the review of income and wealth

Review of Income and Wealth
Series 63, Number 4, December 2017
DOI: 10.1111/roiw.12271

CROSS-SECTIONAL VERSUS PANEL INCOME APPROACHES: ANALYZING INCOME DISTRIBUTION CHANGES FOR THE CASE OF MEXICO

BY ROBERT DUVAL-HERNÁNDEZ*

University of Cyprus, IZA, CIDE

GARY S. FIELDS

Department of Economics, Cornell University, IZA, WIDER

AND

GEORGE H. JAKUBSON

Department of Economics, Cornell University

In this paper we reconcile, both theoretically and empirically, changes in cross-sectional inequality with patterns of panel income changes during periods of economic growth and decline. Using panel earnings data from Mexico, we find that the panel changes are convergent in almost every period, the reason being that a large number of individuals experience small convergent earnings changes while a small number of individuals experience large and convergent earnings changes. We examine what accounts for the inequality of log-earnings at a point in time and for the inequality of the log of earnings averaged over five quarters. We find that the equalization brought about by panel earnings changes is mainly associated with changes in employment status and in sector of employment and not by personal characteristics such as schooling, age, and gender.

JEL Codes: J31, D63

Keywords: income inequality, panel income changes

1. INTRODUCTION

In this paper we compare standard cross-sectional analyses of inequality changes to the analysis of panel income changes. Specifically, we explore how it is possible to have convergent panel income changes even during periods in which inequality is rising. Moreover, using panel earnings data from Mexico, we compare inequality at a given point in time with inequality in earnings averaged over five consecutive quarters and explore what observable factors account for their difference.

The literature analyzing which income groups benefited how much when economic growth or decline took place has devoted a lot of attention to comparing the inequality of income distributions over two or more points in time. By looking

Note: The authors would like to thank two anonymous referees and the journal Editor for their valuable comments and suggestions on an earlier draft of this paper. Needless to say, all remaining errors are our responsibility.

*Correspondence to: Robert Duval-Hernández, Research Officer, Economics Research Centre, University of Cyprus, P.O. Box 20537, CY-1678 Nicosia, Cyprus (rduval@ucy.ac.cy).

at how the shape of this distribution has changed, this literature has used anonymous cross-sectional data to compare inequality 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, whether or not 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 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 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 percent of the income distribution became poorer might change depending on whether we look at the incomes of the anonymous bottom 10 percent, or whether we track with panel data the incomes of those who initially were in the bottom 10 percent. In other words, the standard inequality analysis follows the evolution of incomes of whoever is in the bottom 10 percent, 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 percent, but who might or might not have moved to other points in the income distribution.

In Section 2, we summarize in an accessible manner our recent theoretical findings on how the answers provided by standard inequality analyzes and by panel income change regressions can be reconciled. In Section 3, we turn to an empirical analysis of Mexico. Mexico is of interest for a variety of reasons: because of the availability of standard cross-sectional data as well as five-quarter panel data over several decades; because the data includes periods of economic growth and economic decline and of rising and falling income inequality; and because there are still few studies exploring income dynamics in developing countries.

In addition to comparing the cross-sectional and panel income approaches under very diverse macroeconomic scenarios, we also examine in Section 3.3, how our view of inequality is altered if instead of looking at earnings inequality at a point in time (i.e. monthly earnings in a given quarter), we focus on the inequality of average earnings, where the average is taken over 5 consecutive quarters. 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 earnings changes. Finally, Section 4 summarizes our findings.

2. Reconciling the Anonymous and Panel Income Change Approaches Theoretically

In this section we will present a broad summary of how standard inequality analyzes can be reconciled with income change regressions using panel data. In what follows, "income" will be the generic term that we will use for our variable of interest, which could be total income, labor earnings, hourly wages, consumption, or something else.

There is a large literature on how to measure relative income 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 (Sen, 1997; Cowell, 2011).

To gauge convergence or divergence in incomes it is traditional 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 (or whatever the relevant currency unit is in a particular country), log-dollars, shares of mean (or of total) income, etc.; Δy is the change in that income variable between time 0 and time 1; 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 $\beta_y \gtrless 1$ (e.g. Atkinson, Bourguignon, and Morrisson, 1992).[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. Out of the four combinations listed above, most practitioners tend to accept the validity of i) and iv). 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 and likewise for the rich.

[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.

Empirically though, in analyses of panel income changes at the individual level it is almost always found that such changes are convergent, i.e. the initially disadvantaged individuals tend to have more positive (less negative) income changes than the individuals who were initially rich (Dragoset and Fields, 2008; Fields, 2010b; Fields et al., 2015), on average irrespectively of whether inequality rises or falls.

In this section, we present a non-technical summary of some of the theoretical findings in our work Duval-Hernández, Fields, and Jakubson, (2016), where we explain what underlying conditions need to occur for the reconciliation of convergent panel income changes and rising inequality to take place.

More specifically, in that paper, we reconcile rising inequality as judged by the Lorenz-criterion or by some of the most commonly used inequality indices in the literature, with convergence in regressions like (1) and (2), for incomes measured in dollars (*d*), as shares of mean income, in log-dollars (to approximate proportional income changes), or in a regression with exact proportional changes

$$\frac{d_1-d_0}{d_0}=\phi+\theta d_0+u_{pch}. \tag{3}$$

As previously mentioned, having rising inequality means that the incomes of the *anonymous* rich are moving farther away from the incomes of the *anonymous* poor. Having convergent panel income changes means that on average the *initially* poor are experiencing more positive (or less negative) income changes than the *initially* rich.

In Duval-Hernández, Fields, and Jakubson (2016) we prove that one 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 of how this can occur, consider the simple 5-person income vector in the initial period

$$\text{ExI.} \qquad y_0=[20,\ 41,\ 45,\ 49,\ 70]$$

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. Yet, the coefficient δ_d of regression (1), when expressed in dollars,

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

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. In fact it is easy to verify that for this example, there is convergence if income is measured in shares, and that proportional income changes are also convergent, irrespective of whether they are approximated through a regression in logs or by using the exact proportional changes as a dependent variable, as in equation (3). In other words, the very large income gains of the initially poor coupled with the large income losses of the initially rich lead to convergence and rising inequality, all at the same time.

Another instructive reconciliation in the case of convergence in dollars is the following. Denote by r_d the correlation coefficient between initial and final dollars, i.e.

$$r_d = \frac{cov(d_0, d_1)}{\sqrt{V(d_1)}\sqrt{V(d_0)}} \quad .$$

Letting $CV(d_t)$ 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 (2016) that whether dollar changes are convergent (i.e. $\beta_d < 1$, or equivalently $\delta_d < 0$) or divergent depends on the condition

$$\delta_d \gtreqless 0 \iff r_d \frac{CV(d_1)}{CV(d_0)}(1+g) \gtreqless 1. \tag{4}$$

In other words, Equation (4) shows that dollar changes can be convergent, even when inequality is rising as measured by a Lorenz-consistent index like the coefficient of variation (i.e. if $CV(d_1) > CV(d_0)$), provided that the correlation coefficient between initial and final incomes r_d is small enough or if there is a sufficiently large decline in average income ($g < 0$). Normally, in empirical applications, r_d would be positive. If it is positive but not too large, the expression in (4) could be less than one. Of course, if r_d is negative, the expression in (4) would surely be less than one. In the case of positive income growth ($g > 0$), a small positive correlation between initial and final dollars r_d indicates that there are numerous and/or large income changes such that initial earnings are less important in predicting final earnings (in an R-squared sense).

There are however, other possible ways in which convergent income changes can be reconciled with rising relative inequality. Consider for instance the income transition

ExII. $\quad [1, 1, 1, 1, 1, 1, 1, 1, 6.1, 8.9] \rightarrow [1, 1, 1, 1, 1, 1, 1, 1, 6, 9]$

In this case there is a Lorenz-worsening, yet regression (1) when expressed in log-dollars, exhibits the existence of convergence ($\delta_{log} = -0.0005$).

Similarly, in the transition

ExIII. $\quad [7, 23] \rightarrow [5, 20]$

there is a Lorenz-worsening, yet regression (1) when expressed in dollars exhibits convergence ($\delta_d = -0.0625$).

Interestingly, in the last two examples we can reconcile rising inequality with convergence, *even in the absence of positional changes.* In our aforementioned paper, we show that the reason situations like ExII arise is because log-incomes can be convergent if a rank-preserving *disequalizing* transfer occurs sufficiently high-up in the income distribution. In contrast, situations like ExIII can arise because we have paired an income-change regression (1) specified in dollars with a measure of *relative* inequality. As is well-known, relative inequality comparisons are invariant to proportional changes in incomes, yet the coefficients of a dollar-change regression are affected by such proportional changes. In ExIII, the income share of the rich grew relative to that of the poor, yet the dollar losses of the rich ($3) were larger than the dollar losses of the poor ($2).

In summary, for the reconciliation of rising inequality with convergent panel income changes we require for at least one of the following to occur: i) panel income changes for identified individuals are numerous and in a convergent pattern, or are large enough, as in our example ExI above, or ii) we have a specific combination of panel income change regressions with specific relative inequality measures as in our example ExII with log-dollars, or iii) a strong negative aggregate reduction in incomes takes place along with a dollar change regression like that in our example ExIII above.

In the following section we illustrate the reconciliation of rising relative inequality with convergent panel income changes using an empirical exploration of earnings data for urban Mexican labor markets.

3. Empirical Analysis for Mexico

In the previous section we explained the mechanisms that need to operate in order to reconcile rising inequality with convergent panel income changes. In this section we present a real life example comparing the evolution of inequality and panel changes 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, as well as periods with rising and falling inequality.

3.1. *Data*

The dataset used is the Encuesta Nacional de Empleo Urbano (ENEU) and its successor, the Encuesta Nacional de Ocupación y Empleo (ENOE). These labor market surveys collect information on employment, earnings and sociodemographic characteristics of the population, and they are used to estimate the official unemployment rate.

These surveys are conducted through rotating panels where the same individuals are followed for five consecutive quarters.[2] 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.[3]

[2]Each of the five rotating samples has the same sampling probability.

[3]Over these years there have been changes in the survey instruments. While the samples remain broadly comparable over the years, we cannot conduct an analysis with the specific panels that underwent transitions in sampling frames and/or questionnaires, as the individuals in such panels cannot be consistently be traced over the five quarters. For this reason, the analysis and results presented here exclude certain years, and some of the graphs displayed present gaps in those periods. More specifically, the panels excluded are the ones starting in the 3rd qr of 1993, 4th qr of 1993, 1st qr of 1994, 2nd qr of 1994, as well as all the quarters starting in 2004.

Over the years, the geographical coverage of the survey has changed, including first the main urban centers in the country, 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.[4]

We limit our sample to labor force participants (either employed or unemployed) aged between 18 and 65 years at the end of the panel. Furthermore, we only include individuals who remain in the survey during the five-consecutive quarters. This is done in order to maintain a consistent sample between the estimates presented in Section 3.2, where only quarters 1 and 5 are used, and the ones in Section 3.3, where the full five quarters are used. The basic descriptive statistics of the pooled sample used are presented in Table A-1 in the Appendix.

Our variable of interest is monthly earnings measured in 2010 Mexican Pesos. We assign an earnings level of 0 to unemployed individuals, except 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.[5] All the analysis is performed using the survey sampling weights of the last interview quarter.

We close this data section with two notes of caution. First, although in each panel the initial subsample is representative of the urban population at the city level, attrition and non-reporting in earnings are significant. In particular, about half of the original target sample disappears due to individuals leaving the survey or refusing to answer earnings questions (see the evidence presented in Duval-Hernández, 2006 and Campos-Vázquez, 2013). Unlike other labor surveys, the ENEU/ENOE do not provide a measure of imputed earnings, nor do they provide longitudinal weights that allow one to adjust for attrition in the sample.

The analysis presented in Campos-Vázquez (2013) shows that non-reporting of earnings is positively associated with years of schooling, and hence imputing missing earnings leads to higher levels of cross-sectional inequality than the ones using reported data alone. However, the trends in changes in inequality are unaltered by this non-reporting. Also, as indicated by the analysis in Duval-Hernández (2006), the high levels of attrition over the 5 quarters require us to take with caution any results based on the reported data in the panel. Coming up with a convincing method to address these shortcomings in the data is an important issue pending in the literature. Exploring this topic requires a separate paper on its own; hence we proceed by presenting the results using the data as reported in the survey.

Our second note of caution relates to the possible biases that might arise due to measurement error in the earnings variable. The impacts of classical measurement error are well understood. However, there is no reason to believe that measurement error in earnings follows that restrictive model. There is evidence from the U.S. (Gottschalk and Huynh, 2010) which suggests that the measurement

[4]While excluding certain urban areas added in recent years can lead to a loss of representativeness of the overall urban population at the national level, this bias is likely to be small as the main urban centers remain in the sample throughout the years.

[5]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*. (2015).

error is distinctly non-classical. They had a validation dataset in addition to the survey data from which they could calculate a measurement error and assess its properties. They found that the parameters of equations like those we use to define convergence and divergence were not much affected by the measurement error. Essentially, the attenuation bias from the measurement error was offset by the persistence over time of the errors. Measures of inequality, on the other hand, were affected. Measurement error leads to an underestimate of inequality. Essentially, high income respondents were more likely to underreport and/or to underreport to a greater degree.

We do not have any data available which we could use to construct a validation sample, and are therefore restricted to the survey data. We hope that the lessons from Gottschalk and Huynh apply to the case of Mexico. If so, then our findings on convergence and divergence will be fairly robust to measurement errors. Our inequality comparisons, which find extended periods with first rising and then falling inequality, are likely to be biased downwards in terms of levels but could be accurate in terms of the changes on which we focus.

3.2. *Inequality Changes and Convergent Earnings Reconciled*

In the top two panels of Figure 1 we present the evolution of earnings inequality over the period 1987–2013. There we observe that inequality rose during the years of economic liberalization in Mexico 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 kept falling or started rising, depending on which measure is used to gauge it.

In addition, in this figure we present the δ_y coefficients from regression (1) for yearly changes in earnings—i.e. from the initial quarter to the same quarter one year after—and test for divergence/convergence according to $\delta_y \gtrless 0$. The center-left figure presents convergence in earnings shares because this concept is the one most closely associated to changes in relative inequality. In particular, as mentioned in Section 2, standard relative inequality analyses are concerned with the distribution of anonymous earnings shares. It is apparent from these coefficients that in spite of the ups and downs in inequality displayed in the top two figures, share changes are always convergent, and nearly always significantly so. Since for the most part our interest is to learn whether there has been convergence in pesos, we present in the center-right figure the corresponding coefficients for the linear model of earnings changes measured in pesos. Here too, there is overwhelming evidence of convergence.

Also, at the bottom of Figure 1 we display the coefficients for the panel regressions estimating convergence in proportional earnings changes. The left figure presents the standard logarithmic approximation to these changes, while the right figure depicts the coefficient for the regression using exact proportional changes as a dependent variable (as in equation (3)). Once again, the patterns are overwhelmingly convergent.

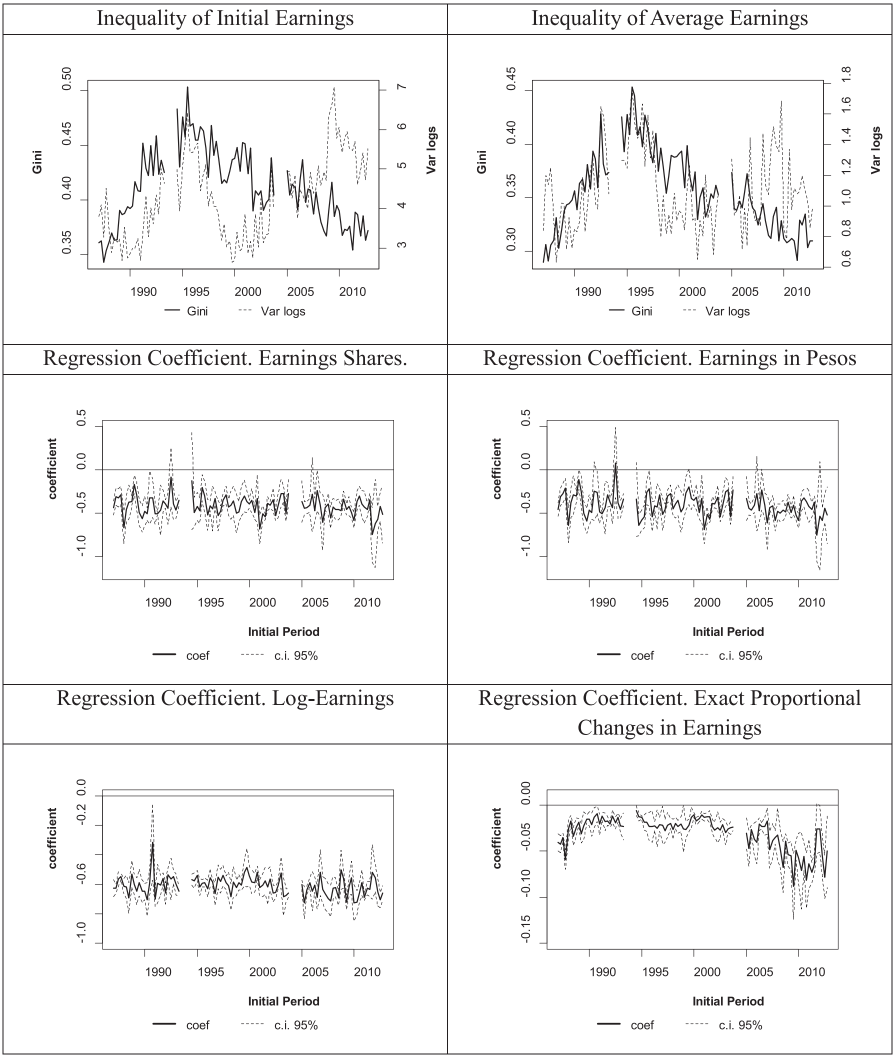

Figure 1. Earnings Inequality and Convergence

Confidence intervals for regression coefficients based on standard errors clustered by period and city

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

A joint look at the results in Figure 1 indicates that regardless of whether inequality rises or falls, all four different ways of earnings changes exhibit convergence. The convergence results are also remarkably stable over the business cycle.

As mentioned in the previous section, this pattern can arise if i) panel earnings changes for identified individuals are numerous and/or large and convergent, or ii) there are changes at the top of the earnings distribution that lead to convergence in log-pesos, or iii) there is a strong economic decline that creates convergence in pesos (but not necessarily in shares or in proportional terms). Given that convergence is found using all four income change concepts, and that convergence occurs in periods with positive and negative growth, the most likely explanation is i).

As further evidence that convergent earnings changes and rising inequality arise from the fact that there are numerous convergent earnings changes, some of them large, we turn to equation (4),

$$\delta_d \gtrless 0 \iff r_d \frac{CV(d_1)}{CV(d_0)}(1+g) \gtrless 1,$$

which offers a way to reconcile convergent peso changes with an increase in the coefficient of variation.

In Figure A-1 in the Appendix we show the different components of this equation. In particular, the top panel shows the ratio of the coefficients of variation. Values greater than one indicate a rise in inequality. Similarly, the second graph illustrates the economy-wide growth in earnings $(1+g)$. It is clear from these graphs that in the data there is a full combination of rising and falling inequality, together with periods of growth and decline. Finally, in the bottom panel we present the correlation coefficient between initial and final earnings in pesos r_d. In that figure, we also include a "divergence bound," which is the largest value that r_d could take before leading to divergent peso changes. In other words, the divergence bound equals

$$\frac{CV(d_0)}{CV(d_1)(1+g)}.$$

This last panel shows that in almost all periods the correlation coefficient is below the bound after which it would register divergent peso changes.[6]

To better understand the nature of these earnings changes, we will look in detail at the data from 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 together with convergent earning changes.

In Table 1 we present a transition matrix between fixed earnings categories. This matrix shows that while most individuals have small earnings 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, ten percent of the labor force experience earnings *changes* larger than 3 thousand pesos in absolute value. Even this small fraction of large changes translates into a low correlation coefficient between initial and final earnings

One 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.

[6]Only in the 3rd quarter of 1992 does r_d take on a value (0.38) greater than the divergence bound (which equals 0.35), and hence there is divergence in Pesos, $\delta_d = 0.07$.

TABLE 1
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)	**2.5**	0.7	0.4	1.2	0.7	0.7	0.0	0.0	0.2	**6.3**
[1,2)	0.2	**1.3**	0.8	0.9	0.7	0.1	0.0	0.0	0.2	**4.2**
[2,3)	0.5	0.8	**1.6**	2.8	1.1	0.5	0.0	0.4	0.4	**8.2**
[3,4)	1.1	0.5	4.6	**11.8**	5.2	3.3	1.2	0.6	1.3	**29.6**
[4,5)	0.3	0.2	0.8	9.6	**5.5**	1.6	1.0	0.3	1.4	**20.7**
[5,6)	0.1	0.0	0.4	2.9	2.2	**1.6**	1.1	0.6	0.9	**9.7**
[6,7)	0.0	0.1	0.2	1.3	0.7	1.1	**1.1**	0.6	1.0	**6.1**
[7,8)	0.0	0.0	0.1	0.4	0.3	0.8	0.3	**0.7**	1.2	**3.9**
[8,)	0.2	0.1	0.1	0.5	0.5	0.8	1.1	0.8	**7.3**	**11.3**
Total	**4.9**	**3.8**	**8.9**	**31.4**	**16.8**	**10.5**	**5.8**	**4.0**	**13.9**	**100.0**

The cells are percentages of the sample population.
Source: Authors' calculation based on ENEU panel q3-1987 to q3-1988.

This can be better appreciated by looking at Figure 2, 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 four quarters 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 initial quartile groups than among the middle quartiles (right panel).

The evidence presented for this specific panel (3rd qr 1987—3rd qr 1988) leads to a final important question: Are the small fraction of large earnings changes responsible for the convergence in earnings presented in Figure 1? To answer that question, we now repeat the four previous convergence regressions, this time excluding in each period the individuals in the top and bottom 10 percent of the distribution of changes. In other words, we re-estimate our convergence coefficients, this time with only the 80 percent of population that had the smallest earnings changes in absolute value. The results of this exercise are displayed in Figure 3. Two important lessons can be drawn from these pictures. First, the degree of convergence is smaller than the one displayed in their un-trimmed counterparts (Figure 1). In other words, a small number of individuals experiencing large earnings changes contribute to a stronger recorded level of convergence. Second, even in the absence of these large earnings changes, the remaining changes are also convergent in most of the panels.

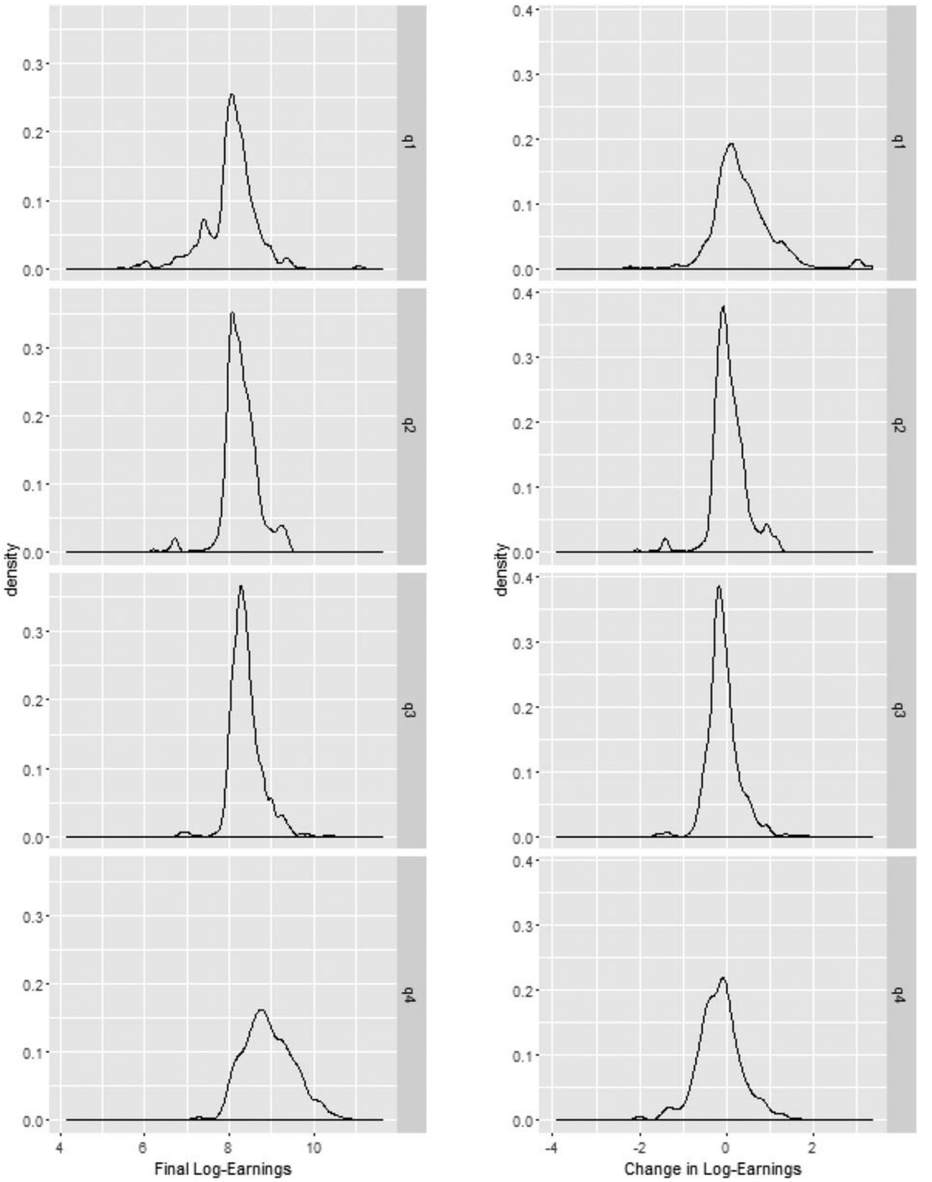

Figure 2. Densities of Final Log-Earnings and Log-Earnings Changes by Quartile Group of the Initial Earnings. Employed Workers Only

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

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, convergent earnings changes, meaning that the initially low-earners experience larger gains, in pesos, in shares, and in proportions, than the initially 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 four quarters later. Instead, while the majority of the population experience moderate convergent earnings changes, there is a small fraction of the population that has large convergent earnings changes.

Research presented in Fields *et al*. (2015) indicates that to an important extent such earnings changes are transitory in nature. If so, then it remains to assess how these changes influence a less transitory measure of inequality. One such measure can be obtained by looking at the inequality of individual average earnings (where the average is taken across the five quarters over which each person is observed). This analysis is presented next.

3.3. *Inequality of Average Earnings and Equalizing Panel Earnings Changes*

The previous sections demonstrated that the evolution of single-period inequality among anonymous individuals does not capture the effects of earnings changes over time. One classic way to incorporate the earnings changes in the analysis of inequality is to look at the inequality of *average* earnings y_a, which in our case will be defined as the average earnings of an individual over the five quarters for which we observe her in the Mexican panels.

Unlike earnings measured at a single point in time, average earnings over several periods capture the effects of earnings changes by averaging out the ups and downs in this variable over time. Hence by focusing on these average earnings, we can obtain a measure of earnings inequality less affected by transitory shocks.

Furthermore, we can analyze whether the individual earnings changes make these average earnings more equally distributed, in comparison to the earnings that would occur in a world without earnings changes. 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 earnings not taken place, i.e. to $I(y_0)$. This measure EQ (for equalization brought about by panel income changes)

$$EQ = I(y_0) - I(y_a) \tag{5}$$

would take positive values if earnings changes *equalized* average earnings relative to initial, and it would take negative values if it *disequalized* them. This measure is just an algebraic transformation of Fields' (2010a) index of mobility as an equalizer of longer-term incomes relative to initial.

It is important to emphasize here that given the short-lived nature of our data, our measure of average earnings mainly averages out the impact of transitory short-run changes. In order to capture a more permanent impact of economic mobility one needs to rely on longer panels. However, it is also important to mention that all the methods presented in this section can be applied to longer panels, and hence can be used to analyze questions pertaining to the long-term impact of earnings changes among panel people.

The top-right graph in Figure 1 plots the evolution of the inequality of individual average earnings, as gauged by the Gini index and by the variance of log-

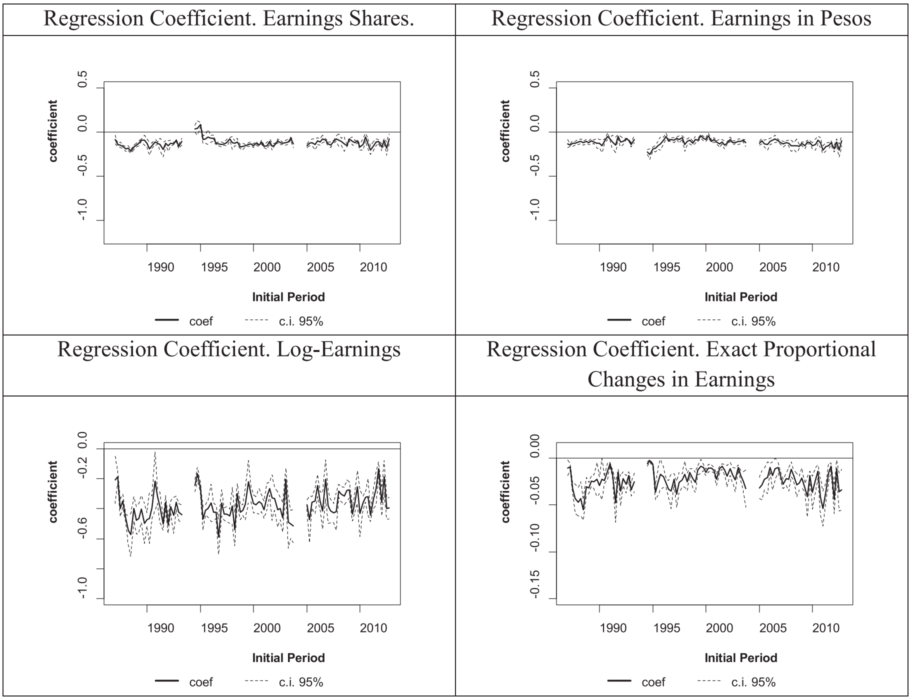

Figure 3. Convergence Coefficients from Linear Regression Model. Sample trimming at each period bottom/top 10% of largest earnings changes

Confidence intervals based on standard errors clustered by period and city.
Source: Authors' illustration based on ENEU/ENOE data.

earnings. A quick comparison of this plot with the one on the top-left reveals that for the most part inequality of average earnings follows the same time path 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 by looking at Figure 4, which shows positive signs for every panel, meaning that average earnings are more equally distributed than are single-period earnings. However, the trends in the equalization measures vary depending on the inequality index used.

3.3.1.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 inequality is measured by the variance of log-earnings. A simple way to do this is to apply the method developed by Fields (2003).

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

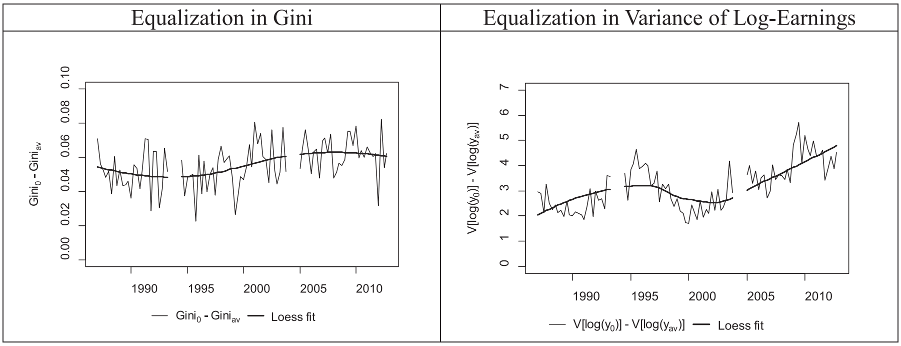

Figure 4. Equalization Brought About by Panel Earnings Changes
Source: Authors' illustration based on ENEU/ENOE data.

$$\ln y = W\gamma + u. \tag{6}$$

Fields (2003) shows that the contribution of a regressor w_k to the variance of logarithms equals

$$\gamma_k cov(w_k, \ln y), \tag{7}$$

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

Table 2 shows the result of applying this decomposition to our Mexican data. In particular, we pooled data from all the available panels into two samples: one including all workers participating in the labor force (irrespective of whether they are employed or not), and another one including only those individuals who remained employed over the five surveyed quarters.

The list of regressors included in equation (6) 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 panel-specific dummies are included as well. In the regressions of average earnings, taking the average across periods of the employment dummy variables (unemployment, sector, industry and occupation) means that we use as independent variables the fraction of time spent in each state by each worker. 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.[7]

The results for the full sample of labor force participants are included in the first two columns of Table 2. There, we observe that the individual's employment/unemployment status is by far the greatest contributor to inequality of initial

[7]The underlying regressions that were used to generate this decomposition are reported in Table A-2 in the Appendix.

TABLE 2

Accounting for Levels of Single-Period and Log-Average 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
$V(\ln\ y)$	4.30 [100]	1.11 [100]	2.26 [100]	1.02 [100]
Gender	0.05 [1.2]	0.035 [3.2]	0.065 [2.9]	0.04 [3.9]
Age	0.039 [0.9]	0.01 [0.9]	0.025 [1.1]	0.011 [1.1]
Yrs. of Schooling	0.057 [1.3]	0.058 [5.2]	0.072 [3.2]	0.069 [6.8]
Unemployment	2.22 [51.7]	0.088 [7.9]		
Sector of Employment	0.138 [3.2]	0.083 [7.4]	0.225 [9.9]	0.096 [9.5]
Occupation	0.083 [1.9]	0.048 [4.4]	0.056 [2.5]	0.043 [4.2]
Industry	0.005 [0.1]	0.017 [1.6]	0.063 [2.8]	0.023 [2.2]
City dummies	0.014 [0.3]	0.01 [0.9]	0.015 [0.6]	0.011 [1.1]
Panel-specific dummies	0.011 [0.3]	0.01 [0.9]	0.012 [0.5]	0.009 [0.9]
Residuals	1.678 [39.1]	0.749 [67.5]	1.731 [76.5]	0.712 [70.2]

All earnings measures are in natural logarithms.
Source: Authors' calculation based on ENEU/ENOE data.

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 percent of inequality. After that, occupation, years of schooling, gender and age each contribute between 1 and 2 percent to the level of variance of log-earnings. Finally, around 40 percent of the variance of log-earnings remains unexplained by the preceding observable characteristics.

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 total variance remains unaccounted for by observables, which only explain little more than 30 percent of the variance of log-average earnings. Among the observables, the employment/unemployment status and sector of employment still account for the greatest share of variation, followed by schooling and occupation dummies, which account for about 5 percent of the variation each. The significant drop in the explanatory power of unemployment in the average earnings equation reflects the fact 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 2 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 almost 10 percent in each case). The years of schooling accounts for between 3 and 7 percent of inequality, with a larger impact on the dispersion of average earnings. In both cases, more than 70 percent of the variation in earnings is unaccounted for by our observable variables.

It is interesting to note that both in the model with all the workers (columns 1 and 2) and the one with full-year employed workers only (columns 3 and 4), the years of schooling account for a greater share of the inequality in average earnings, than of inequality of initial earnings. This indicates that schooling is a better predictor of dispersion of longer-term earnings than of single-period earnings.

Before closing this subsection it is important to remark that the large share of the variation in earnings (initial or average) attributed to the residuals is a natural consequence of the fact that this decomposition method is built using a standard log-earnings regression like the ones commonly used in the labor economics literature. In particular, the fraction of inequality accounted for by the residuals equals, by construction, one minus R-square for the specific log-earnings regression. Since in this kind of "Mincer-like" regressions one rarely sees R-squares greater than 30–35 percent, it then follows that the largest portion of inequality remains unaccounted for by observables. In other words, unlike other decompositions in the inequality literature that fully account for inequality (according to income source, or a subgroup decompositions), this regression-based decomposition leaves a term that cannot be accounted for by the observables within the underlying econometric estimations. However, the advantage of this method is that the contribution obtained for each observable factor is one that holds all other observable factors constant, the same as in a standard linear regression.

3.3.2.Accounting for Equalizing Panel Earnings Changes

So far the previous decomposition accounted for the levels of both single-period and average log-earnings. However, we can also use this method to explore what factors account for our equalization measure EQ in equation (5).

In performing the accounting of the gap in (5) 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 the 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, i.e.

$$\ln y_c = W_a \gamma_0. \tag{8}$$

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_c^2 the variance of the counterfactual log-earnings in (8). Finally, denote by σ_{r0}^2 and σ_{ra}^2 the residual variance of initial and average log-earnings, respectively.[8] Then, we can decompose the gap $EQ = V(ln\, y_0) - V(ln\, y_a)$ as

$$EQ = \left(\sigma_{w0}^2 - \sigma_c^2\right) + \left(\sigma_c^2 - \sigma_{wa}^2\right) + \left(\sigma_{r0}^2 - \sigma_{ra}^2\right). \tag{9}$$

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, $\left(\sigma_c^2 - \sigma_{wa}^2\right)$, 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 the two models. For any of these terms a negative value would mean a dis-equalization 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 (9).

This decomposition 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 the application of this decomposition to analyze the equalization of average earnings relative to initial earnings due to panel earnings changes, rather than the changes in inequality between two anonymous distributions. The full derivation of this decomposition is included in the Appendix of the paper. This decomposition is presented for the Mexican data in Table 3. Several interesting findings arise from this exercise.

Looking at the sample with all labor force participants, we observe that the largest contribution to equalizing earnings over time comes from changes in the employment status of the workers (44.3 percent of the equalization), and from changes in the coefficient associated to this employment status (22.6 percent of the equalization). To understand the reason behind these large effects notice first that in the single-period earnings regressions, the unemployment status variable is a very strong predictor of single-period earnings. Similarly, in the average earnings regression, knowing the percentage of periods in which the individual is unemployed is also strong predictor of his average earnings. However, average earnings become distributed more equally than initial earnings because the burden of unemployment is shared across different individuals in the population. To wit, out of those individuals that started in unemployment in the first period, less than 5 percent remained unemployed over the four subsequent quarters. This explains the equalization reported in the first column of the table. The remaining contribution to equalization reported in the second column (22.6 percent) occurs because the cost of unemployment (i.e. the earnings lost by being unemployed) varies over time.

All other observable factors play a negligible role in this equalization, and about 30 percent 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 less than 20 percent of the equalization can be accounted for

[8]In other words, $V(\ln\ y_0) = \sigma_{w0}^2 + \sigma_{r0}^2$, and $V(\ln\ y_a) = \sigma_{wa}^2 + \sigma_{ra}^2$.

TABLE 3

Equalization of Average Earnings Relative to Initial Earnings Due to Panel Earnings Changes. Percentage Shares of V(ln y_0) – V(ln y_A) are Reported in Square Brackets

	All Workers		Employed Full-year	
$V(\ln\ y_0)-V(\ln\ y_a)$	3.19	[100]	1.25	[100]
	Chars	Coeff	Chars	Coeff
Gender	0.003	0.012	0.001	0.024
	[0.1]	[0.4]	[0.1]	[2.0]
Age	0.008	0.02	0.004	0.01
	[0.3]	[0.6]	[0.3]	[0.8]
Yrs. of Schooling	−0.002	0.001	−0.001	0.004
	[−0.1]	[0.02]	[−0.1]	[0.3]
Unemployment	1.41	0.721		
	[44.3]	[22.6]		
Sector of Employment	−0.001	0.057	0.067	0.061
	[−0.04]	[1.8]	[5.4]	[4.9]
Occupation	0.023	0.012	0.007	0.006
	[0.7]	[0.4]	[0.5]	[0.5]
Industry	−0.023	0.01	0.009	0.031
	[−0.7]	[0.3]	[0.7]	[2.5]
City dummies	−2.0E-05	0.004	−0.0001	0.004
	[−0.001]	[0.1]	[−0.01]	[0.3]
Panel-specific dummies	0.002	−0.0005	0.0003	0.002
	[0.1]	[−0.02]	[0.02]	[0.2]
Residuals	0.93		1.019	
	[29.2]		[81.6]	

ln y_0 denotes initial log-earnings, ln y_a denotes average log earnings.
Char and Coeff are the effects associated to changes in characteristics and coefficients, respectively.
Source: Authors' calculation based on ENEU/ENOE data.

by observable factors, and almost three quarters of this amount is attributable to changes in the sector of employment (formal/informal/self-employed), and industry, and their respective coefficients.

In modern industrial economies, jobs are simultaneously being created and destroyed (Davis and Haltiwanger, 1999), and Mexico is no exception (Kaplan *et al.*, 2007). Job creation and job destruction generate a constant flow of workers moving from one sector/industry/occupation to another. Our results indicate that, at least in the Mexican case, such dynamics contribute to the equalization of longer-term earnings relative to single-period earnings. It would be interesting to explore whether this finding also occurs in other countries with different economic and institutional structures.

4. 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 inequality can be reconciled theoretically with convergent income changes. This seemingly counterintuitive combination is commonly observed in the data. We show that this combination can arise depending on the sizes of the income changes, where the income changes occur in the distribution, and the link between initial and final incomes. One specific version of the relationship can be found in equation (4) in the text.

Our theoretical discussion of possibilities was empirically illustrated using a panel survey with 96 short-lived panels, each of which tracks the earnings of workers for five quarters in urban Mexico. In the empirical analysis we observed that while earnings inequality sometime rises and sometimes falls, earnings changes in Mexico are almost 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. In general, though, most people tend to experience small to moderate convergent 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 stable measure of earnings than 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 3.3 of the paper.

In that section, we showed that individual earnings averaged over 5 quarters are more equally distributed than earnings in any single quarter. This can occur because transitory shocks get averaged out, although it could also reflect the effects of measurement error in earnings. 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 inequality as measured by the variance of log-earnings. Permanent characteristics like gender and years of schooling only account for a small fraction of the observed dispersion.

Turning to the factors accounting 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 labor force participants. In the sample of full-year employed workers, sector of employment variables account for most of the equalization explained by observables, but 80 percent 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 analyze changes in the distribution of many economic variables of interest including earnings, income, wealth, consumption, etc. These analyses could be conducted in other countries and in other economic contexts. These methods could also be used to analyze longer panels. One could use these methods to analyze the impacts of labor market policies on economic inequality. We leave such explorations to future work.

References

Atkinson, A. B., F. Bourguignon, and C. Morrisson, *Empirical Studies of Earnings Mobility*, Vol. 52, Fundamentals of Pure and Applied Economics, Harwood Academic Publishers, Amsterdam, 1992.

Campos-Vázquez, R. M., "Efectos de Los Ingresos No Reportados En El Nivel Y Tendencia de La Pobreza Laboral En México," *Ensayos. Revista de Economía*, XXXII, 23–54, 2013.
Cowell, F. A., *Measuring Inequality*, 3rd ed., Oxford University Press, Oxford 2011.
Davis, S. J. and J. Haltiwanger, "Gross Job Flows," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, 2711–805, Elsevier, Amsterdam, 1999.
Dragoset, L. M. and G. S. Fields, "U.S. Earnings Mobility: Comparing Survey-Based and Administrative-Based Estimates," Unpublished manuscript, 2008.
Duval-Hernández, R., "Dynamics of Labor Market Earnings and Sector of Employment in Urban Mexico, 1987–2002," Cornell University, Ithaca, NY, 2006.
Duval-Hernández, R., G. S. Fields, and G. H. Jakubson, "Panel Income Changes and Changing Relative Income Inequality," Unpublished manuscript, 2016.
Fields, G. S., "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, 2003.
———, "Does Income Mobility Equalize Longer-Term Incomes? New Measures of an Old Concept," *Journal of Economic Inequality*, 8, 409–27, 2010a.
———, "What We Know (and Want to Know) about Earnings Mobility in Developing Countries," Unpublished manuscript, 2010b.
Fields, G. S., R. Duval-Hernández, S. Freije, and M. L. Sánchez Puerta, "Earnings Mobility, Inequality, and Economic Growth in Argentina, Mexico, and Venezuela," *Journal of Economic Inequality*, 13, 103–28, 2015.
Gottschalk, P. and M. Huynh, "Are Earnings Inequality and Mobility Overstated? The Impact of Nonclassical Measurement Error," *Review of Economics and Statistics*, 92, 302–15, 2010.
Juhn, C., K. M. Murphy, and B. Pierce, "Wage Inequality and the Rise in Returns to Skill," *The Journal of Political Economy*, 101, 410–42, 1993.
Kaplan, D. S., G. Martinez Gonzalez, and R. Robertson, "Mexican Employment Dynamics: Evidence from Matched Firm-Worker Data," 4433, World Bank Policy Research Working Paper, 2007.
Oaxaca, R., "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, 693–709, 1973.
Sen, A., *On Economic Inequality*. Expanded Edition with a Substantial Annexe by James E. Foster and Amartya Sen, Oxford University Press, Oxford, 1997.
Yun, M. S., "Earnings Inequality in USA, 1969–99: Comparing Inequality Using Earnings Equations," *Review of Income and Wealth*, 52, 127–44, 2006.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table A-1. Descriptive Statistics
Table A-2. Log-Earnings Linear Regression Models
Figure A-1. Changes in the Coefficient of Variation, Economy-Wide Earnings Growth, and Correlation Coefficient