise from this exercise.

Looking at the sample with all labour force participants, we observe that the largest contribution to equalizing earnings over time comes from changes in the employment status of the workers (44 per cent of the equalization), and from changes in the coefficient associated to this employment status (22 per cent of the equalization). 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 better-educated individuals (see for instance, Duval-Hernández and Orraca Romano 2011), mainly because poor uneducated workers cannot afford

[9] The full derivation of this decomposition is included in the Appendix of the paper. Other decompositions that are similar in nature are surveyed in Fortin, Lemieux, and Firpo (2011).

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 earnings.

Table 4: Accounting for earnings equalizing mobility. Percentage shares of V(ln y_0) (V(ln y_a) are reported in square brackets)

	All workers		Employed full year	
V(ln y_0) - V(ln y_a)	3.30	[100]	1.30	[100]
	Chars	Coeff	Chars	Coeff
Gender	0.003	0.011	0.001	0.024
	[0.1]	[0.3]	[0.05]	[1.9]
Age	0.005	0.031	0.001	0.018
	[0.2]	[0.9]	[0.1]	[1.4]
Education	-0.001	-6.62E-05	-2.66E-05	0.003
	[-0.04]	[-0.002]	[-0.002]	[0.21]
Unemployment	1.455	0.738		
	[44.1]	[22.4]		
Sector of employment	-0.003	0.061	0.068	0.066
	[-0.1]	[1.8]	[5.2]	[5.0]
Occupation	0.024	0.014	0.007	0.007
	[0.7]	[0.4]	[0.6]	[0.6]
Industry	-0.026	0.009	0.010	0.031
	[-0.8]	[0.3]	[0.7]	[2.41]
City	3.10E-05	0.004	-6.76E-05	0.004
	[0.001]	[0.1]	[-0.005]	[0.3]
Period	0.002	-0.0003	0.0002	0.002
	[0.1]	[-0.01]	[0.01]	[0.2]
Residuals	0.970		1.062	
	[29.4]		[81.4]	

Notes: ln y_0 denotes initial log-earnings, ln y_a denotes average log-earnings. Char and Coeff are the effects associated with changes in characteristics and coefficients, respectively.

Source: Authors' calculations based on ENEU/ENOE data.

All other observable factors play a negligible role in this equalization, and about 30 per cent of the equalization remains accounted for by the residuals.

The results for the sample of employed workers in the last two columns of the table show that only 20 per cent of the equalization can be accounted for by observable factors, and half of this amount is attributable to changes in the sector of employment (formal/informal/self-employed) and its coefficients.[10]

5 Conclusions

This paper showed how our view of who benefits and who is hurt as the economy changes over time 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 2 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 with 96 panels, each of which tracks the earnings of workers for one year in urban Mexico.

In the empirical analysis we observed that while earnings inequality sometimes 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 experience large losses, while another small fraction of the initially poor experience large gains. On average, though, most people tend to experience small to moderate 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 less transitory 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 4 of the paper.

In that section, we showed that individual earnings averaged over five quarters are more equally distributed than 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. Permanent characteristics like gender and years of schooling only account for a small fraction of the observed dispersion in earnings, mainly for individual average earnings.

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 for the sample of labour force participants. However, for the sample of full-year employed workers, sector of employment variables account for most of the equalization explained by observables, but 80 per cent of the total equalization remains unaccounted for by the observable factors.

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

[10] The fact that time-invariant factors, like gender and age (which are invariant in these one-year panels), contribute to the impact of *changing* characteristics on the inequality gap EqM (the column labelled 'Chars'), is due to the correlation that these factors have with time-varying characteristics. For more details, refer to the Appendix.

initial earnings. Also, it would be interesting to assess the role that labour market policies play in such equalization/disequalization.

References

Duval-Hernández, R., G.S. Fields, and G.H. Jakubson (2014). 'Changing Income Inequality and Panel Income Changes', unpublished manuscript.

Fields, G.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*, 22: 1-38.

Fields, G.S. (2010). 'Does Income Mobility Equalize Longer-Term Incomes? New Measures of an Old Concept'. *Journal of Economic Inequality*, 8: 409–27.

Fields, G.S., R. Duval-Hernández, S. Freije, and M.L. Sánchez-Puerta (forthcoming). 'Earnings Mobility, Inequality, and Economic Growth in Argentina, Mexico, and Venezuela'. *Journal of Economic Inequality*.

Fortin, N., T. Lemieux, and S. Firpo (2011). 'Decomposition Methods in Economics'. In D. Card and O. Ashenfelter (eds), *Handbook of Labor Economics*, Volume 4. Amsterdam: Elsevier Science.

Furceri, D. (2005). 'β and σ-convergence: A Mathematical Relation of Causality'. *Economics Letters*, 89: 212–15.

Juhn, C., K.M. Murphy, and B. Pierce (1993). 'Wage Inequality and the Rise in Returns to Skill'. *Journal of Political Economy*, 101(3): 410–42.

Oaxaca, R. (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, M.S (2006). 'Earnings Inequality in USA, 1969–99: Comparing Inequality Using Earnings Equations'. *Review of Income and Wealth*, 52(1): 127-44.

Appendix

Appendix Table A1. Descriptive statistics of regression sample

Variable	Obs	Mean	Std. Dev.	Min	Max
Initial earnings	295,417	5,721	8,197	0	2,531,458
Earnings change	295,417	33.3	7,770.9	-350,188	2,522,127
Male	295,417	0.69	0.46	0	1
Age	295,417	36.15	11.24	18	65
Years schooling	295,417	9.42	4.22	0	25
Unemployed	295,417	0.03	0.18	0	1
Sector of employment					
Formal employee	295,417	0.52	0.50	0	1
Informal employee	295,417	0.26	0.44	0	1
Self-employed	295,417	0.18	0.39	0	1
Occupation					
Professional, manager	295,417	0.14	0.35	0	1
Supervisor, operator	295,417	0.12	0.33	0	1
Production, Craft	295,417	0.22	0.41	0	1
Transport, mechanic	295,417	0.07	0.25	0	1
Clerical, sales	295,417	0.26	0.44	0	1
Services	295,417	0.13	0.34	0	1
Security	295,417	0.03	0.17	0	1
Industry					
Agricultural	295,417	0.01	0.08	0	1
Extractive, electricity	295,417	0.01	0.09	0	1
Construction	295,417	0.07	0.25	0	1
Manufacturing	295,417	0.23	0.42	0	1
Trade	295,417	0.18	0.39	0	1
Services	295,417	0.41	0.49	0	1
Public admin	295,417	0.06	0.24	0	1

Source: Authors' calculations based on ENEU/ENOE data.

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,

$$\ln y_0 = Z\alpha_0 + X_0\beta_0 + u_0 \tag{A1}$$

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 (5) in the text provides a compact way of expressing this equation as

$$\ln y_0 = W_0\gamma_0 + u_0 \tag{A2}$$

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

Furthermore, assume the error term is uncorrelated with the regressors

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

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

$$\begin{gathered} \ln y_a = Z\alpha_a + X_a\beta_a + u_a \Leftrightarrow \\ \ln y_a = W_a\gamma_a + u_a \end{gathered} \tag{A3}$$

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

$$\gamma_{k0} cov(w_{k0}, \ln y_0) \tag{A4}$$

for the initial period equation, and as

$$\gamma_{ka} cov(w_{ka}, \ln y_a) \tag{A5}$$

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_{r0}^2 and σ_{ra}^2, 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

$$\sigma_{ws}^2 = \Sigma_{k=1}^{K} \gamma_{ks} cov(w_{ks}, \ln y_s) \quad for\ s \in \{0, a\}$$

we can then express the log-variance of earnings as

$$V(\ln y_s) = \sigma_{ws}^2 + \sigma_{rs}^2 \quad for\ s \in \{0, a\} \tag{A6}$$

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

$$\ln y_c = W_a \gamma_0 \tag{A7}$$

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 an Oaxaca decomposition.

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

$$\sigma_c^2 = \Sigma_{k=1}^{K} \gamma_{k0} cov(w_{ka}, \ln y_c) \tag{A8}$$

Using equations (A6) and (A8) we can rewrite the total equalization gap as

$$V(\ln y_0) - V(\ln y_a) = (\sigma_{w0}^2 - \sigma_c^2) + (\sigma_c^2 - \sigma_{wa}^2) + (\sigma_{r0}^2 - \sigma_{ra}^2) \tag{A9}$$

The first term, $\sigma_{w0}^2 - \sigma_c^2$, represents the equalization brought about by changes in the regressors, 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

$$\gamma_{l0} cov(w_{l0}, \ln y_0) - \gamma_{l0} cov(w_{la}, \ln y_c)$$

$$= \gamma_{l0} cov(w_{l0}, w_{10}\gamma_{10} + \cdots + w_{K0}\gamma_{K0}) - \gamma_{l0} cov(w_{la}, w_{1a}\gamma_{10} + \cdots + w_{Ka}\gamma_{K0}) \quad \text{(A10)}$$

$$= \gamma_{l0} \Sigma_{k=1}^{K} \gamma_{k0} [cov(w_{l0}, w_{k0}) - cov(w_{la}, w_{ka})]$$

while the ‘coefficient effect’ can be written as

$$\gamma_{l0} cov(w_{la}, \ln y_c) - \gamma_{la} cov(w_{la}, \ln y_a)$$

$$= \gamma_{l0} cov(w_{la}, w_{1a}\gamma_{10} + \cdots + w_{Ka}\gamma_{K0}) - \gamma_{la} cov(w_{la}, w_{1a}\gamma_{1a} + \cdots + w_{Ka}\gamma_{Ka}) \quad \text{(A11)}$$

$$= \Sigma_{k=1}^{K} (\gamma_{l0}\gamma_{k0} - \gamma_{la}\gamma_{ka}) cov(w_{la}, w_{ka})$$

Expression (A10) 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.