Supplement to “Inequality, income dynamics, and worker transitions: The case of Mexico”

(*Quantitative Economics*, Vol. 13, No. 4, November 2022, 1669–1705)

Daniela Puggioni
Dirección General de Investigación Económica, Banco de México

Mariana Calderón
Dirección General de Investigación Económica, Banco de México

Alfonso Cebreros Zurita
Dirección General de Investigación Económica, Banco de México

León Fernández Bujanda
Dirección General de Investigación Económica, Banco de México

José Antonio Inguanzo González
Department of Economics, U.C. Los Angeles

David Jaume
Dirección General de Estabilidad Financiera, Banco de México

In Appendix A, we provide additional information regarding the structure and the specific features of the administrative data used to carry out the descriptive analysis in Section 3 of the paper and report relevant summary statistics of our master sample that should facilitate the comparison of the results across countries. Appendix B contains additional results that complement those presented in Section 3 of the paper. We offer more details regarding the comparison between administrative and survey data discussed in Section 4.1 of the paper in Appendix C where we also present additional details based exclusively on survey data. Appendix D includes additional information and results regarding worker transitions from and back into formal employment.

Appendix A: IMSS data and descriptive statistics from the master sample

The Mexican administrative data, the IMSS data, are available on a monthly basis from January 2005 to December 2019 and cover, approximately, between 13 million workers

Daniela Puggioni: dpuggionih@banxico.org.mx
Mariana Calderón: mcalderon@banxico.org.mx
Alfonso Cebreros Zurita: carlos.cebreros@banxico.org.mx
León Fernández Bujanda: lfernandezb@banxico.org.mx
José Antonio Inguanzo González: joseinguanzo@ucla.edu
David Jaume: djaumep@banxico.org.mx

Available at http://qeconomics.org. https://doi.org/10.3982/QE1863

at the start of the sample period and 20 million workers toward the end. The information available for each worker is:[1]

- ◇ Social Security Number (SSN)
- ◇ Unique population registry* (CURP for its acronym in Spanish)
- ◇ Gender
- ◇ Type of employment (permanent vs. temporary contract)[2]
- ◇ Daily (taxable) wage
- ◇ Employer id
- ◇ Firm tax id* (RFC for its acronym in Spanish)
- ◇ Sector of economic activity
- ◇ Geographic location of employer (county where the employer registers the employees with IMSS)

Although not directly provided by IMSS, the worker's SSN provides enough information to infer the year of birth (age) and the year of first enrollment in social security.[3]

While our social security data contain sufficient information to characterize the patterns of income dynamics and inequality for Mexican workers, there are a few issues regarding their limitations that are relevant for understanding how they may compare to or differ from administrative records and/or employer–employee matched data in other countries. First, the IMSS data do not contain any information on workers employed in the informal sector which means, as already mentioned, that they miss a very substantial fraction of the Mexican labor force. Second, two additional issues are worthy of mention:

a. Employer id versus Firm id. In the IMSS data, the employer id does not correspond to the firm id as it is usually the case in employer–employee matched data sets. The Mexican social security system allows firms to have multiple "registros patronales" (i.e., employer ids) that are used to register their workers with IMSS. The same firm could use multiple employer ids for several reasons such as operating multiple plants, or employing groups of workers with different risk profiles and there is no official source of information providing a concordance between employer ids and the firms these belong to.[4] The variable firm id in the data corresponds to the id

[1]These fields of information correspond to those that IMSS has agreed to share with the General Directorate of Economic Research at Banco de México. The fields identified with asterisks are only available from November 2018 onward.

[2]Temporary contracts are those that are specified with start and end dates, while permanent contracts do not include a prespecified end date.

[3]For some of the observations, the age variable (inferred from the SSN) corresponds to an age that cannot possibly be correct. These observations represent a negligible fraction of monthly observations and are eliminated once the age restrictions are applied for constructing the master sample.

[4]In Mexico, social security contributions are paid based on the risk profile of the worker's occupation.

with which each firm is registered in the the "Registro Federal de Contribuyentes" (RFC), a tax identity code assigned by the Servicio de Administración Tributaria, the Mexican tax authority. This code is used by both firms and individuals engaging in economic activities subject to taxes. Analogous to the case of the employer id, firms may legally register multiple RFCs and there is no information regarding which RFCs belong to the same firm.[5]

b. Demographics. Mexican social security data do not provide information on a worker's educational attainment, occupation of employment, or foreign-born status, nor allows for identifying households (neither partners, parents, nor children of a worker). IMSS data cannot be linked with other sources of information, such as household surveys, that contain some of these demographic characteristics.

We now turn to the specifics of the sample construction and sample statistics. Original IMSS records are collapsed to a yearly frequency by summing all the wage observations for a given worker within the year.[6] For the period 2005–2019, this results in over 315 million worker-year pairs with between 17 and 26 million observations per year for workers aged 14–75 years old. Imposing age restrictions on the sample to only include workers aged 25–55, which is the relevant group for all the results of Section 3 of the paper, between 23 and 26% of yearly observations and 24% of the observations in the whole sample are lost. Excluded observations mainly consist of workers aged 24 and below (see Figure A.1).

Regarding the composition by gender, the age-restricted sample is consistent with the composition of the original data: on average, throughout the sample period, 63% of

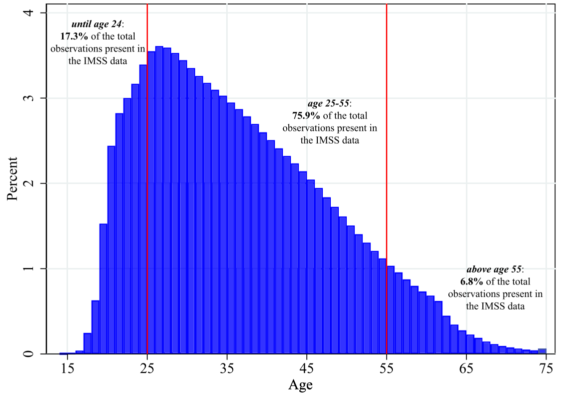


FIGURE A.1. Age distribution in the master sample. *Note*: Based on authors' calculations with data from IMSS. The distribution is calculated over the entire sample, without age restrictions, consisting of 315 million worker-year pairs.

[5]For example, it could be the case that a firm uses one RFC for its taxable domestic operations and another RFC for its foreign sector operations. But there is no information on how many RFCs each firm possesses and how it uses them.

[6]At this stage, the only observations that are dropped from the original records are those with a missing value for wage. These observations represent a negligible share of monthly observations.

TABLE A.1. Gender and age composition of the age-restricted sample.

	By Gender		By Age Group in % Share		
Year	Men	Women	[25–29]	[30–44]	[45–55]
2005	65.2	34.8	25.1	56.3	18.6
2010	63.6	36.4	23.3	55.8	20.9
2015	63.0	37.0	22.9	54.1	23.0
2019	61.3	38.7	22.6	52.4	25.0

Note: Based on authors' calculations with IMSS data.

observations are men, with the share of women rising steadily from 35% in 2005 to 39% in 2019. This implies that the absolute number of men grew by 50% between 2005 and 2019, while the absolute number of women grew by 76% during the same period.[7]

Within the group of workers in question (25–55 years of age), the bulk of the observations are concentrated among workers aged 30 to 44 that, on average, command 55% of yearly observations. By splitting the observations into three age groups, we see that there is a significant change in the distribution of yearly observations across these groups, with a noticeable increase in the participation of the eldest workers (45–55), at the expense of the participation of both the youngest workers (25–29) and workers aged 30 to 44 (see Table A.1).

An important characteristic of our administrative data is the frequent movement of workers in and out of jobs affiliated with social security. These movements represent transitions between formal employment and nonformal employment, with this latter state being either employment in the government, employment in the informal sector, unemployment, or exit from the labor force. We highlight some relevant statistics regarding these transitions. Based on an analysis conducted with a random sample of 4 million workers aged 25 to 55, we document that (see also Table A.2):

TABLE A.2. Distribution of workers by number of active job spells in the formal sector.

	Share %		
N. of spells	Men	Women	All
1	74.5	76.8	75.4
2	19.2	18.5	18.9
3	5.1	4.0	4.7
4	1.1	0.6	0.9
5 or more	0.2	0.1	0.1

Note: Based on authors' calculations with a random sample from IMSS data consisting of 4 million workers aged 25 to 55.

[7] In the original monthly records spanning over January 2005 to December 2019, roughly 64% of observations are men, with the share of women rising steadily throughout the sample, from about 35% at the outset to 37% by the end of the period.

i. 8.8% are present during the entire sample period from 2005 to 2019.

ii. 75.4% have only one active spell that has an average duration of 5.9 years.[8]

iii. 18.9% have only two active spells, the first of these having an average duration of 3.2 years, and the second having an average duration of 3.6 years.[9]

iv. 24.6% of workers have at least one inactive spell during the sample period. Among these, 76.9% have only one inactive spell that has an average duration of 2.9 years.

v. The average duration of the first active spell is 5.2 years, regardless of the total number of active spells.

vi. The average duration of the first inactive spell is 2.7 years, regardless of the total number of inactive spells.

The share of individuals with only one active spell as formal workers, 75.4%, is the result of a combination of: workers that stayed in the database during the whole sample period (8.8%); workers who entered formal employment after 2005 and continuously kept a formal job until 2019 (29.2%); and workers who entered in or after 2005, ended their formal employment relationship before 2019, and did not regain formal employment before or in 2019, that is, they did not come back into the database (37.4%). We refer to the first two groups of workers as "stayers" and to the third group as "leavers." The distribution of workers that have only one formal spell according to these two categories is characterized in Table A.3. Stayers and leavers are almost equally split in the sample and, in general, Mexican workers tend to have a tenuous connection with employment in the formal sector, with this being more evident for women.

Figure A.2 shows the distribution of leavers by age of exit. About 16.9% of these workers leave the IMSS data set when they are 55 years old. That is, they exit either because they reach the upper age limit we imposed to be included in the master sample or because they retire altogether. At the other extreme, a significant share of workers exit formal employment at a young age: 30.1% of all workers that leave are 30 years old or younger.

TABLE A.3. Distribution of workers with only one job spell in the formal sector.

	Share %		
	Men	Women	All
Stayers	51.9	48.2	50.4
Leavers	48.1	51.8	49.6

Note: Based on authors' calculations with a random sample from IMSS data consisting of 4 million workers aged 25 to 55.

[8]Recall that a spell is defined as a sequence of contiguous years in which we observe the worker's income (wage).

[9]94% of workers have at most two active spells.

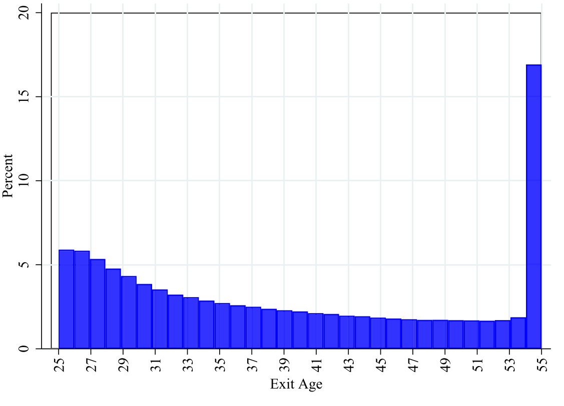


FIGURE A.2. Distribution of leavers by age of exit. *Note*: Based on authors' calculations with data from IMSS.

Appendix B: Additional results for inequality, mobility, and income dynamics

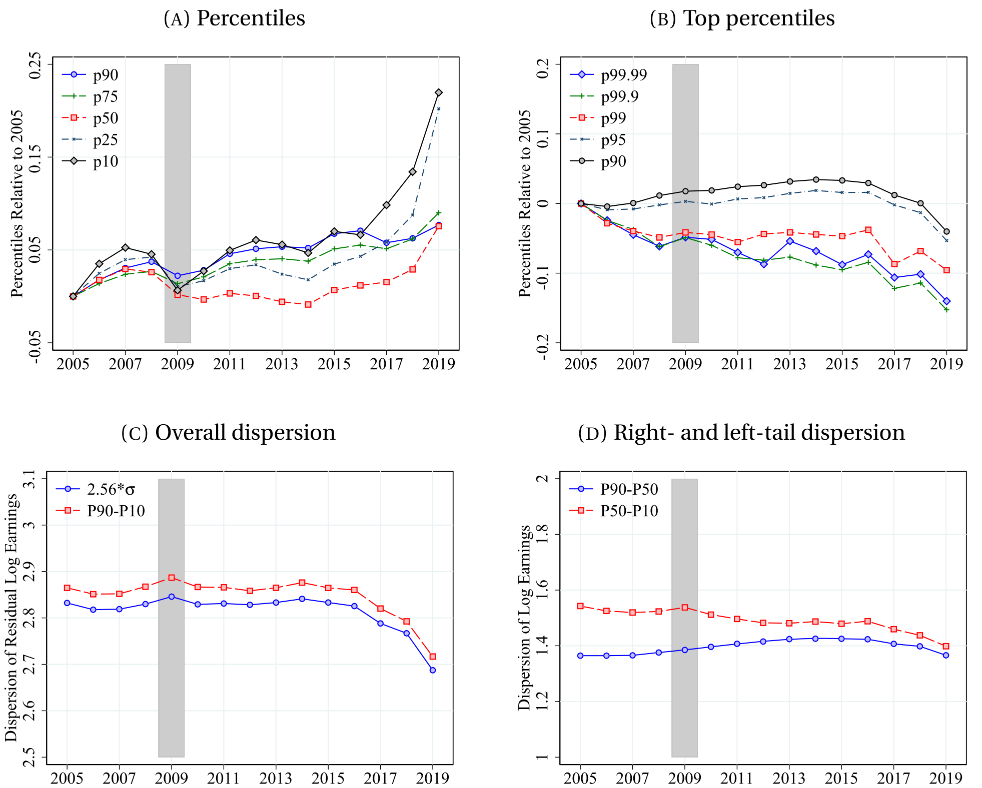


Figure B.1. Distribution of earnings in the population. *Note*: Based on authors' calculations with data from IMSS. Using the CS+TMax sample, this figure plots against time the following statistics of the distribution of log real earnings for the whole population: (A) P10, P25, P50, P75, P90; (B) P90, P99, P99.9, P99.99; (C) P90–P10 and $2.56 * \sigma$ that corresponds to the P90–P10 differential for a Gaussian distribution; (D) P90–P50 and P50–P10. Since the data are top coded, the percentiles above P95 are imputed by fitting a Pareto distribution around the top code. All percentiles are normalized to 0 in 2005, the first available year. Shaded areas are recessions.

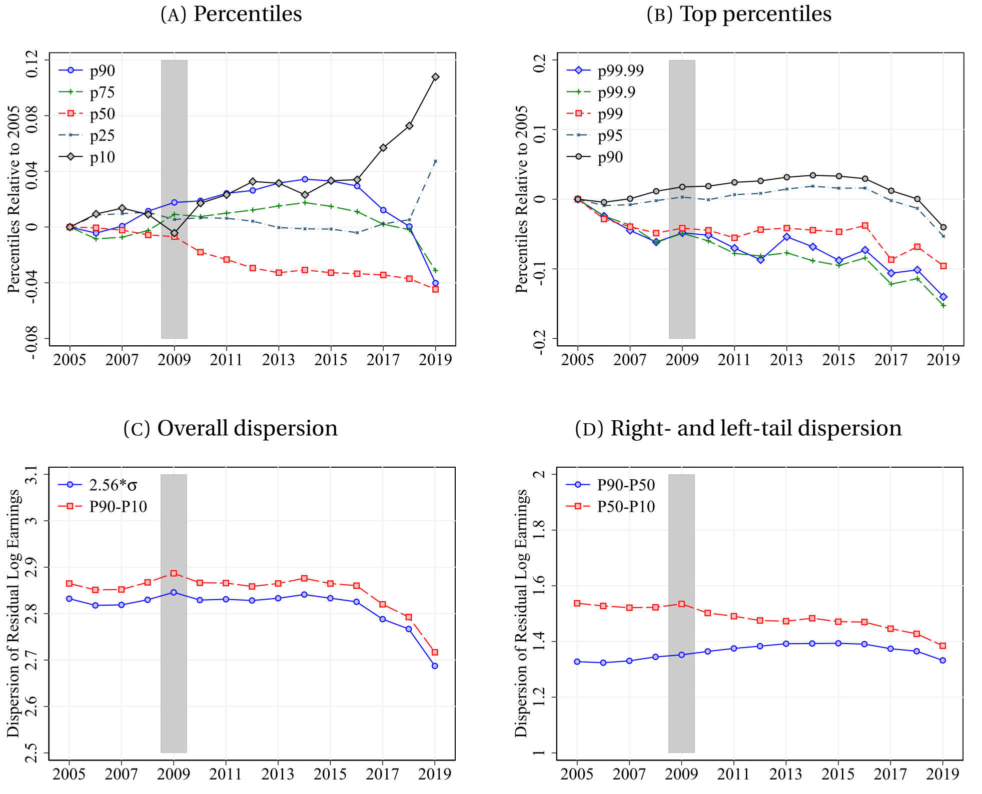


FIGURE B.2. Distribution of residual earnings in the population after controlling for age. *Note*: Based on authors' calculations with data from IMSS. Using the CS+TMax sample, this figure plots against time the following statistics of the distribution of residual earnings for the whole population: (A) P10, P25, P50, P75, P90; (B) P90, P99, P99.9, P99.99; (C) P90–P10 and $2.56 * \sigma$ that corresponds to the P90–P10 differential for a Gaussian distribution; (D) P90–P50 and P50–P10. Residual earnings are obtained regressing log earnings against a full set of age dummies, separately by gender and year, and are computed to avoid trends being affected by individuals being at different stages of their life cycles, or by the business cycle. Shaded areas are recessions.

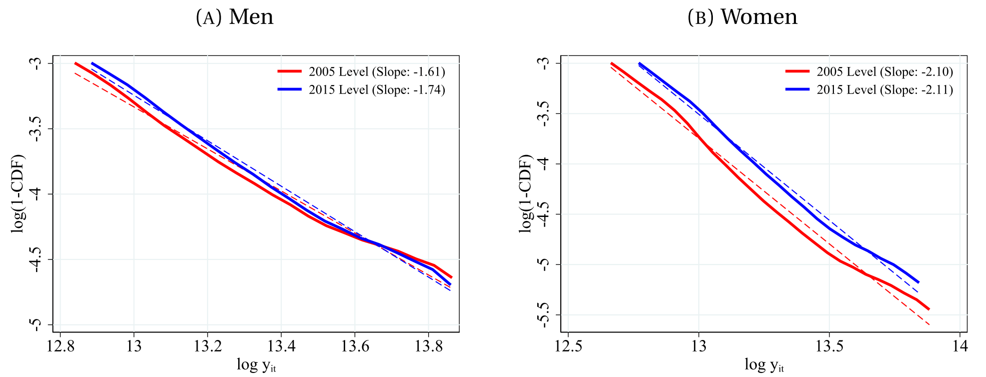


FIGURE B.3. Top income inequality: Pareto tail at top 5%. *Note*: Based on authors' calculations with data from IMSS. Using the CS+TMax sample, this figure plots against log earnings and for selected years in the sample the following variables: (A) Men: log counter cumulative distribution of earnings; (B) Women: log counter cumulative distribution of earnings. The log counter cumulative distribution is calculated as log(1–CDF). The estimated tail index for a power law distribution in the upper tail is reported in parentheses. Since the data are top coded and the top percentiles imputed, the figure reports top income inequality at the top 5% of the distribution of log earnings.

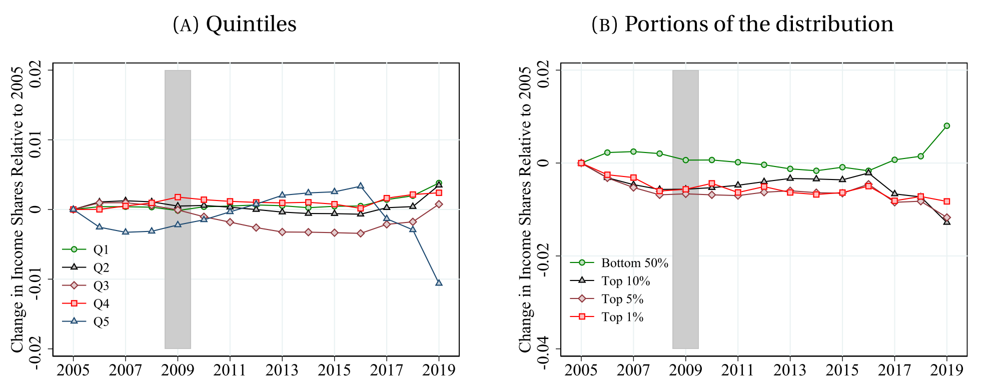


FIGURE B.4. Changes in income share relative to 2005. *Note*: Based on authors' calculations with data from IMSS. Using the CS+TMax sample, this figure plots against time changes in the distribution of income shares relative to 2005 for the whole population: (A): changes in the quintiles of the income shares distribution; (B) changes in selected portions of the income shares distribution. Quintiles are normalized to 0 in 2005, the first available year. Since the data are top coded and the top percentiles imputed, the figure reports changes in income shares only up until the top 1% of the distribution. Shaded areas are recessions.

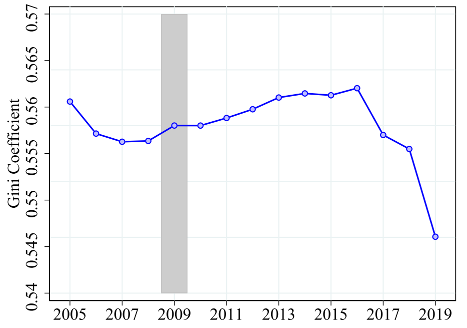


FIGURE B.5. Evolution of the Gini coefficient. *Note*: Based on authors' calculations with data from IMSS. Using the CS+TMax sample, this figure plots against time the Gini coefficient for the whole population. A Gini coefficient equal to 0 expresses perfect equality in the income distribution, while a Gini coefficient equal to 1 expresses maximal inequality. Shaded areas are recessions.

(A) Men (B) Women

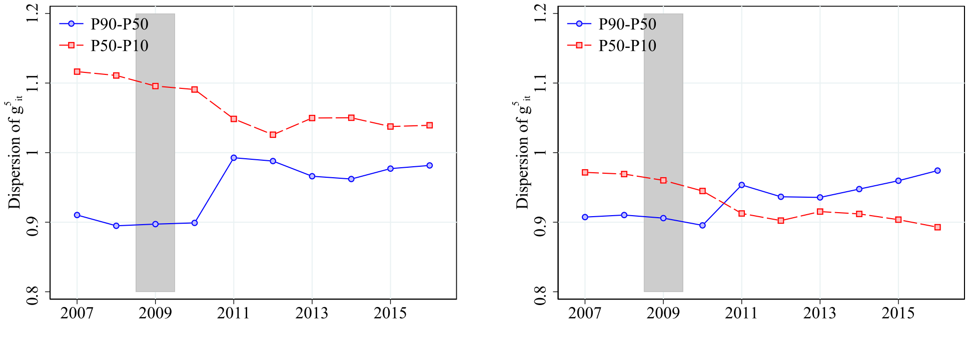
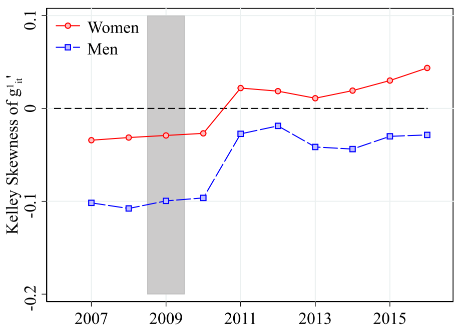


FIGURE B.6. Dispersion of 5-year earnings changes. *Note*: Based on authors' calculations with data from IMSS. Using the LS+TMax sample, this figure plots against time the following measures of top- and bottom-tail dispersion of the distribution of 5-year earnings changes: (A): Men: P90–P50 and P50–P10 differentials; (B) Women: P90–P50 and P50–P10 differentials. Shaded areas are recessions.

(A) Kelley skewness



(B) Excess Crow–Siddiqui kurtosis

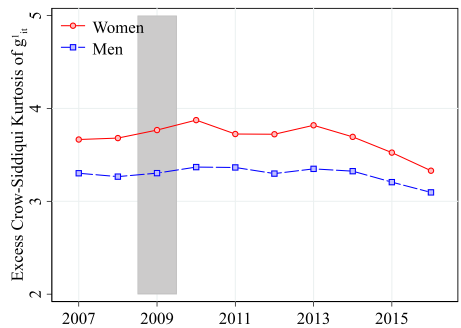


FIGURE B.7. Skewness and kurtosis of 5-year earnings changes. *Note*: Based on authors' calculations with data from IMSS. Using the LS+TMax sample, this figure plots against time the following higher-order moments of the distribution of 5-year earnings changes: (A) Men and Women: Kelley skewness calculated as $\frac{(P_{90}-P_{50})-(P_{50}-P_{10})}{P_{90}-P_{10}}$; (B) Men and Women: Excess Crow–Siddiqui kurtosis calculated as $\frac{P_{97.5}-P_{2.5}}{P_{75}-P_{25}}-2.91$, where the first term is the Crow–Siddiqui measure of kurtosis and 2.91 corresponds to the value of this measure for the Normal distribution. Shaded areas are recessions.

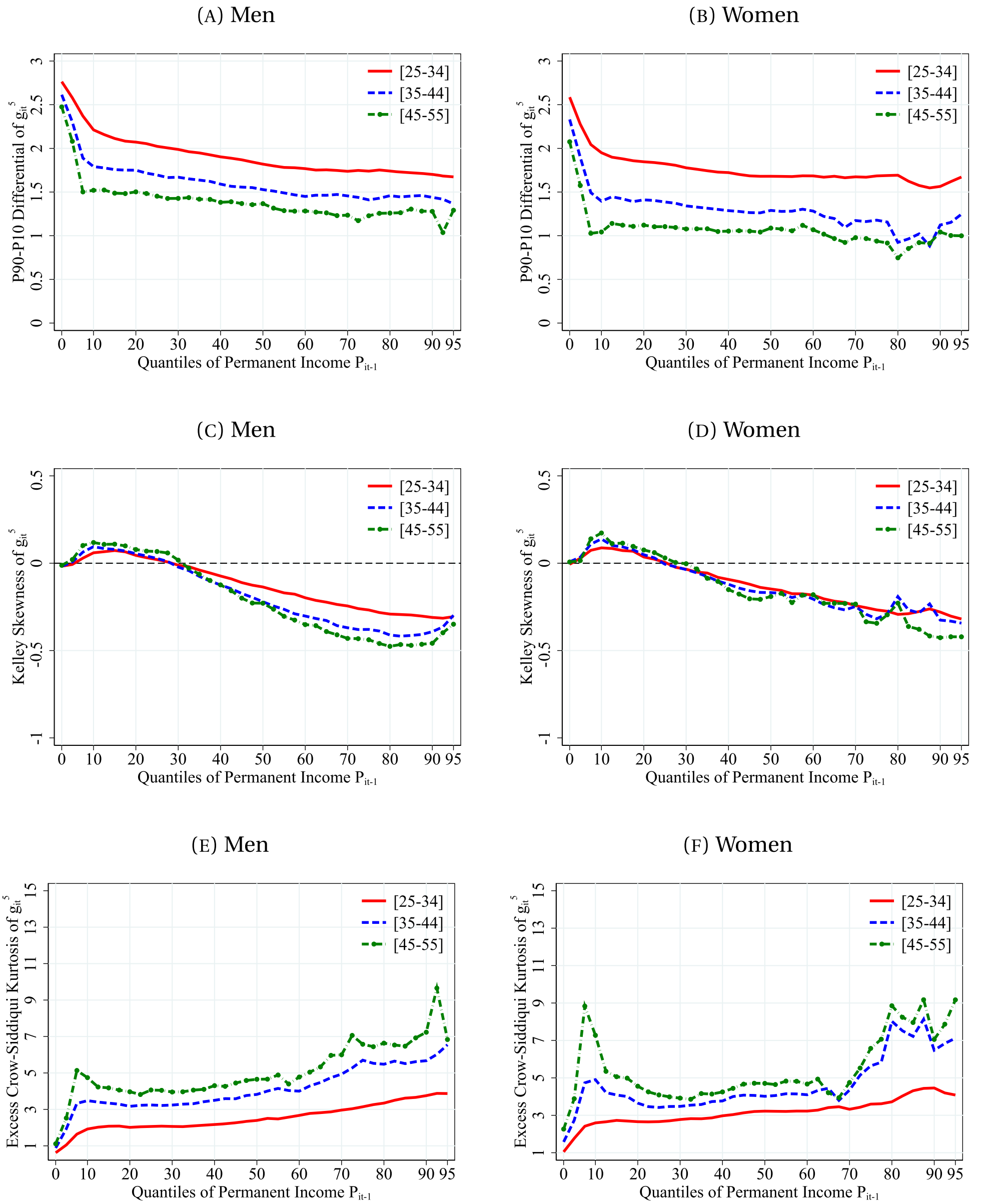


FIGURE B.8. Dispersion, skewness, and kurtosis of 5-year earnings changes. *Note*: Based on authors' calculations with data from IMSS. Using the H+TMax sample, this figure plots against percentiles of the permanent income distribution, and for three different age groups, the following moments of the distribution of 5-year earnings changes: (A) and (B) Men and Women: P90–P10 differential; (C) and (D) Men and Women: Kelley skewness; (E) and (F) Men and Women: Excess Crow–Siddiqui kurtosis. The permanent income is calculated aggregating over a period of 15 years, the maximum number of years available in our sample, from 2005 to 2019. Since the data are top coded, the percentiles of the permanent income distribution are plotted only until P95.

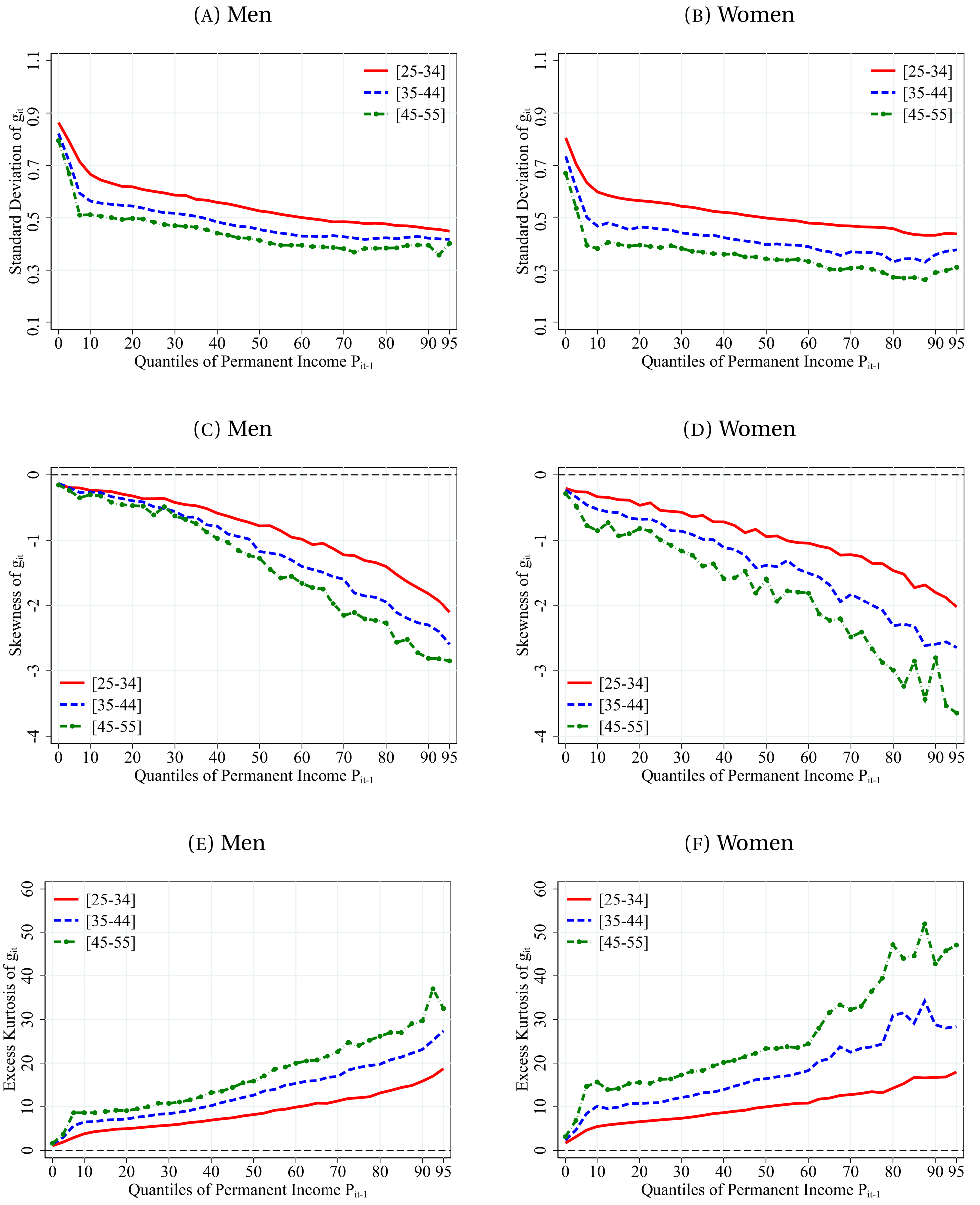


FIGURE B.9. Standardized moments of earnings changes. *Note*: Based on authors' calculations with data from IMSS. Using the H+TMax sample, this figure plots against percentiles of the permanent income distribution, and for three different age groups, the following standardized moments of the distribution of 1-year earnings changes: (A) and (B) Men and Women: Standard deviation; (C) and (D) Men and Women: Coefficient of skewness; (E) and (F) Men and Women: Excess kurtosis. Excess kurtosis calculated as $\gamma - 3$, where γ is the standard measure of kurtosis (i.e., fourth standardized moment) and 3 corresponds to the value of this measure for the Normal distribution. The permanent income is calculated aggregating over a period of 15 years, the maximum number of years available in our sample, from 2005 to 2019. Since the data are top coded, the percentiles of the permanent income distribution are plotted only until P95.

(A) Men (B) Women

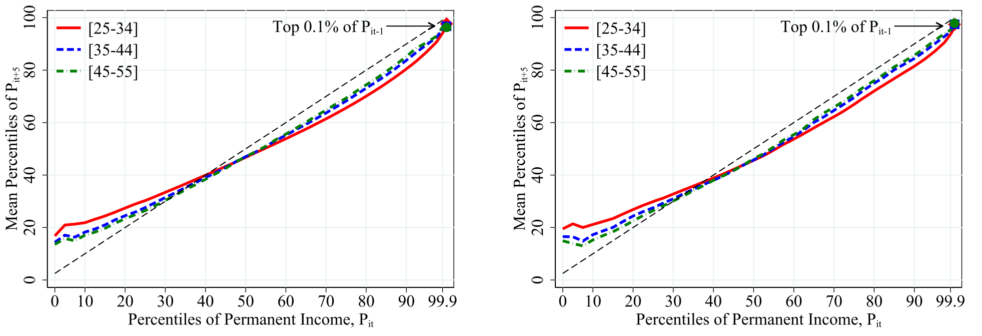


FIGURE B.10. Evolution of 5-year mobility over the life cycle. *Note*: Based on authors' calculations with data from IMSS. The figure shows average rank-rank short-term (5-year) mobility for male (A) and female (B) workers of different ages.

(A) Men (B) Women

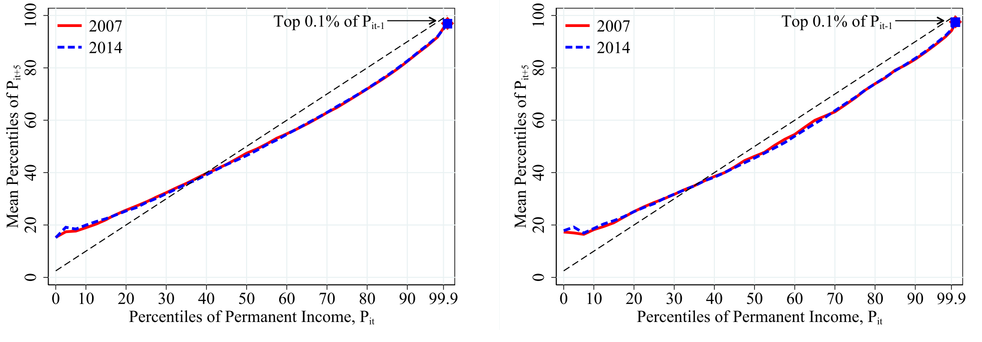


FIGURE B.11. Evolution of 5-year mobility over time. *Note*: Based on authors' calculations with data from IMSS. The figure shows average rank-rank short-term (5-year) mobility for male (A) and female (B) workers in selected years of the sample, 2007 and 2014.

(A) Men (B) Women

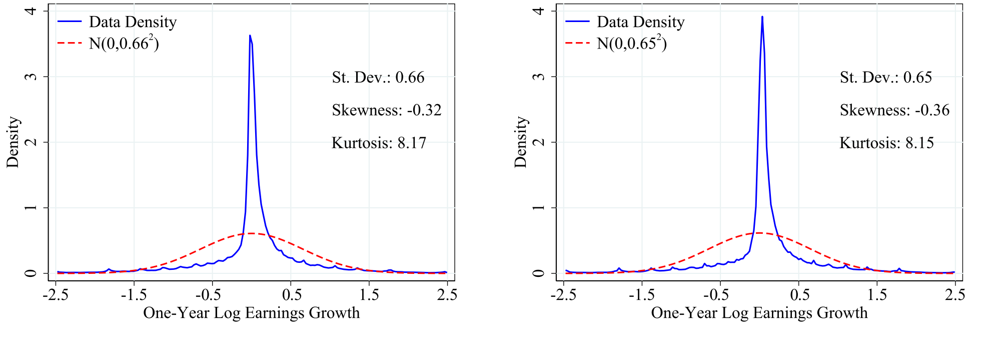


FIGURE B.12. Empirical log-densities of 1-year earnings growth. *Note*: Based on authors' calculations with data from IMSS. The figure shows the log-density of the distribution of 1-year earnings growth for men (A) and women (B) in 2005.

(A) Men (B) Women

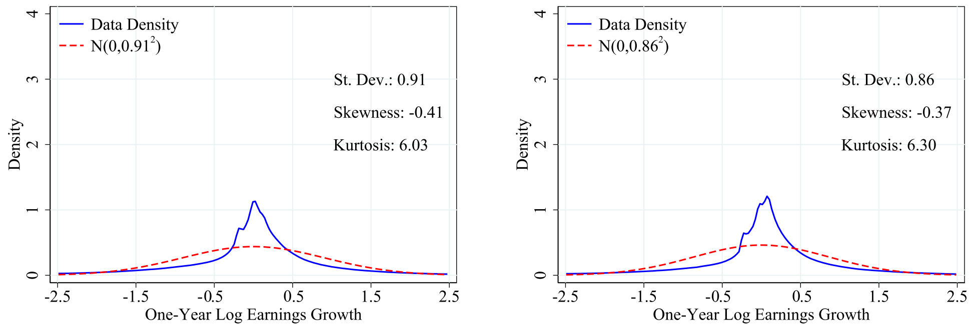


FIGURE B.13. Empirical log-densities of 5-year earnings growth. *Note*: Based on authors' calculations with data from IMSS. The figure shows the log-density of the distribution of 5-year earnings growth for men (A) and women (B) in 2005.

(A) Men (B) Women

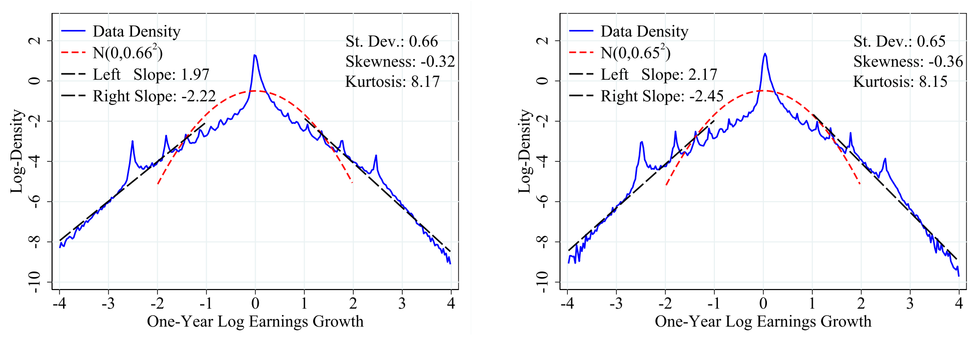


FIGURE B.14. Empirical log-densities of 1-year earnings growth. *Note*: Based on authors' calculations with data from IMSS. The figure shows the log-density of the distribution of 1-year earnings growth for men (A) and women (B) in 2005.

(A) Men (B) Women

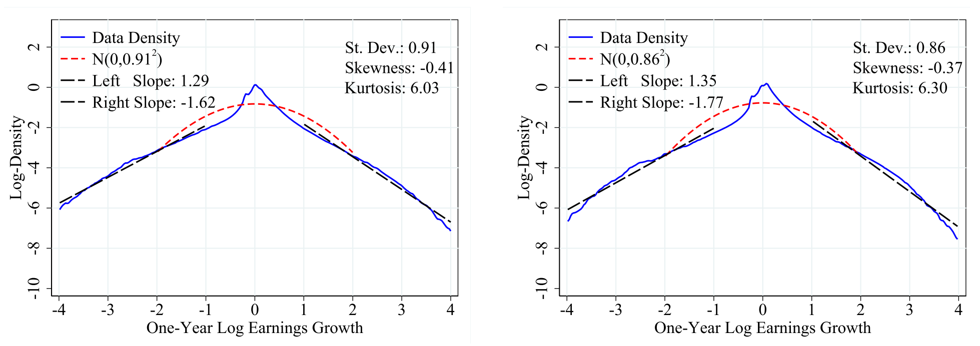


FIGURE B.15. Empirical log-densities of 5-year earnings growth. *Note*: Based on authors' calculations with data from IMSS. The figure shows the log-density of the distribution of 5-year earnings growth for men (A) and women (B) in 2005.

Appendix C: Additional results from the household survey

In this Appendix, we analyze more in depth the issue of nonresponse in the ENOE—the household survey we used—discussed in Section 4.1 of the paper and present additional results based exclusively on survey data for informal workers and the whole workforce.

Nonresponse is a well-known drawback of survey data and we believe that it is likely to be the reason why the characterization of the upper part of the distribution of log earnings differs significantly between our administrative and survey data. To illustrate how this issue affects the ENOE data, we focus on nonresponse limited to a specific question in the survey regarding earnings by considering only individuals who report "invalid" earnings. That is, respondents who declared to have remunerated employment and have worked a positive number of hours, but decided to not provide information about their earnings.

Table C.1 presents the average characteristics of individuals in the survey with both valid and invalid earnings and tests for their differences.[10] The main messages of this table are that: (i) the number of individuals with invalid earnings has been significantly growing over time; (ii) based on observable characteristics, individuals with valid earnings are (statistically) different than individuals with invalid earnings; (iii) the characteristics that differentiate these two groups of individuals have been changing over time. Regarding this last point, Figure C.1 shows that the evolution of nonresponse regarding earnings has not only significantly increased over time, but that it has become more prevalent among highly educated workers who live in cities, are employed in the formal sector, and have a full-time job. As these characteristics are usually associated with higher earnings, we conclude that higher earners are those who more frequently choose to withhold information about their income and that they have increasingly chosen to do so. This points to the fact that the ENOE may be particularly inadequate for providing an accurate picture of the top percentiles of the earnings distribution and that we should be cautious when interpreting the statistics that use information from these percentiles.

Figures C.2–C.4 are analogous to Figures 13–15 presented and discussed in Section 4.1 of the paper. The difference is that here we show the evolution of the percentiles of log earnings and measures of inequality for formal and informal workers, and for the whole pool of workers using exclusively information from the household survey. We find comparable trends indicating that, with the necessary caveats that we have already discussed, the statistics we calculate for formal workers with both administrative and household survey data are relevant and are also relatively in line with those calculated for other categories of workers that are not present in the administrative records.

[10] The information comes from the third quarter of selected years.

Table C.1. Differences in means in observable characteristics.

Characteristics	2005			2012			2019			Difference in Invalid Across Periods		
	Valid	Invalid	Difference	Valid	Invalid	Difference	Valid	Invalid	Difference	2005–2012	2012–2019	2005–2019
Age	36.86	40.25	−3.40	37.80	40.80	−2.99	39.07	41.03	−1.96	−0.54	−0.23	−0.77
			(0.123)			(0.091)			(0.081)	(0.144)	(0.106)	(0.136)
Woman	0.37	0.33	0.04	0.39	0.37	0.02	0.40	0.38	0.02	−0.04	−0.01	−0.05
			(0.004)			(0.003)			(0.002)	(0.005)	(0.004)	(0.005)
Minimum wage stratum	2.96	2.64	0.32	2.84	2.57	0.27	2.40	2.28	0.12	0.07	0.29	0.36
			(0.016)			(0.011)			(0.009)	(0.018)	(0.013)	(0.016)
Formal	0.48	0.56	−0.09	0.44	0.58	−0.15	0.47	0.62	−0.15	−0.02	−0.03	−0.05
			(0.004)			(0.003)			(0.003)	(0.005)	(0.004)	(0.005)
Rural	0.14	0.09	0.04	0.16	0.09	0.07	0.13	0.09	0.05	0.00	0.00	0.00
			(0.003)			(0.002)			(0.002)	(0.003)	(0.002)	(0.003)
Full-time	0.79	0.83	−0.04	0.76	0.83	−0.07	0.77	0.84	−0.07	0.00	−0.01	−0.01
			(0.004)			(0.003)			(0.002)	(0.004)	(0.003)	(0.003)
No schooling completed	0.18	0.12	0.05	0.13	0.08	0.06	0.09	0.05	0.04	0.05	0.03	0.08
			(0.003)			(0.002)			(0.002)	(0.003)	(0.002)	(0.002)
Primary school	0.23	0.17	0.06	0.20	0.14	0.07	0.17	0.11	0.05	0.04	0.02	0.06
			(0.004)			(0.003)			(0.002)	(0.004)	(0.003)	(0.003)
Middle school	0.33	0.29	0.04	0.35	0.29	0.06	0.37	0.28	0.09	0.00	0.01	0.01
			(0.004)			(0.003)			(0.003)	(0.005)	(0.003)	(0.004)
Secondary school	0.10	0.12	−0.02	0.14	0.16	−0.02	0.18	0.19	−0.02	−0.04	−0.03	−0.07
			(0.003)			(0.002)			(0.002)	(0.004)	(0.003)	(0.004)
Postsecondary school	0.16	0.30	−0.14	0.17	0.34	−0.17	0.20	0.37	−0.17	−0.04	−0.03	−0.07
			(0.003)			(0.003)			(0.002)	(0.005)	(0.004)	(0.005)
N. of Observations	132,849	13,977		120,491	29,344		121,850	40,451				

Note: Based on authors' calculations with data from ENOE for the third quarter of 2005, 2012, and 2019. The table reports the means of observable characteristics of individuals whose response to the household survey question about earnings are valid or invalid, and the significance of the difference in these means. The sample includes salaried workers 12 years or older. Standard errors are in parentheses. The table is an updated version of Cuadro 1 in Campos Vázquez (2013)
.

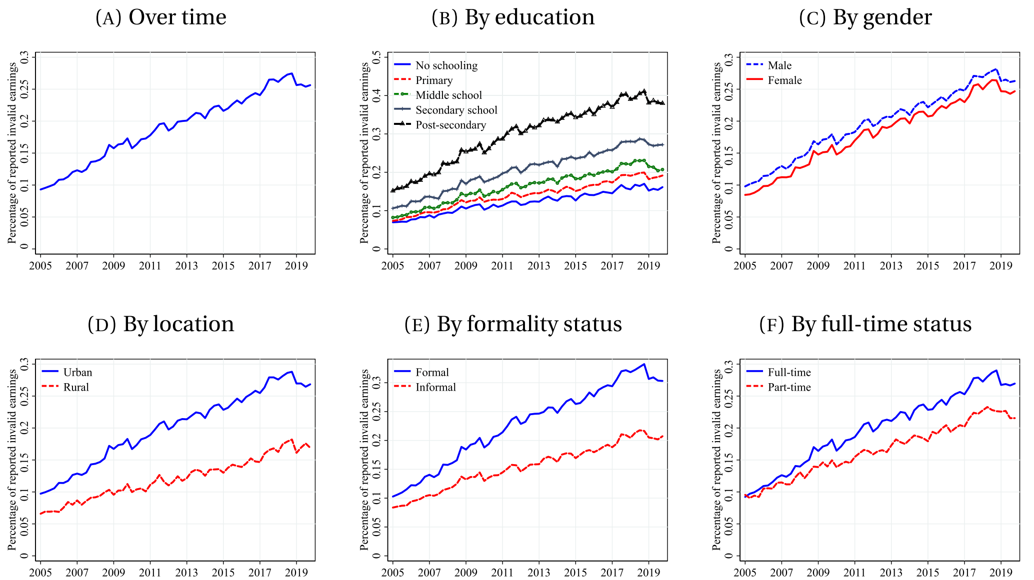
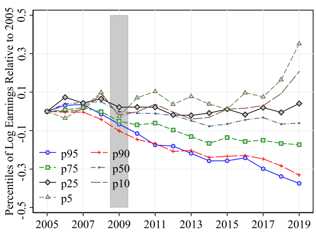
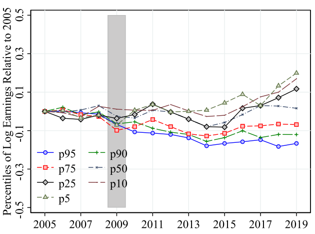


FIGURE C.1. Evolution of the percentage of workers who do not report earnings by sociodemographic groups. *Note*: Based on authors' calculations with data from ENOE. The figure plots the evolution of the percentage of individuals in the household survey with invalid earnings: (A) over time; (B) over time by educational levels; (C) over time by gender; (D) over time by location; (E) over time by formality status; (F) over time by full-time status.

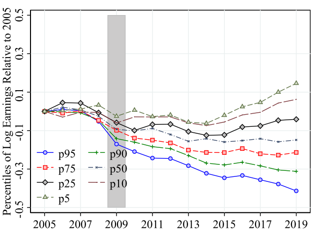
(A) ENOE formal workers



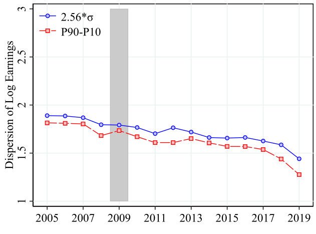
(B) ENOE informal workers



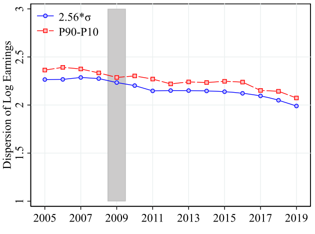
(C) ENOE all workers



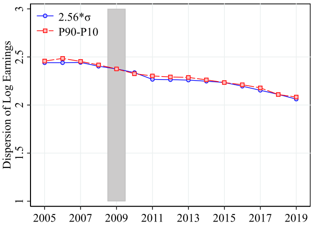
(D) ENOE formal workers



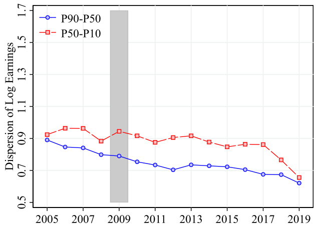
(E) ENOE informal workers



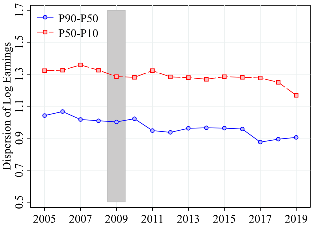
(F) ENOE all workers



(G) ENOE formal workers



(H) ENOE informal workers



(I) ENOE all workers

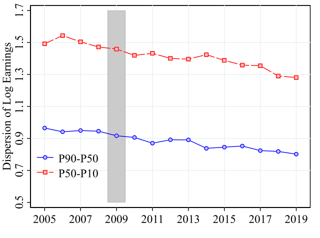


FIGURE C.2. Comparison between subsamples in the survey data: Distribution of log real earnings. *Note*: Based on authors' calculations with data from ENOE. Using a sample from ENOE constructed to match the CS+TMax sample as closely as possible, this figure plots against time the following statistics of the distribution of log earnings: (A) ENOE formal workers: P5, P10, P25, P50, P75, P90, P95; (B) ENOE informal workers: P5, P10, P25, P50, P75, P90, P95; (C) ENOE all workers: P5, P10, P25, P50, P75, P90, P95; (D) ENOE formal workers: P90–P10 and 2.56*σ; (E) ENOE informal workers: P90–P10 and 2.56*σ; (F) ENOE all workers: P90–P10 and 2.56*σ; (G) ENOE formal workers: P90–P50 and P50–P10; (H) ENOE informal workers: P90–P50 and P50–P10; (I) ENOE all workers: P90–P50 and P50–P10. Shaded areas are recessions.

(A) ENOE formal workers (B) ENOE informal workers (C) ENOE all workers

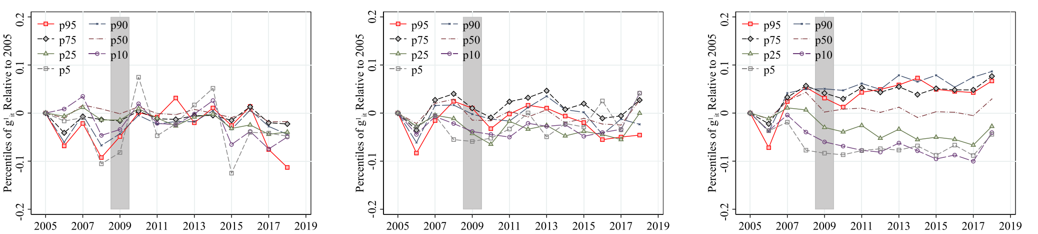


FIGURE C.3. Comparison between subsamples in the survey data: Evolution of the percentiles of the distribution of 1-year log earnings changes. *Note*: Based on authors' calculations with data from ENOE. Using a sample from ENOE constructed to match the CS+TMax sample as closely as possible, this figure plots against time the following statistics of the distribution of 1-year log earnings changes: (A) ENOE formal workers: P5, P10, P25, P50, P75, P90, P95, P99, P99.9; (B) ENOE informal workers: P5, P10, P25, P50, P75, P90, P95; (C) ENOE formal workers: P5, P10, P25, P50, P75, P90, P95, P99, P99.9. All percentiles are normalized to 0 in 2005, the first available year. P99 and P99.9 are omitted because, due to the lack of a sufficient number of observations, they are too noisy to be informative. Shaded areas are recessions.

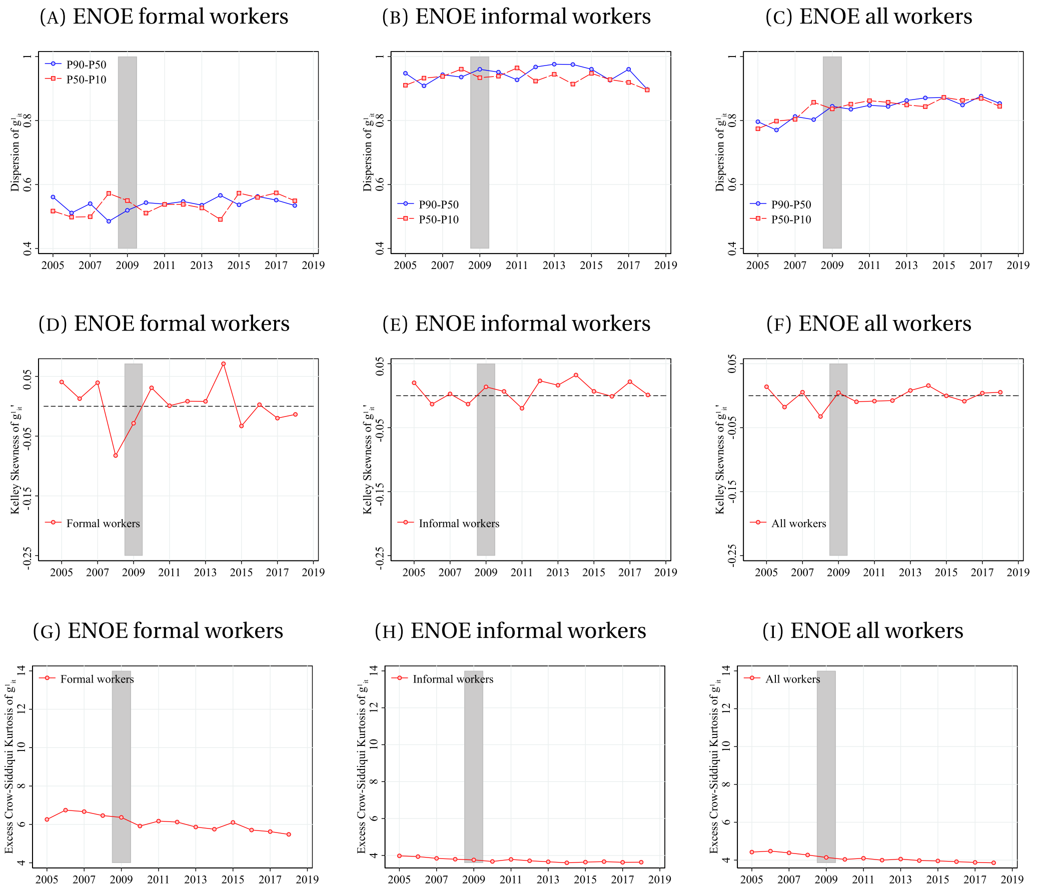


FIGURE C.4. Comparison between subsamples in the survey data: Measures of dispersion, symmetry, and peakedness of the distribution of 1-year log earnings changes. *Note*: Based on authors' calculations with data from ENOE. Using a sample from ENOE constructed to match the CS+TMax sample as closely as possible, this figure plots against time the following statistics of the distribution of 1-year log earnings changes: (A) ENOE formal workers: P90–P50 and P50–P10; (B) ENOE informal workers: P90–P10 and P50–P10; (C) ENOE all workers: P90–P50 and P50–P10; (D) ENOE formal workers: Kelly skewness; (E) ENOE informal workers: Kelly skewness; (F) ENOE all workers: Kelly skewness; (G) ENOE formal workers: Excess Crow–Siddiqui kurtosis; (H) ENOE informal workers: Excess Crow–Siddiqui kurtosis; (I) ENOE all workers: Excess Crow–Siddiqui kurtosis. Shaded areas are recessions.

Appendix D: Additional results for transitions in and out of formal employment

This Appendix includes additional information and results that complement those of Section 4.2 of the paper. We present the regression output for the regressions in equations (4.2) and (4.2) in the main text together with a graphical representation of the results of a wide set of robustness exercises we perform. All these exercises confirm that the main results of the analysis in Section 4.2 are robust to different specifications.

Table D.1. Estimates of wages trajectories for workers who exit and reenter formal employment.

Dependent Variable: Log Wages					
Independent Variables:	Men	Women	Independent variables:	Men	Women
$t=-2$	0.062 (0.007)	0.034 (0.005)			
$t=-1$	0.047 (0.007)	0.032 (0.004)			
$t=-0$	0.000 $(\cdot)$	0.000 $(\cdot)$			
$t=1$	−0.154 (0.011)	−0.146 (0.008)			
$t=2$	−0.081 (0.009)	−0.063 (0.008)			
$t=3$	−0.045 (0.009)	−0.013 (0.008)			
$k=1$	0.000 $(\cdot)$	0.000 $(\cdot)$			
$k=2$	−0.019 (0.007)	−0.035 (0.006)			
$k=3$	−0.030 (0.008)	−0.047 (0.010)			
$k=4$	−0.048 (0.011)	−0.047 (0.011)			
$k=5$	−0.047 (0.014)	−0.060 (0.013)			
$k=6$	−0.040 (0.016)	−0.080 (0.017)			
$k=7$	−0.051 (0.022)	−0.085 (0.021)			
$k=8$	−0.097 (0.021)	−0.105 (0.018)			
$k=9$	−0.008 (0.025)	−0.102 (0.021)			
$t=-2\times k=1$	0.000 $(\cdot)$	0.000 $(\cdot)$	$t=1\times k=1$	0.000 $(\cdot)$	0.000 $(\cdot)$
$t=-2\times k=2$	−0.001 (0.008)	−0.004 (0.006)	$t=1\times k=2$	0.001 (0.008)	0.006 (0.009)
$t=-2\times k=3$	0.003 (0.008)	−0.002 (0.010)	$t=1\times k=3$	−0.011 (0.009)	−0.009 (0.011)
$t=-2\times k=4$	−0.000 (0.012)	−0.008 (0.012)	$t=1\times k=4$	−0.020 (0.013)	−0.028 (0.014)
$t=-2\times k=5$	−0.013 (0.016)	−0.005 (0.013)	$t=1\times k=5$	−0.031 (0.016)	−0.024 (0.015)
$t=-2\times k=6$	−0.008 (0.016)	0.009 (0.011)	$t=1\times k=6$	−0.049 (0.018)	−0.001 (0.021)
$t=-2\times k=7$	0.005 (0.021)	−0.003 (0.024)	$t=1\times k=7$	−0.061 (0.027)	−0.031 (0.023)
$t=-2\times k=8$	0.038 (0.026)	0.030 (0.018)	$t=1\times k=8$	−0.004 (0.027)	0.045 (0.030)
$t=-2\times k=9$	−0.029 (0.029)	0.026 (0.021)	$t=1\times k=9$	−0.070 (0.039)	−0.029 (0.036)
$t=-1\times k=1$	0.000 $(\cdot)$	0.000 $(\cdot)$	$t=2\times k=1$	0.000 $(\cdot)$	0.000 $(\cdot)$
$t=-1\times k=2$	−0.002 (0.007)	0.000 (0.006)	$t=2\times k=2$	−0.005 (0.008)	0.004 (0.009)
$t=-1\times k=3$	−0.002 (0.008)	−0.004 (0.009)	$t=2\times k=3$	−0.013 (0.011)	−0.014 (0.011)
$t=-1\times k=4$	−0.001 (0.011)	−0.001 (0.010)	$t=2\times k=4$	−0.024 (0.012)	−0.028 (0.013)
$t=-1\times k=5$	−0.005 (0.013)	−0.009 (0.013)	$t=2\times k=5$	−0.040 (0.017)	−0.026 (0.014)
$t=-1\times k=6$	−0.000 (0.015)	0.006 (0.015)	$t=2\times k=6$	−0.058 (0.020)	−0.010 (0.021)
$t=-1\times k=7$	0.000 (0.021)	−0.010 (0.021)	$t=2\times k=7$	−0.067 (0.027)	−0.039 (0.023)
$t=-1\times k=8$	0.030 (0.024)	0.018 (0.021)	$t=2\times k=8$	−0.002 (0.033)	0.025 (0.032)
$t=-1\times k=9$	−0.020 (0.028)	0.010 (0.026)	$t=2\times k=9$	−0.079 (0.039)	−0.048 (0.036)

(*Continues*)

TABLE D.1. *Continued.*

Dependent Variable: Log Wages					
Independent Variables:	Men	Women	Independent variables:	Men	Women
$t=0\times k=1$	0.000 ($\cdot$)	0.000 ($\cdot$)	$t=3\times k=1$	0.000 ($\cdot$)	0.000 ($\cdot$)
$t=0\times k=2$	0.000 ($\cdot$)	0.000 ($\cdot$)	$t=3\times k=2$	−0.003 (0.008)	0.00 (0.008)
$t=0\times k=3$	0.000 ($\cdot$)	0.000 ($\cdot$)	$t=3\times k=3$	−0.019 (0.010)	−0.014 (0.010)
$t=0\times k=4$	0.000 ($\cdot$)	0.000 ($\cdot$)	$t=3\times k=4$	−0.016 (0.012)	−0.039 (0.012)
$t=0\times k=5$	0.000 ($\cdot$)	0.000 ($\cdot$)	$t=3\times k=5$	−0.032 (0.018)	−0.020 (0.015)
$t=0\times k=6$	0.000 ($\cdot$)	0.000 ($\cdot$)	$t=3\times k=6$	−0.051 (0.020)	−0.001 (0.022)
$t=0\times k=7$	0.000 ($\cdot$)	0.000 ($\cdot$)	$t=3\times k=7$	−0.083 (0.028)	−0.037 (0.024)
$t=0\times k=8$	0.000 ($\cdot$)	0.000 ($\cdot$)	$t=3\times k=8$	−0.008 (0.032)	0.027 (0.026)
$t=0\times k=9$	0.000 ($\cdot$)	0.000 ($\cdot$)	$t=3\times k=9$	−0.090 (0.027)	−0.025 (0.028)
Constant	8.164 (0.030)	8.230 (0.034)			
N. of Observations	682,248 (Men)	389,856 (Women)			

Note: Based on authors' estimates with data from IMSS. The table reports estimates of the coefficients from equation (4.1). The specification includes sector of economic activity, state, and year fixed effects. Standard errors (in parentheses) are clustered at the worker and sector-year levels.

TABLE D.2. Estimates of wages trajectories: treatment versus control group with a 3-year window.

Dependent Variable: Log Wages		
Independent Variables:	Men	Women
$t=-2$	−0.018 (0.004)	−0.014 (0.003)
$t=-1$	−0.011 (0.003)	−0.007 (0.002)
$t=0$	0.000 ($\cdot$)	0.000 ($\cdot$)
$t=1$	−0.001 (0.006)	−0.020 (0.005)
$t=2$	0.005 (0.007)	0.025 (0.006)
$t=-3$	0.008 (0.008)	0.0300 (0.007)
treated = 1	−0.351 (0.012)	−0.393 (0.013)
$t=-2\times$ treated = 1	0.102 (0.008)	0.078 (0.011)
$t=-1\times$ treated = 1	0.067 (0.006)	0.053 (0.007)
$t=0\times$ treated = 1	0.000 ($\cdot$)	0.000 ($\cdot$)
$t=1\times$ treated = 1	−0.189 (0.016)	−0.222 (0.016)
$t=2\times$ treated = 1	−0.131 (0.016)	−0.154 (0.018)
$t=3\times$ treated = 1	−0.101 (0.016)	−0.118 (0.019)
Constant	8.306 (0.036)	8.463 (0.036)
N. of Observations	1,705,014	1,029,924

Note: Based on authors' estimates with data from IMSS. The table reports estimates of the coefficient $\beta_{\tau}^{\text{treated}}$ from equation (4.2). The specification includes sector of economic activity, state, and year fixed effects. Standard errors (in parentheses) are clustered at the worker and sector-year levels.

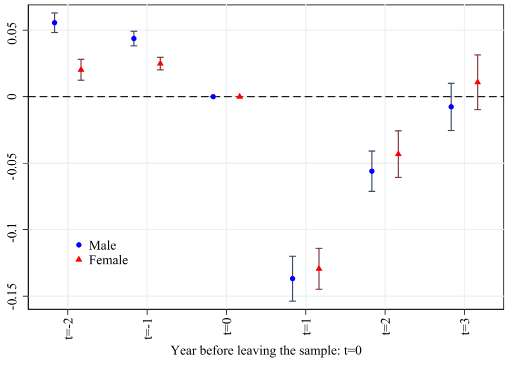


FIGURE D.1. Estimates of wages trajectories (log differences) of workers who exit and reenter formal employment adding worker fixed effects. *Note*: Based on authors' estimates with data from IMSS. The figure plots differences of log wages obtained by estimating equation (4.1) with worker fixed effects using a subsample of workers with only two spells of formal employment and including worker fixed effects in the specification. Markers for men and women are positioned to the left and right, respectively, of each event year t. Standard errors are clustered at the worker and sector-year levels and 95% confidence intervals are plotted together with point estimates.

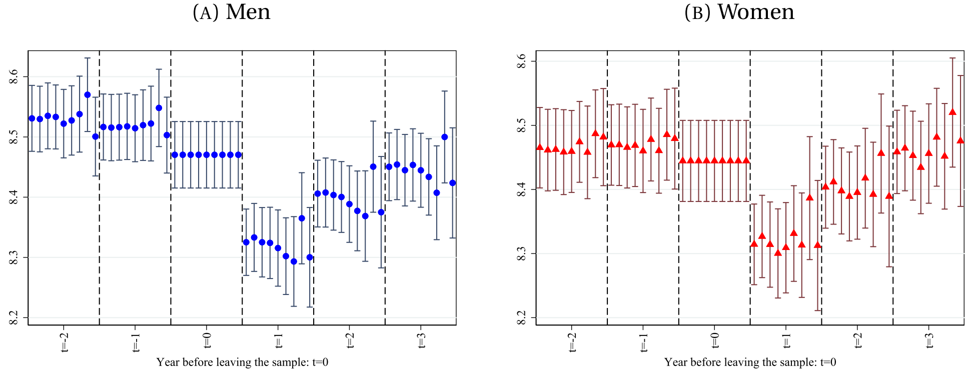


FIGURE D.2. Estimates of wages trajectories (levels) of workers who exit and reenter formal employment including worker fixed effects. *Note*: Based on authors' estimates with data from IMSS. The figure plots the conditional means of log wages using the estimated coefficients from equation (4.1) where worker fixed effects are added. The coefficients β_κ and β_τ^κ are omitted in this estimation as they are collinear with worker fixed effects. Standard errors are clustered at the worker and sector-year levels and 95% confidence intervals are plotted together with point estimates. Standard errors and confidence intervals are obtained with the delta method.

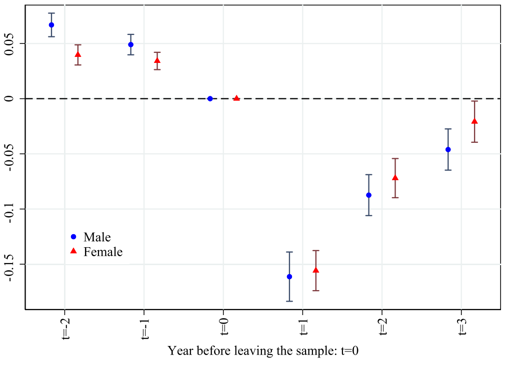


FIGURE D.3. Estimates of wages trajectories (log differences) of workers who exit and reenter formal employment including cohort-year fixed effects. *Note*: Based on authors' estimates with data from IMSS. The figure plots differences of log wages obtained by estimating equation (4.1) using a subsample of workers with only two spells of formal employment and including cohort-year fixed effects. A cohort c is defined as the cohort of workers who turned 18 in year c. The coefficients β_κ and β_τ^κ are omitted in this estimation as they are collinear with cohort-year fixed effects.Markers for men and women are positioned to the left and right, respectively, of each event year t. Standard errors are clustered at the worker and sector-year levels and 95% confidence intervals are plotted together with point estimates.

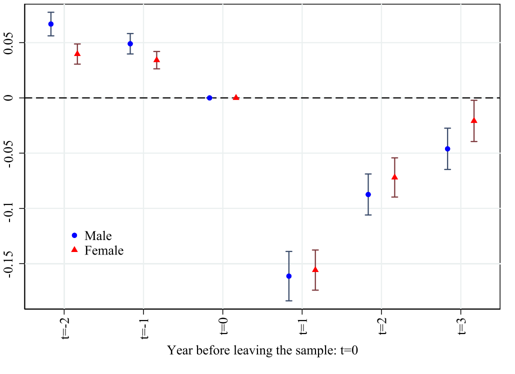


FIGURE D.4. Estimates of wages trajectories (log differences) of workers who exit and reenter formal employment with a 5-year event window. *Note*: Based on authors' estimates with data from IMSS. The figure plots differences of log wages obtained by estimating equation (4.1) using a subsample of workers with only two spells of formal employment and widening the event window to 5 years. Markers for men and women are positioned to the left and right, respectively, of each event year t. Standard errors are clustered at the worker and sector-year levels and 95% confidence intervals are plotted together with point estimates.

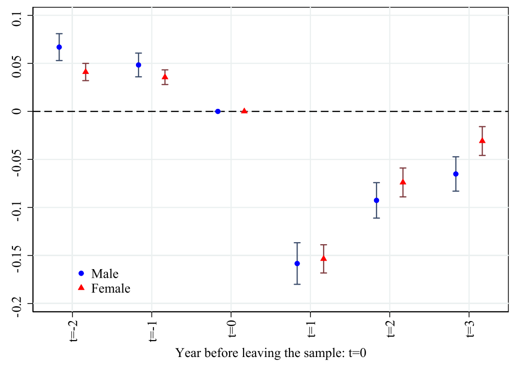


FIGURE D.5. Estimates of wages trajectories (log differences) of workers who exit and reenter formal employment with at least 2 spells of formal employment. *Note*: Based on authors' estimates with data from IMSS. The figure plots differences of log wages obtained by estimating equation (4.1) using a subsample of workers with at least two spells of formal employment. For workers with more than two spells, only the first two are considered. Markers for men and women are positioned to the left and right, respectively, of each event year t. Standard errors are clustered at the worker and sector-year levels and 95% confidence intervals are plotted together with point estimates.

As a final robustness check, we assess whether our baseline results from Section 4.2 could be driven by the preexit and post reentry wage trajectories of workers whose first spell of formal employment came to an end due to the 2009 financial crisis. We address this concern by estimating the following alternative specification:

$$\ln(w_{it}) = \sum_{\tau=-2}^{3} \beta_{\tau}^{\text{crisis}} \mathbb{I}_{\tau} \mathbb{I}_{i,\text{crisis}} + \gamma_g \mathbb{I}_g + \alpha_e + \alpha_s + \alpha_t + \varepsilon_{it}. \tag{D.1}$$

In this case, $\mathbb{I}_{i,\text{crisis}}$ is an indicator variable that equals 1 if the last year we observe the worker in the database before exit is 2008 or 2009. The coefficients $\beta_{\tau}^{\text{crisis}}$ capture the average wage in every year of the event window for workers that exited during the financial crisis as compared to the average wages of those who left in any other year and their estimates are shown in Figure D.6. For both genders, we find that 1 year before exit occurred, workers who left during the financial crisis had, on average, slightly higher wages than those who left in any other year. This difference is statistically significant for men and marginally significant for women. In contrast, the average wage difference among these two groups of workers is not statistically significant upon reentry. Hence, we conclude that the wage patterns documented with our benchmark specification are a general feature of the transitions out and back into formal employment and do not seem to be driven by the specific exit/reentry that occurred during the large shock of the financial crisis.

(A) Men (B) Women

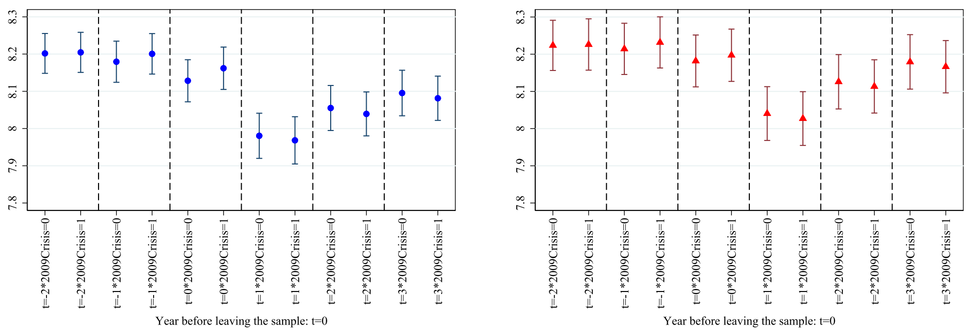


FIGURE D.6. Estimates of wages trajectories of workers who left formal employment during the 2009 financial crisis. *Note*: Based on authors' estimates with data from IMSS. The figure plots the conditional means of log wages computed as $\mathbb{E}[\ln w_{it}|X = X_0, \tau = t, \mathbb{I}_{i,\text{crisis}} = 1] = \hat{\beta}_{\tau}^{\text{crisis}}$ using the estimated coefficients from equation (D.1). Standard errors are clustered at the worker and sector-year levels and 95% confidence intervals are plotted together with point estimates.

References

Campos Vázquez, Raymundo M. (2013), "Efectos de los ingresos no reportados en el nivel y tendencia de la pobreza laboral en México." *Ensayos Revista de Economía*, 32. [17]

Co-editor Giovanni L. Violante handled this manuscript.

Manuscript received 2 March, 2021; final version accepted 9 April, 2022; available online 16 June, 2022.