

# Labor Market Trajectories and Earnings Dynamics in Colombia<sup>\*</sup>

Leonardo Fabio Morales      Sergio Ocampo

Brenda Samaniego de la Parra

## Abstract

This paper analyzes labor earnings dynamics in Colombia using administrative Social Security records for the universe of formal workers from 2009 to 2024. We present stylized facts on earnings dispersion, growth, and volatility following the Global Repository of Income Dynamics framework. We find a modest decline in earnings concentration and stable overall dispersion, patterns comparable to those observed in the United States and Brazil. We show how individuals' histories of attachment to the formal sector shape the levels of earnings inequality and volatility. We identify five distinct groups of individuals based on their formal sector attachment trajectories that explain most of the variation in formal participation. Earnings dispersion is concentrated among groups with low formal attachment, while those with high attachment display substantially lower, and more stable, earnings dispersion and are more likely to work at larger firms. Our findings highlight the central role of long-run formal attachment in shaping earnings dynamics.

*Keywords:* Formal labor market, Earnings dynamics, Informality.

*JEL Codes:* D31, J31, J46.

---

<sup>\*</sup>Morales: Banco de la República; [lmoralzu@banrep.gov.co](mailto:lmoralzu@banrep.gov.co)

Ocampo: University of Western Ontario; [socampod@uwo.ca](mailto:socampod@uwo.ca)

Samaniego de la Parra: University of Notre Dame Keough School of Global Affairs; [bsamanie@nd.edu](mailto:bsamanie@nd.edu)

The opinions contained in this document are the sole responsibility of the authors and do not commit Banco de la República nor its Board of Directors. We are particularly grateful for the excellent research assistance of Felipe Saavedra Torres, David Salgado, and Simón Yarce Arango. We are indebted to the staff at Banco de la República for making access to Colombian administrative data possible and to Serdar Ozkan and Sergio Salgado for preparing the GRID masterfiles used in the preparation of the GRID database. We are thankful to Paulina Restrepo Echavarría for her involvement with the GRID project and her comments on the preparation of this paper and to Andrea Otero and Juan Esteban Carranza for valuable comments.

## 1. Introduction

The study of earnings inequality and earnings dynamics has long recognized that cross-sectional measures provide only a partial view of workers' economic outcomes. This has motivated a growing body of research that uses longitudinal administrative datasets in developing and developed countries to follow workers over long horizons and study how labor market histories shape earnings dynamics (see, for example, [Guvenen, Pistaferri, and Violante 2022b](#)). One critical dimension of workers' labor market trajectories, particularly in developing countries, is their presence in the formal sector. Formal jobs not only pay more but also provide legal protections and entitlements to workers, yet formal employment is often intermittent, with workers transitioning in and out of formal work arrangements. Consequently, understanding patterns of formal sector attachment, that is, how formality fluctuates over individuals' work lives, is essential for understanding the distribution, growth, and volatility of earnings.

In this paper, we use Colombia's administrative records to document and analyze earnings dynamics in the formal sector. These results are part of the *Global Repository of Income Dynamics* (GRID) project.<sup>1</sup> We proceed in two stages. First, we document earnings inequality and growth among all formal workers in Colombia between 2009 and 2024 following [Guvenen, Pistaferri, and Violante \(2022b\)](#). Second, we exploit the longitudinal dimension of the data to characterize individuals' histories of formal attachment and show how these histories shape the distribution and volatility of workers' formal sector earnings.

Specifically, we use the Integrated Record of Social Security Contributions (PILA, for its Spanish acronym), a registry that contains the universe of formal employment records from 2009 to 2024. The PILA is a matched employer-employee panel constructed

---

<sup>1</sup>The repository is an open-access international database with harmonized longitudinal statistics on the distribution of earnings and earnings growth computed from administrative datasets ([Guvenen, Pistaferri, and Violante 2022a](#)).

from mandatory social security contribution records that encompass the universe of formal sector participants. Accordingly, we define formality as making contributions to social security, either as a wage-earning/salaried employee or as a registered self-employed worker. The longitudinal structure of the PILA allows us to track individuals across formal jobs and over time.<sup>2</sup>

We begin by calculating the harmonized statistics from the GRID project on earnings inequality and volatility in Colombia. The 2009–2024 period saw substantial earnings growth across the distribution and a small decrease in earnings concentration driven by a modest rise in the share of earnings of the bottom 50 percent. Earnings dispersion among formal workers was remarkably stable, with the P90–P10 differential fluctuating around 2.4 log points—a level comparable to the U.S., Brazil, and Argentina. Crucially, we find that formal earnings in Colombia are left-skewed, so that earnings dispersion is concentrated in the lower half of the distribution. We later link these findings to formal attachment trends among low earners.

Earnings growth was similarly stable, with limited variation except around the major 2012 labor market reform and the COVID-19 pandemic.<sup>3</sup> Formal earnings growth in Colombia exhibited relatively low volatility, approximately 0.3 log points below that of the U.S. and 0.4 log points below that of France. However, volatility is substantially higher for young workers and those at the bottom of the permanent income distribution.

We next characterize the dynamics of formal attachment over individuals' work lives and examine their role in shaping earnings dynamics. We measure formal attachment as the annual share of months in which an individual holds a formal job. We then identify the most salient patterns of attachment histories. There are stark differences

---

<sup>2</sup>The informal economy is by nature hidden from administrative registries such as the PILA (Ulyssea 2020; Flórez, Morales-Zurita, Medina, and Lobo 2021). Therefore, we cannot distinguish between outflows to informal work and non-employment.

<sup>3</sup>The dispersion of below-median earnings growth increased while above-median dispersion declined after the 2012 labor market reform. See Fernández and Villar (2017); Kugler, Kugler, and Prada (2017); Morales and Medina (2017) for studies on the effects of this reform.

in individuals' patterns of participation in the formal sector and their corresponding earnings trajectories. Approximately two-thirds of individuals exhibit either persistently low or persistently high formal attachment. The remaining third is divided between individuals with increasing formal attachment over their work lives, and a group of individuals who drop out of the formal labor market after the COVID-19 pandemic. Groups with low formal attachment exhibit the greatest dispersion and volatility in earnings over time.

We characterize the dynamics of formal attachment through *agglomerative hierarchical clustering*.<sup>4</sup> This methodology groups individuals based on their whole realized work histories, which is key for identifying changes in formal attachment over time. We are therefore able to capture differences in both the prevalence and timing of formal work, distinguishing individuals who enter formality early in life from those who formalize later, those who oscillate in and out year by year, and those whose attachment collapses during large shocks such as the 2020 pandemic. We apply the method to histories of formal attachment for the cohort of workers aged 25–30 in 2009.

We recover five distinct groups that explain 60 percent of the total variation in formal attachment and over 80 percent of the between-group variation. The largest group (42 percent of the cohort) maintains formal employment for just over one month per year, with earnings close to the minimum monthly wage. The second largest (22 percent) exhibits stable formal attachment (10 months per year on average), paired with consistent and high earnings growth. Two additional groups (representing a combined 28 percent of the cohort) show increasing formal attachment, though the onset of this increase occurs at different points of the life cycle. The final group of individuals held

---

<sup>4</sup>For an introduction to clustering, see [Hastie, Tibshirani, and Friedman \(2009, ch. 14\)](#). See [Audoly, M<sup>c</sup>Gee, Ocampo, and Paz Pardo \(2025\)](#) and [Lagakos, Michalopoulos, and Voth \(2025\)](#) for applications of agglomerative hierarchical clustering in economics. This method allows us to directly study labor market histories and complements alternative methods used to elicit underlying trends in labor markets, such as sequence analysis, as in [Humphries \(2022\)](#), K-means clustering, as in [Gregory, Menzio, and Wiczer \(2025\)](#), and hidden Markov chain models, as in [Ahn, Hobijn, and Şahin \(2023\)](#).

formal jobs, on average, for 6 months per year but then experienced sharp declines in formal attachment and earnings, first in 2015 and again during the COVID-19 pandemic.

The connection between formal attachment and earnings dynamics is striking. Earnings dispersion is concentrated in groups with low formal attachment and low earnings, primarily among those in the bottom half of these groups' earnings distributions. Groups with high attachment display substantially lower and more stable earnings dispersion. The pattern extends to earnings growth: growth volatility is highest among low-attachment groups. Crucially, even within groups, earnings dispersion and volatility are negatively correlated with formal attachment, rising as a group's attachment declines and falling as it increases. These dynamics provide a link between transitions into and out of formality, the left-skewness in earnings, and the high earnings volatility observed among low earners. They further show that the stability of cross-sectional earnings inequality masks substantial heterogeneity between groups and over time.

The literature on the relationship between transitions in and out of formality and earnings inequality and dynamics is scarce, with most studies focusing on the effects of average job-to-job transition rates or on isolated job-loss episodes. Such prior work finds that job flows are positively correlated with inequality (Jolly 2013; Carranza, Prieto, and Sehnbruch 2025; Menezes-Filho and Narita 2025) but does not examine the effect of transitions on earnings dynamics or study long-run trends in workers' labor-market transitions. We fill this gap by examining individuals' formal sector participation across the life cycle and studying the consequences of these trajectories on labor earnings dispersion and volatility. To this end, we use workers' entire formal sector labor histories. We show that the degree of formal attachment is a persistent feature for some individuals, while for others it changes over their life cycle and in response to aggregate shocks.

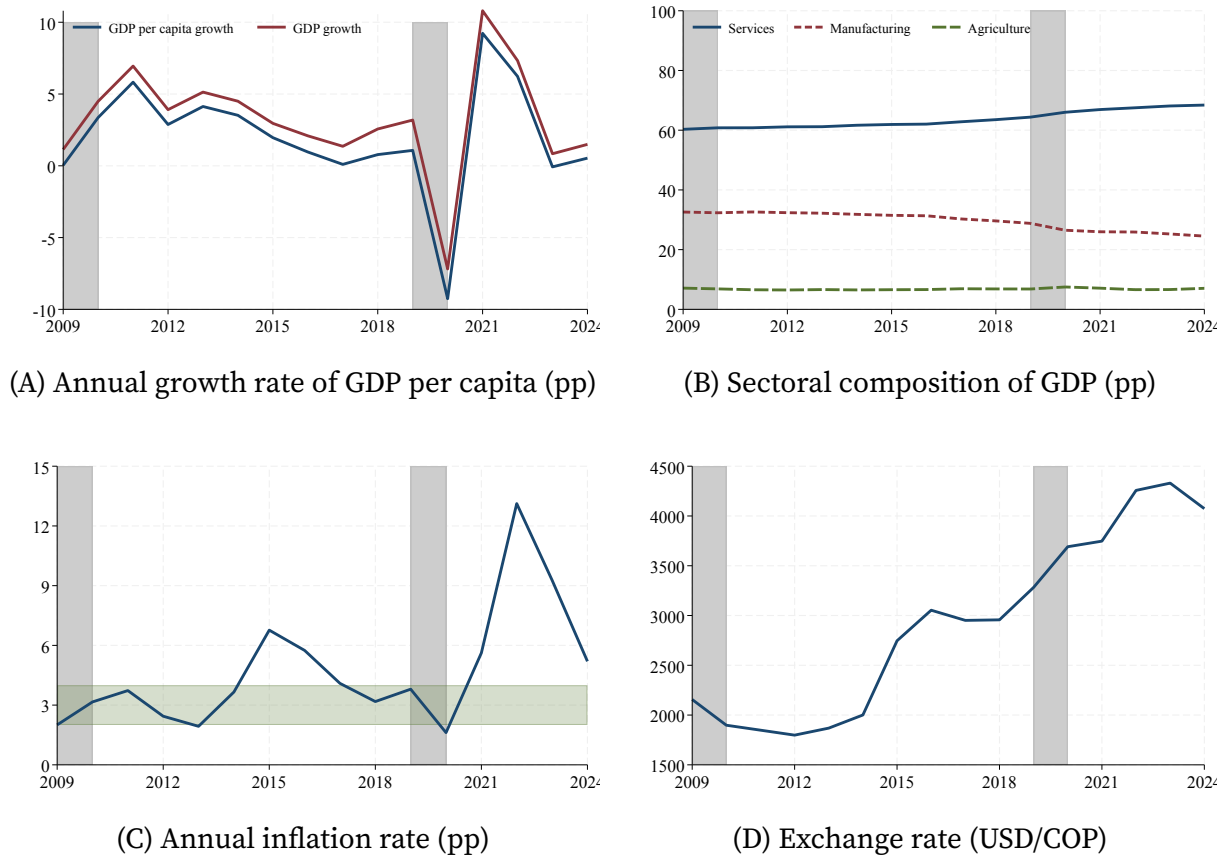
We also contribute to the broader literature on earnings dynamics, particularly the

growing body of work on earnings inequality and mobility across countries documented by the GRID project (Güvenen, Pistaferri, and Violante 2022b). We provide the GRID project's harmonized statistics for Colombia, describe their aggregate patterns and evolution over time, and show that long-run heterogeneity in formal labor market participation plays a central role in shaping individual earnings trajectories. While prior work has emphasized transitory and permanent shocks, our results highlight the importance of low-frequency, long-run features of individuals' employment histories in explaining the dispersion and volatility of earnings. Our focus on these long-term patterns is essential because it shifts attention from instantaneous, one-time shocks to career trajectories that build lifetime inequality. Moreover, this approach recognizes that participation in the formal sector is not a fixed worker characteristic. Instead, most individuals participate in the formal sector at some point during their working lives, but differences in timing and attachment to formal employment translate to substantial differences in earnings trajectories. In this context, our clustering methodology departs from static "formal" versus "informal" labels, providing a framework that captures the heterogeneity and dynamism in formal sector attachment.

## **2. Context on the Colombian Economy and Labor Market**

Our study period includes two economic deceleration periods and the recoveries that followed: namely, the Global Financial Crisis, 2008–2009, and the pandemic crisis in 2020. After the Global Financial Crisis, Colombian GDP per capita increased substantially, growing by 4 percent on average between 2010 and 2014 (see Figure 1A). The pandemic crisis was severe in Colombia. GDP per capita declined by 9.2 percent, which had a profound impact on labor markets. Nevertheless, the recovery was fast, as in many other countries, and GDP returned to its average growth trend after 2021. These business cycles were accompanied by a continued process of structural transformation that

FIGURE 1. Macroeconomic Indicators for Colombia

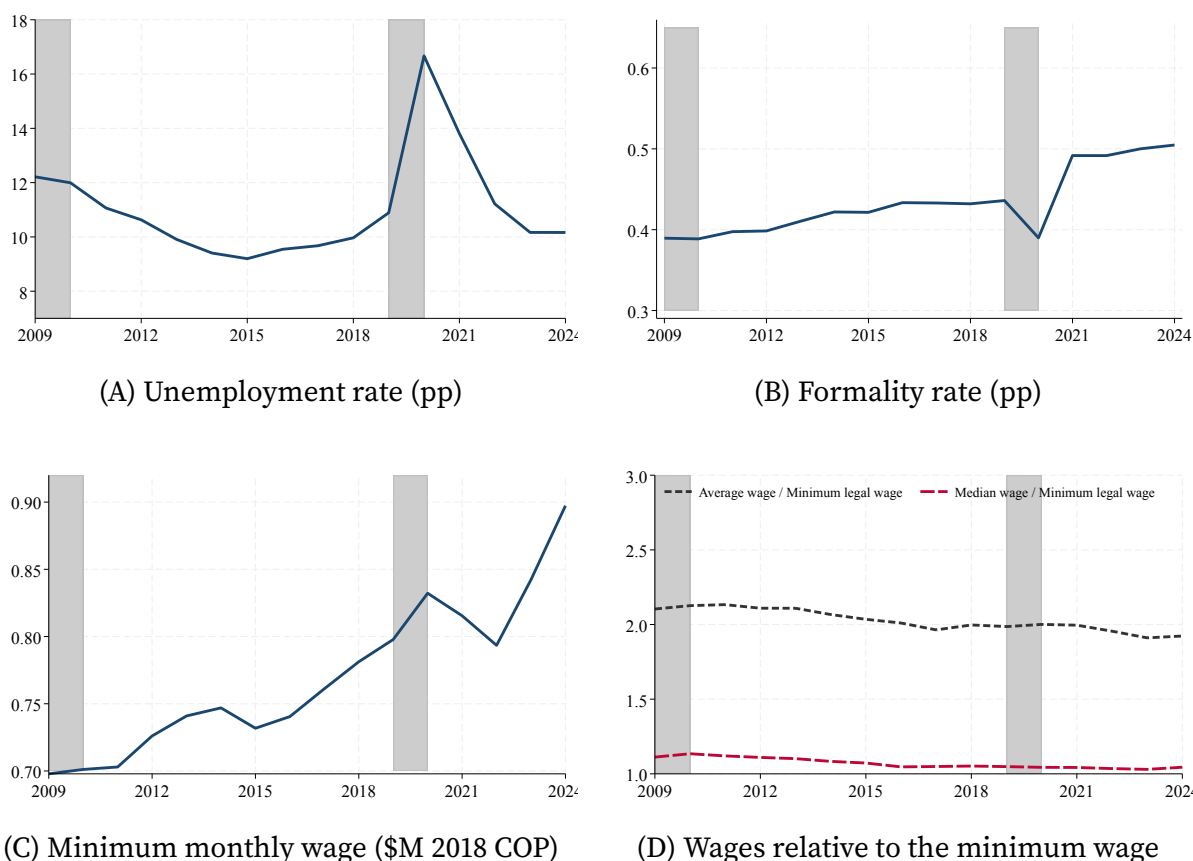


Note: Panel A shows annual growth of GDP per capita and GDP in percentage points computed from the national accounts and population files of the National Statistical Department (DANE). Panel B shows the composition of economic activity across sectors in percentage points computed from the national accounts files of the DANE. Panel C shows the year-on-year inflation rate of the Consumer Price Index (CPI) in percentage points computed by the DANE. Panel D shows the annual average exchange rate (USD/COP) reported by the Colombian central bank (Banco de la República). Shaded areas correspond to recession years as defined by Colombia’s business cycle chronology (Arango and Ospina 2025).

accelerated after the 2020 pandemic, as seen in Figure 1B. The share of services in GDP increased from 60 to 68 percent between 2009 and 2024, while manufacturing decreased from 32 to 25 percent. Agriculture’s share remained stable at 8 percent.

Colombia’s central bank, *Banco de la República*, has an inflation-targeting regime accompanied by a floating exchange rate (with interventions in case of heightened volatility). Figures 1C and 1D show, respectively, the inflation rate and exchange rate. The inflation target for this period was held constant within a 2–4 percent band, with

FIGURE 2. Labor Market Indicators for Colombia



Note: Panel A shows the annual average of the unemployment rate (not seasonally adjusted) as reported by the National Statistical Department (DANE) from its national household survey (GEIH). Panel B shows the formality rate measured using GEIH. Panel C shows the legal monthly minimum wage in millions of 2018 Colombian pesos as reported by the Colombian central bank (Banco de la República) and deflated using the Consumer Price Index. Panel D shows mean and median wages of formal workers relative to the legal monthly minimum wage computed with data from the PILA. Shaded areas correspond to recession years as defined by Colombia’s business cycle chronology (Arango and Ospina 2025).

inflation rising well above that range after the pandemic. The exchange rate (expressed in pesos per dollar) has depreciated almost continuously, more than doubling between 2009 and 2024.

Unemployment in Colombia is persistently high, as shown in Figure 2A, where we report unemployment relative to the active labor force that includes formal and informal workers using data from the Colombian household survey, the *Gran Encuesta Integrada de Hogares*. While unemployment was declining at the beginning of our sample,

starting at 11.4 percent and reaching a historical minimum of 9 percent in 2015, the slowdown experienced by the Colombian economy afterward pushed unemployment back up to 11 percent by 2019. The labor market disruption brought by the pandemic resulted in unemployment peaking at over 16 percent. The urban unemployment rate rose even higher, reaching 23 percent in the second quarter of 2020—the highest in history (Morales-Zurita, Bonilla-Mejía, Pulido, Flórez, Hermida, Pulido-Mahecha, and Lasso-Valderrama 2022). Nevertheless, the recovery was fast. Unemployment returned to its pre-pandemic levels by 2023.

The Colombian labor market also exhibits high informality, exceeding the regional average (Otero-Cortés 2025). Figure 2B reports the formality rate computed from the GEIH. These are employed (or self-employed) workers who pay payroll taxes and contribute to the public social security system. Fewer than half of workers have formal jobs at any point in time, with formality dipping during the pandemic, reaching a low of 38 percent in 2021 (Grupo de Análisis del Mercado Laboral 2025). These low levels of formality are accompanied by the prevalence of small and medium-sized enterprises—less than 50 employees—that account for 90 percent of formal firms in Colombia and 30 percent of employment (Flórez, Morales-Zurita, Medina, and Lobo 2021).

The relatively high and binding minimum wage plays an important role in Colombia's high levels of informality and unemployment. Figure 2C shows the inflation-adjusted monthly minimum wage expressed in millions of 2018 Colombian pesos. The real minimum wage has increased roughly in line with per capita GDP between 2009 and 2024, maintaining its relative level with respect to the distribution of wages. As shown in Figure 2D, the minimum monthly wage is nearly equal to the median monthly wage and roughly half of the average monthly wage among formal workers. This places Colombia's

minimum wage above that of other OECD countries (Arango and Flórez 2021).<sup>5</sup>

In fact, 40 percent of formal workers in a given month report earning the minimum wage, as reported by Becerra-Camargo and Morales-Zurita (2025). Nevertheless, fluctuations in formality over the year, like those studied in Section 5, as well as transitions into unemployment imply that the minimum wage is much less binding at the annual frequency that we study in Sections 4–6. Moreover, Colombian labor contracts have very limited flexibility in hours. Over 90 percent of wage-earning workers in the official household survey (GEIH) report working full-time.<sup>6</sup> This implies that the dispersion in annual earnings (particularly in the lower half of the distribution) reflects variation in months of formal attachment over the year, not variation in hours within a month.

## **2.1. The Colombian social security system and formal employment**

Colombian labor law requires employers to register their workers with the social security administration and to make payroll tax contributions. The same applies to independent (self-employed) formal workers, who must make their own contributions.<sup>7</sup> Registration in the social security system entitles the individual to a comprehensive set of benefits and protections: a health-insurance plan, a pension plan (administered through either the public defined-benefit regime, *Régimen de Prima Media*, or the private individual-account regime, *Régimen de Ahorro Individual con Solidaridad*), occupational-risk insurance, and paid maternity and disability leave.

---

<sup>5</sup>Moreover, most wage contracts, as well as taxes, fees, and other prices in the economy, are indexed to annual increases in the minimum wage (Grupo de Análisis del Mercado Laboral 2024).

<sup>6</sup>We define full-time work based on the maximum number of hours in a legal work week (48 hours up to 2022, 47 in 2023, and 46 in 2024).

<sup>7</sup>Most formal self-employed workers in Colombia operate under service-provision contracts, which require monthly proof of social-security contributions as a condition for payment from the client. This third-party validation may explain why self-employment is a larger share of formal employment in Colombia (about 20% in both PILA and survey data) than elsewhere. For instance, in Mexico it accounts for only 1.8% of administrative records but 8% in surveys.

Formal workers also accrue *cesantías*—a mandatory severance-savings scheme equivalent to one monthly wage per year of service—and gain access to the network of *Cajas de Compensación Familiar* (CCFs), private non-profit entities financed by a 4 percent employer payroll tax that administer a monthly family subsidy per dependent and provide subsidized food, pharmaceuticals, school supplies, training, recreation, and housing programs (Samaniego de la Parra, Otero-Cortés, and Morales-Zurita 2024). Since the enactment of Law 1636 of 2013, workers who have contributed to a CCF for at least twelve months over the previous three years are additionally covered by the *Mecanismo de Protección al Cesante*, an unemployment-insurance scheme that pays up to six months of contributions to health and pensions and continues the family subsidy in exchange for participation in active labor market policies coordinated through the public employment service (Astorga-Rojas, Guida-Johnson, and Morales-Zurita 2026). These benefits complement the substantially higher labor earnings of formal workers relative to those of informal workers. Using the Colombian household survey (GEIH), Bonilla-Mejía et al. (2025) estimate an unconditional formality wage premium of roughly USD \$129 per month in 2021.

This bundle of benefits is financed through mandatory contributions from employers and workers, with some components proportional to workers' wages and others levied on a per-worker basis. Payroll taxes, non-wage labor surcharges, and firing expenses increase the cost of formal hiring in the Colombian labor market, with labor costs exceeding the OECD average (Flórez, Morales-Zurita, Medina, and Lobo 2021); these factors have been mentioned in the literature as possible explanations for labor market rigidity in Colombia. Recognizing this, several labor market reforms were implemented during our study period, aiming, among other things, to reduce effective payroll tax rates, payroll costs, and social security contributions for part-time workers.

More specifically, Law 1607 of 2012 substituted part of the payroll contributions used to finance public health, family, and training programs for general corporate taxation,

lowering the non-wage cost of formal labor and contributing to documented increases in formal employment (Bernal, Eslava, Meléndez, and Pinzón 2017; Fernández and Villar 2017; Kugler, Kugler, and Prada 2017; Morales and Medina 2017). In addition, Decree 2616, in effect since 2014, allowed prorated social security contributions for workers employed for less than a full month, facilitating formal contracts for part-time jobs (Samaniego de la Parra, Otero-Cortés, and Morales-Zurita 2024). A set of fiscal reforms implemented between 2012 and 2015 cut employer payroll contributions by 13.5 percentage points for workers earning less than ten monthly minimum wages and replaced the lost revenue with a new 9 percent corporate income tax (CREE). This corporate income tax was then supplemented with a wealth tax on firms, and later replaced with increases in personal income taxes. Additional labor market policies during our study period include the 2011 First Job Act, which reduced payroll taxes for hires of workers under the age of 28 (Becerra-Camargo and Morales-Zurita 2025), and a comprehensive program of payroll subsidies as part of a fiscal stimulus package following the COVID-19 crisis in 2021 (Bonilla-Mejía et al. 2025), both with positive effects on formal employment growth. Finally, there was an additional increase in unemployment insurance benefits in 2022 (Astorga-Rojas, Guida-Johnson, and Morales-Zurita 2026).

### **3. PILA: Longitudinal data on formal workers**

Our analysis of earnings dynamics relies on annual earnings series aggregated from monthly administrative records in Colombia's social security system, managed by the Ministry of Health. Firms and self-employed workers are required by law to report all social security contributions through the *Integrated Record of Contributions to Social Security* (PILA, by its Spanish acronym). The data in the PILA system are compiled from payroll tax filings and therefore constitute a census of all formal employment matches in Colombia. Accordingly, we categorize a worker as formally employed in a

given month whenever PILA records a social security contribution, either filed by the worker directly (formal self-employed) or through an employer (salary- or wage-earning workers). In 2024, 80.5 percent of formal workers were wage-earning employees. The remaining 19.5 percent were self-employed ([Grupo de Análisis del Mercado Laboral 2025](#)).

The PILA has been used as the primary data source in a sizable body of work on the Colombian labor market (see, among others, [Kugler, Kugler, and Prada 2017](#); [Bernal et al. 2017](#); [Fernández and Villar 2017](#); [Samaniego de la Parra, Otero-Cortés, and Morales-Zurita 2024](#); [Eslava, Haltiwanger, and Pinzón 2022](#); [Kugler et al. 2020](#); [Bonilla-Mejía et al. 2025](#)). Moreover, this literature has compared the PILA to Colombia’s official household survey, the *Gran Encuesta Integrada de Hogares* (GEIH).<sup>8</sup> A direct comparison conducted by the Colombian central bank ([Banco de la República 2016](#)) found that the number of formal workers and the average labor income recovered from the PILA track their GEIH counterparts closely once differences in definitions and reporting incentives are accounted for. Both data sources also show common cyclical movements, including the 2020 contraction and the subsequent recovery, and exhibit the same long-run upward trend in formality. Appendix E provides additional context and compares the distributions of monthly earnings in PILA and GEIH.

We construct a monthly employer–employee panel using PILA’s administrative records from 2009 to 2024. We aggregate observations to an annual frequency, yielding, for each worker and calendar year, the worker’s annual labor income, an indicator of whether they made at least one social security contribution during the year, and the share of months in which they contributed. We also track firms’ average size, industry, and location (at the municipality level), as well as the workers’ age and gender.

In Section 4, we use the panel of all formal employment spells between 2009 and

---

<sup>8</sup>The GEIH is a monthly household sample administered by DANE, in which respondents self-report their employment status, sector of activity, and labor income. The survey is entirely cross-sectional, with retrospective questions that are used to study labor market transitions.

TABLE 1. Summary Statistics by Year

| Year | Obs.<br>(Mill.) | Earnings (Millions 2018 COP) |      |       |       |       |       |       |           |       | Log Earnings |       |           |       | Age Shares |       |       |
|------|-----------------|------------------------------|------|-------|-------|-------|-------|-------|-----------|-------|--------------|-------|-----------|-------|------------|-------|-------|
|      |                 | Percentiles                  |      |       |       |       | Mean  |       | Std. Dev. |       | Mean         |       | Std. Dev. |       | 25-35      | 36-45 | 45-55 |
|      |                 | p10                          | p25  | p50   | p75   | p90   | Men   | Women | Men       | Women | Men          | Women | Men       | Women |            |       |       |
| 2009 | 7.03            | 1.38                         | 4.29 | 7.61  | 9.77  | 17.91 | 9.79  | 9.71  | 15.49     | 12.62 | 1.83         | 1.90  | 0.95      | 0.89  | 46.5       | 32.7  | 20.7  |
| 2012 | 8.15            | 2.17                         | 5.07 | 7.96  | 11.36 | 21.25 | 11.26 | 11.24 | 17.74     | 13.96 | 1.97         | 2.05  | 0.95      | 0.89  | 46.6       | 31.5  | 21.9  |
| 2015 | 9.39            | 2.24                         | 5.97 | 8.95  | 12.55 | 23.08 | 12.29 | 12.41 | 19.17     | 15.32 | 2.07         | 2.15  | 0.94      | 0.89  | 46.9       | 31.1  | 22.0  |
| 2018 | 10.24           | 2.34                         | 6.25 | 9.37  | 12.97 | 24.16 | 12.67 | 12.87 | 20.22     | 15.89 | 2.10         | 2.18  | 0.94      | 0.89  | 46.4       | 32.1  | 21.5  |
| 2021 | 10.68           | 2.48                         | 6.62 | 9.93  | 13.73 | 25.66 | 13.42 | 13.66 | 19.95     | 16.54 | 2.16         | 2.24  | 0.95      | 0.90  | 46.3       | 33.0  | 20.8  |
| 2024 | 11.48           | 2.71                         | 7.21 | 10.82 | 14.67 | 27.11 | 14.45 | 14.44 | 19.94     | 16.87 | 2.26         | 2.31  | 0.93      | 0.88  | 45.7       | 33.5  | 20.8  |

**Notes:** The table reports summary statistics for workers in the sample following the GRID project's sample selection criteria. Income percentiles correspond to the cross-sectional distribution of annual labor income in each year. Mean and standard deviation statistics are reported separately for men and women. Age shares report the fraction of workers in each age group. Earnings levels are reported in millions of 2018 Colombian pesos, deflated using the Colombian Consumer Price Index. **Source:** Authors' calculations using PILA 2009–2024.

2024 to examine labor earnings dynamics and inequality in Colombia following the GRID project's methodology and sample criteria. Our primary variable of interest is real annual earnings,  $y_{it}$ , constructed following Guvenen, Pistaferri, and Violante (2022b). We deflate nominal earnings using the Colombian Consumer Price Index (CPI).<sup>9</sup> We also analyze residualized earnings,  $\varepsilon_{it}$ , permanent earnings, and one- and five-year changes in residualized earnings,  $g_{i,t}^h$  with  $h \in [1, 5]$ .

$$g_{i,t}^h = \Delta \varepsilon_{i,t} = \varepsilon_{i,t+h} - \varepsilon_{i,t}; \quad h \in \{1, 5\}. \quad (1)$$

We calculate residualized earnings as the residual of a regression of the natural log of real annual earnings,  $\log(y_{it})$ , on a set of age indicator variables separately by year and gender. Permanent earnings are mean real earnings over three-year rolling windows.

Table 1 presents summary statistics of the main variables of interest for formal workers that meet the GRID sample guidelines (Guvenen, Pistaferri, and Violante 2022b). That is, workers aged 25 to 55 who display *meaningful attachment* to the formal labor force as captured by a minimum annual earnings threshold  $Y_{min,t}$ .

We define this threshold so that it corresponds to a month of full-time minimum

<sup>9</sup>Series available at <https://www.banrep.gov.co/es/glosario/indice-precios-al-consumidor-ipc>

wage employment. This threshold follows the GRID guidelines while accommodating Colombia's labor market conditions. In the standard GRID guidelines, the minimum annual labor income threshold is set at 260 hours at the hourly minimum wage. This represents 13 weeks working 20 hours per week, approximating part-time employment. In Colombia, the minimum wage is expressed at a monthly frequency and the legal work week consists of 48 hours, which leads us to adjust the threshold.<sup>10</sup> This leads to a minimum annual earnings threshold of 1.25 monthly minimum wages. Two conditions hold in Colombia's labor market that justify lowering the threshold to one minimum wage. First, the monthly minimum wage is binding for a large share of labor contracts. As we discussed in Figure 2, 40 percent of formal workers earn the minimum wage in any given month. Second, an important share of individuals have formal sector earnings for less than two months per year, as the analysis of Section 5 shows. Setting  $Y_{min,t}$  to one monthly minimum wage in annual earnings ensures that our analyses speak to this large and economically meaningful part of the Colombian workforce. This leaves us with a sample that grows from 7 to almost 11.5 million workers, composed mostly of younger workers (about 45 percent are between 25 and 35 years of age), consistent with Colombia's national demographic structure.

In Section 5, we study the labor market trajectories of formal workers over their work lives. We select a cohort of workers aged 25 to 30 in 2009 and follow them through 2024, when they are between 40 and 45 years old. Starting the cohort at age 25 ensures that most workers have completed their education by the beginning of the sample, and it lets us observe them throughout their prime working years. We impose the same GRID sample requirements used in Section 4, so that this cohort is a strict *subsample* of the main GRID sample. Concretely, we build a balanced panel of formal-sector earnings and the share of months with social security contributions for every individual in the

---

<sup>10</sup>See, for example, the case of Mexico's GRID Phase 1 project, [Puggioni et al. \(2022\)](#). Since Mexico has a daily minimum wage, the threshold was set at 45 minimum wages.

cohort who meets the GRID minimum-income threshold at least once between 2009 and 2024. Because the panel is balanced, if an individual does not contribute in any month of a given year, we record both their annual formal earnings and formal attachment in that year as zero. In the years when they have at least one formal job, their earnings coincide with those in the main GRID sample.

## **4. Earnings dynamics of Colombia's Formal Workers**

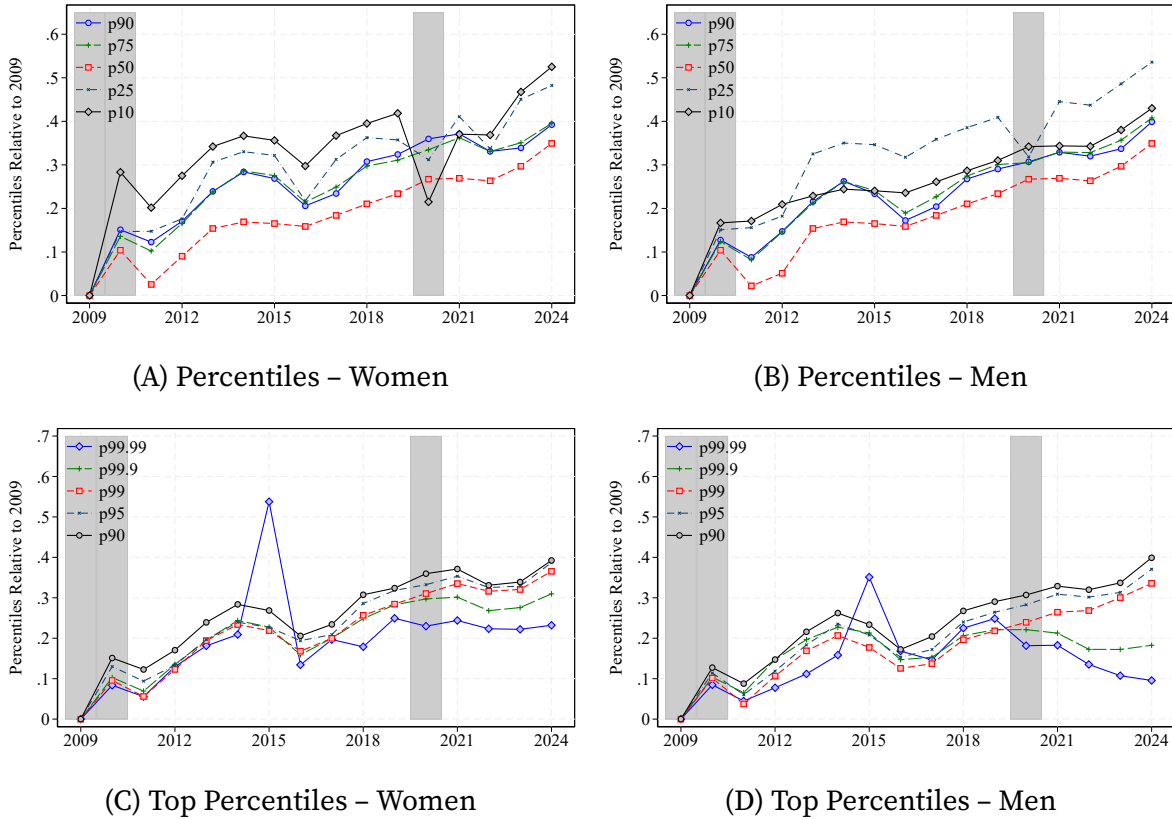
This section presents key statistics describing earnings inequality and the dynamics of earnings among Colombia's formal workers using PILA's administrative matched employer-employee data from 2009 to 2024. In constructing these statistics, we follow [Guvenen, Pistaferri, and Violante \(2022b\)](#).

### **4.1. Earnings inequality**

We start by highlighting the key takeaways from the evolution of the distribution of real log-earnings among formal workers presented in Figure 3. Panels A and B display the change for various percentiles of the distribution of the real annual log-earnings of formal workers, normalized to 2009 levels.

A general upward trend is visible across the whole distribution for women and men alike. For women, the bottom of the distribution grew at a faster pace; see the 10<sup>th</sup> and 25<sup>th</sup> percentiles (P10, P25), followed by the top half of the distribution. This fast growth resulted in a 0.5-log-point increase between 2009 and 2024, despite a substantial dip during the pandemic for the P10. For men, the growth at the bottom is not as pronounced, even though the 25<sup>th</sup> percentile exhibits the largest increase. Most percentiles of the men's log-earnings distribution grew at a common pace. The median (P50) grew at a lower pace for both women and men. By 2024, median earnings were 0.3 log points above their 2009 level. The differences are nevertheless small compared to the other

FIGURE 3. Evolution of Log-Earnings Percentiles



**Notes:** Panels A and B plot the bottom percentiles of the distribution of log real annual earnings, normalized to 2009. Panels C and D plot the top percentiles of the distribution of log real annual earnings, normalized to 2009. All results are for workers aged 25 to 55. Shaded areas correspond to recession years as defined by Colombia's business cycle chronology (Arango and Ospina 2025).  
**Source:** authors' calculations using PILA 2009–2024.

parts of the distribution. This results in a stable dispersion of log-earnings during our sample, as we report below in Figure 5.

**Growth at the top.** Panels C and D zoom into the evolution of the top of the log-earnings distribution for women and men, respectively. These panels show how the 90<sup>th</sup> through 99.99<sup>th</sup> percentiles have grown between 2009 and 2024. Echoing the results from panels A and B, growth at the top has been stronger for the lower parts of the right tail (the 90<sup>th</sup> and 95<sup>th</sup> percentiles) relative to the very top. It has also been weaker overall than among the bottom 90<sup>th</sup> percent of the distribution, particularly for men. Taken together, the top percentiles have gone up by 0.2 to 0.4 log points for women and 0.1 to 0.4 log points

for men (while the 10<sup>th</sup> percentile went up by over 0.5 and 0.4 log points, respectively). In fact, the very top of the distribution (the 99.9<sup>th</sup> and 99.99<sup>th</sup> percentiles) stalled for women after 2018 and actually decreased for men starting that same year.<sup>11</sup>

**Concentration of earnings.** The growth in earnings across the distribution described above led to modest changes in earnings shares and a slight decrease in earnings concentration. Figure 4 shows the cumulative change in the share of total earnings held by different segments of the distribution. Panels A and B divide the population by quintiles. While the series move over time, the most notable change for men and women is the decrease in the earnings share of the top quintile (Q5), which declined by three-quarters of a percentage point for women and over two percentage points for men, and the increase in the share of the second quintile (Q2) for men, which rose by three-quarters of a percentage point.

Panels C and D zoom into the top 10 percent of the distribution and contrast it with the earnings share of the bottom half of earners. The evolution at the top differs between women and men. Earnings shares for women were relatively stable, with top shares decreasing early by less than half a percentage point and holding relatively steady until 2022, when the top 5% and top 10% shares dropped another half a percentage point. At the same time, the earnings share of the bottom half decreases by similar magnitudes until 2020, when it starts growing post-pandemic, ending three-quarters of a percentage point above its original level.

Men have much larger changes in their earnings share. Until 2014, the labor earnings share for the bottom 50% of workers was constant at its 2009 level, while the top groups (those above the P90) experienced small declines. After 2014, the earnings share for

---

<sup>11</sup>The 99.99<sup>th</sup> percentile exhibits a sharp rise in 2015, and a subsequent fall in 2016, for both men and women. This is potentially in part a response to a payroll, personal income, and corporate tax reform from 2014 that took effect between 2015 and 2016. These changes in fiscal costs made it possible for employers to increase the compensation of high earners, while still reducing costs.

the bottom 50% of the distribution increased rapidly until 2017—for a total rise of about one percentage point that was maintained until 2022. This was followed by another increase, so that by the end of the sample, the bottom half of earners increased their earnings share by 2 percentage points relative to 2009. Conversely, top earners saw their share of total earnings decline, first for the lower part of the top tail (the top 10, 5, and 1%), and then for the very top, with the exception of the top 0.01% whose share did not change. These declines were significant for the lower part of the tail, going down one to two percentage points, and more modest at the very top.

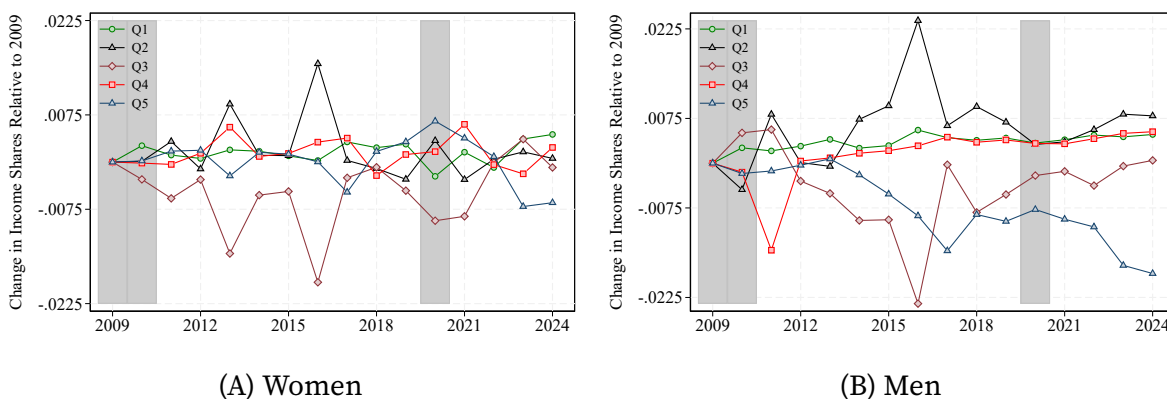
***Dispersion of log-earnings.*** Next, we turn to the dispersion in the distribution of log-earnings in Figure 5. The key takeaway is the marked stability of overall log-earnings dispersion throughout the entire period. Panels A and B plot the P90–P10 log-earnings gap and a scaled standard deviation ( $2.56 \times \sigma$ ) for log real annual earnings between 2009 and 2024. Both metrics track each other closely for women and men alike, with men showing slightly higher dispersion. Despite some minor year-to-year volatility, neither statistic shows a clear or sustained upward or downward trend. The one notable change occurs for women during the pandemic, when the distribution widens (as seen in the P90–P10 differential), then returns to its pre-pandemic levels.

Panels C and D decompose the overall dispersion of log-earnings (captured by the P90–P10 log-earnings gap) into the dispersion driven by individuals above (P90–P50) and below (P50–P10) the median. Both the P90–P50 and the P50–P10 earnings gaps are remarkably stable over the period we study. For women, the distribution is nearly symmetric around the median with very similar dispersion in the bottom and upper halves. Panel C shows that the spike in the dispersion of income for women during the pandemic is entirely driven by larger dispersion in the bottom half of the distribution (see the P50–P10 differential).

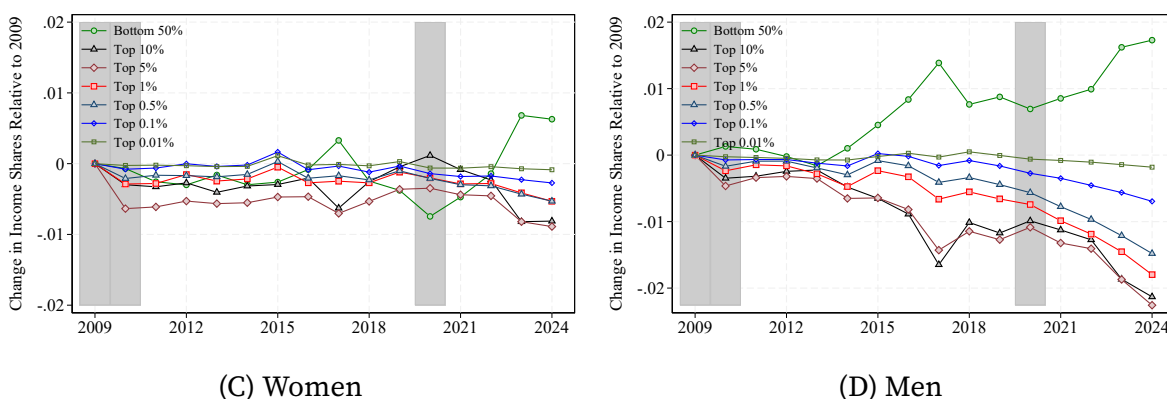
For men, the distribution of log-earnings is asymmetric. Dispersion below the

FIGURE 4. Evolution of Earnings Shares

Earnings Shares by Quintiles



Earnings Shares by Percentiles



**Notes:** Panels A and B plot changes in the income shares of quintiles of the earnings distribution relative to 2009. Panels C and D plot changes in the income shares of selected percentiles of the earnings distribution relative to 2009. A value of 0.01 corresponds to a 1 percentage-point change. All results are for workers aged 25 to 55. Shaded areas correspond to recession years as defined by Colombia’s business cycle chronology (Arango and Ospina 2025). **Source:** authors’ calculations using PILA 2009–2024.

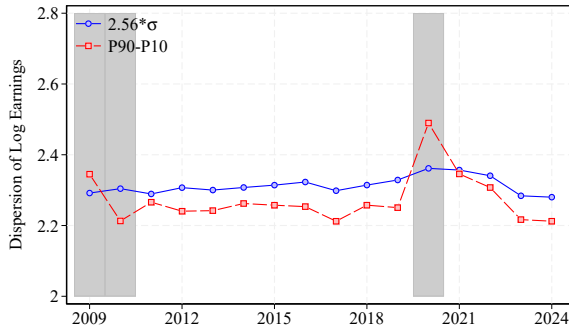
median is consistently and significantly larger than the dispersion above the median—workers at the bottom are further from the median than high earners.<sup>12</sup> While the P50–P10 log-earnings gap consistently hovers around 1.4 (except for a short-lived decline in 2011–2012), the P90–P50 log-earnings gap is notably smaller, close to 1. Thus, a key insight is that left-skewness in the distribution of log-earnings is an important driver of overall inequality for men.

<sup>12</sup>We return to this fact in Section 6 where we link earnings dynamics to formal attachment.

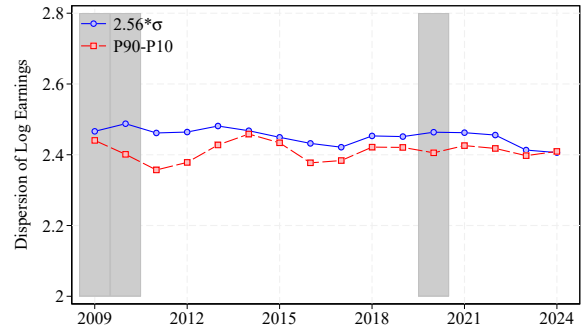
***Initial conditions.*** Finally, we look at the dispersion in log-earnings for workers aged 25 in Panels E and F of Figure 5. This gives us an idea of the dispersion in log-earnings early in workers' careers, near their entry to the labor market. The results are similar for women and men and are maintained throughout our sample period. The log-earnings of young workers are as dispersed as the log-earnings of the population at large, but this dispersion is much more concentrated in the bottom half of the log-earnings distribution. This is most striking for women, whose overall distribution of log-earnings is nearly symmetrical in the population, and yet their early log-earnings are much more dispersed below the median.

FIGURE 5. Log-Earnings Inequality

Dispersion of Log-Earnings

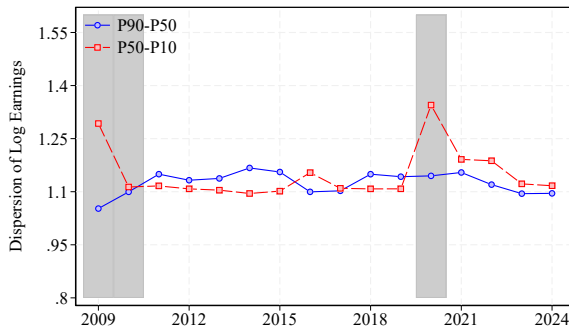


(A) Women

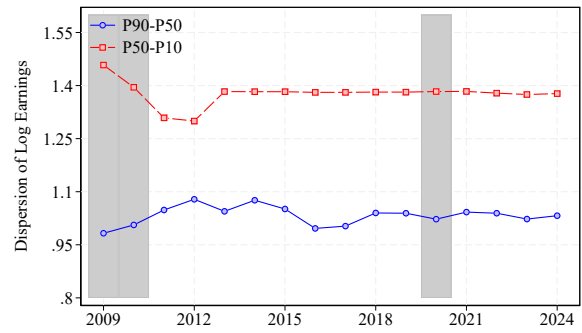


(B) Men

Dispersion Above and Below Median

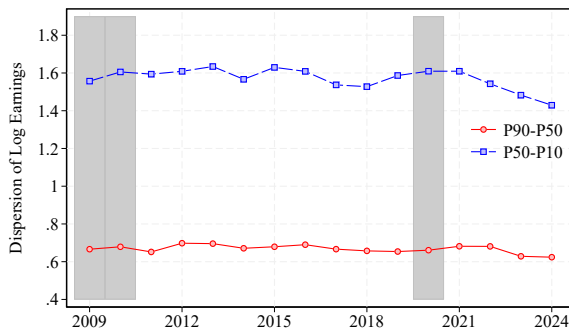


(C) Women

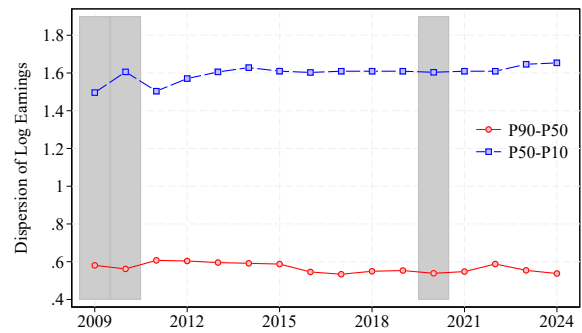


(D) Men

Dispersion Above and Below Median at Age 25



(E) Women



(F) Men

**Notes:** Distribution of log real earnings of formal workers aged 25 to 55. Panels A and B plot the P90–P10 differential and the scaled standard deviation ( $\sigma$ ). Panels C and D plot the P90–P50 differential (blue circles) and the P50–P10 differential (red squares). Panels E and F plot the P90–P50 differential (red squares) and the P50–P10 differential (blue circles) for workers aged 25. Shaded areas correspond to recession years as defined by Colombia’s business cycle chronology (Arango and Ospina 2025). **Source:** authors’ calculations using PILA 2009–2024.

## 4.2. Earnings dynamics

We next move to examine the dynamics of formal labor earnings. Again following [Guvenen, Pistaferri, and Violante \(2022b\)](#), we define residualized changes in log-earnings for individual  $i$  between periods  $t$  and  $t + k$  as  $g_{i,t}^k = \varepsilon_{i,t+k} - \varepsilon_{i,t}$  as in equation (1), where  $\varepsilon_{i,t}$  is the residual of a regression of log-earnings on a set of age dummies, run separately by year and gender. Here, we examine one-year changes in earnings (that is,  $k = 1$ ) to first consider transitory earnings shocks. Figure 6 looks at the distribution of these changes for women and men. We reproduce our results in Appendix A for five-year changes in earnings ( $k = 5$ ) and analyze the distribution of earnings dynamics, focusing on more permanent shocks.

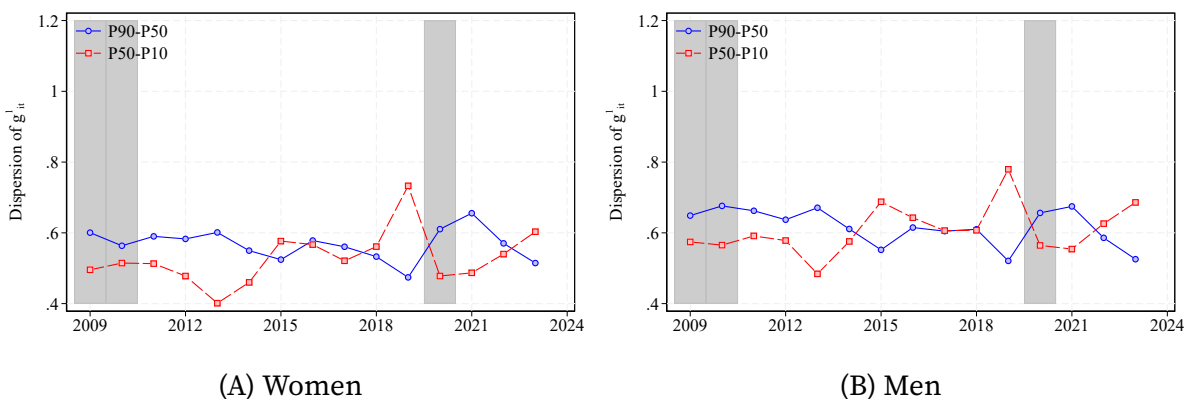
Panels A and B in Figure 6 present the difference in the annual growth rate of residualized log earnings ( $g^1$ ), relative to the median, for the upper (P90–P50) and lower (P50–P10) halves of the distribution. The results are similar for women and men and show an almost symmetric dispersion around the median in the distribution of residual log-earnings growth, although women show slightly lower dispersion. Prior to 2012, the spread at the top of the log-earnings growth distribution lies above that of the bottom half, and both are stable just above and below 0.6 log points, respectively.

However, there is a reversal of this pattern in 2012, coinciding with a major tax and labor market reform in Colombia.<sup>13</sup> After 2013, the dispersion in below-median earnings growth rises, while that in above-median growth drops slightly, leading to an overlap in the upside and downside spreads. Around the onset of the pandemic, increased volatility raised the downside spread, consistent with large negative shocks such as job loss or pay cuts becoming a more significant driver of one-year earnings volatility around that time; from the perspective of 2019, the drop in income during 2020

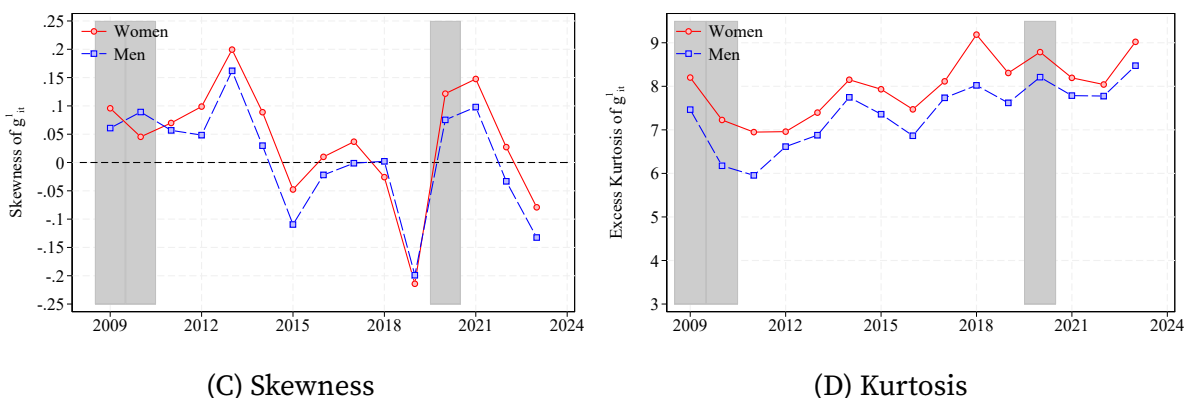
---

<sup>13</sup>See, for example, [Morales and Medina \(2017\)](#), [Fernández and Villar \(2017\)](#), and [Bernal, Eslava, Meléndez, and Pinzón \(2017\)](#) for various analyses of the reform’s impact on formality, employment, and wages.

FIGURE 6. Distribution of 1-Year Residual Log-Earnings Growth  
Dispersion Above and Below Median



Kelley Skewness and Crow-Siddiqui Excess Kurtosis of 1-Year Changes,  $g_{i,t}^1$



**Notes:** Panels A and B plot the P90–P50 differential of the annual change in residualized log-earnings (blue circles) along with the corresponding P50–P10 differential (red squares) for formal workers in each year. Panel C plots the Kelley Skewness measure of the annual change in residualized log-earnings for women (red circles) and men (blue squares). Panel D plots the Crow-Siddiqui Excess Kurtosis measure of the annual change in residualized log-earnings for women (red circles) and men (blue squares). All results are for workers aged 25 to 55. Shaded areas correspond to recession years as defined by Colombia’s business cycle chronology (Arango and Ospina 2025). **Source:** authors’ calculations using PILA 2009–2024.

appears as negative one-year growth in residualized log earnings. In the recovery period after the pandemic, there is a renewed increase in the dispersion of the below-median earnings growth and a decrease in that of above-median earnings growth, similar in magnitude to the changes observed after the 2012 labor market reform.

Panels C and D plot the Kelley skewness measure and the Crow-Siddiqui excess kurtosis measure of the annual change in residualized log-earnings. The moments of the distribution are similar for women and men, although women show consistently

higher skewness and kurtosis throughout the sample. The series for skewness in panel C reflect the evolution of the dispersion in panels A and B. Skewness is positive at the beginning of the sample and starts falling after 2012, turning negative by 2015. It then rebounds, maintaining levels close to symmetry, before dipping strongly in 2019, reflecting the earnings losses of the subsequent pandemic year. The recovery comes along with earnings growth and a rebound to positive skewness until 2022. The kurtosis series shows the now standard excess kurtosis of log-earnings changes (see, among others, Guvenen, Ozkan, and Song 2014; Guvenen, Pistaferri, and Violante 2022b). There are, nevertheless, meaningful changes over time, with kurtosis decreasing during the economic contraction at the beginning of the sample period, then increasing by 2 points over the decade that follows.

Finally, we move away from the time-series perspective and instead analyze dispersion in log-earnings growth across workers' life cycles. Figure 7 presents the P90-P10 differential in one-year residualized log-earnings growth for different age groups across percentiles of permanent earnings, allowing us to examine which groups of workers experience more volatility.<sup>14</sup> Panels A and B plot the P90–P10 differential in the log-earnings growth. Results are similar between women and men, although men show higher dispersion. Overall, we find that earnings volatility is substantially higher for individuals at the very bottom of the permanent income distribution (the lowest 10 percentiles). Volatility then declines sharply until the 40<sup>th</sup> percentile and remains flat for the vast majority of the income distribution. Furthermore, earnings volatility declines with age across the permanent income distribution. Dispersion is highly concentrated among the youngest workers and those with the lowest permanent income (similar to the results in panels E and F of Figure 5).<sup>15</sup>

---

<sup>14</sup>We define permanent income for person  $i$  in period  $t$  as the average labor earnings reported in PILA, including any zeros, in 3-year rolling windows. We exclude individuals with earnings below the minimum earnings cutoff for two or more periods within the rolling window. For the analyses that use permanent income quantiles, we average the statistics of interest across the 2009 to 2024 period.

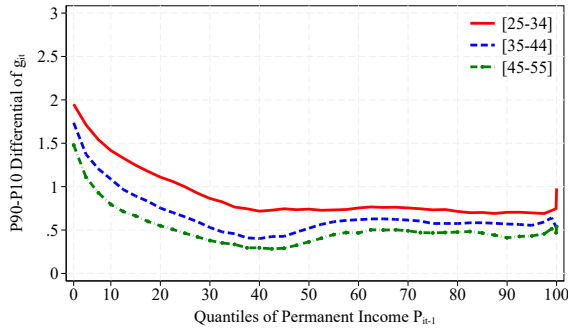
<sup>15</sup>The same patterns are present when plotting the standard deviation of one-year log-earnings growth.

Panels [C](#) and [D](#) plot the Kelley skewness measure. Consistent with international evidence, the distribution of residualized log-earnings growth is positively skewed at the bottom and negatively skewed at the top. We find small differences between age groups (except for women aged 45 to 55). What is noteworthy is the difference between the distributions of women's and men's log-earnings growth. Women's distribution is positively skewed below the 80<sup>th</sup> percentile of permanent income, while men's distribution is positively skewed only below the median. This can, in part, reflect the lower earnings of women in the Colombian formal labor market.

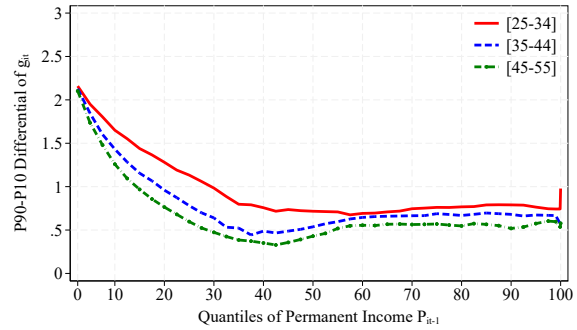
Panels [E](#) and [F](#) plot the Crow-Siddiqui excess kurtosis measure. While the distribution of log-earnings changes is leptokurtic across the distribution of permanent income, there is a clear increase in kurtosis for higher-earning workers with a spike at the very top and a spike between the 40<sup>th</sup> and 50<sup>th</sup> percentiles of permanent income. The spike at the top is consistent with the international evidence on the distribution of log-earnings changes. The spike near the middle of the distribution is consistent with the outsized impact of the minimum wage for Colombian workers, to which we return in Sections [5](#) and [6](#).

FIGURE 7. Distribution of Residualized Log-Earnings Growth by Permanent Income

P90–P10 Differential

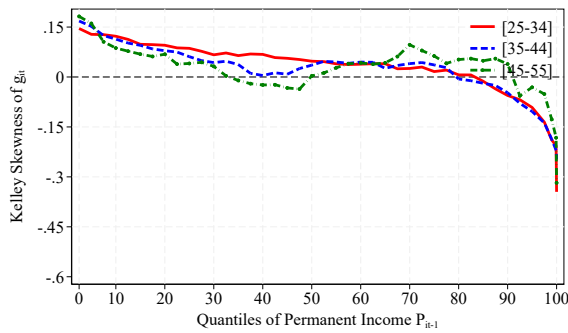


(A) Women

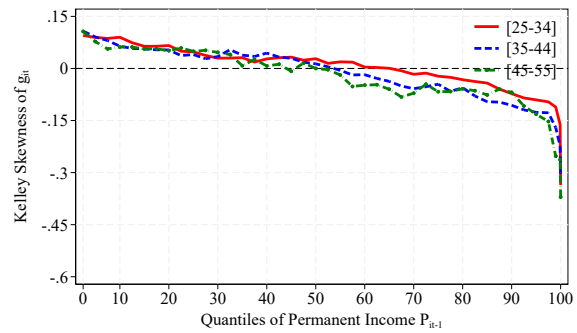


(B) Men

Kelley Skewness

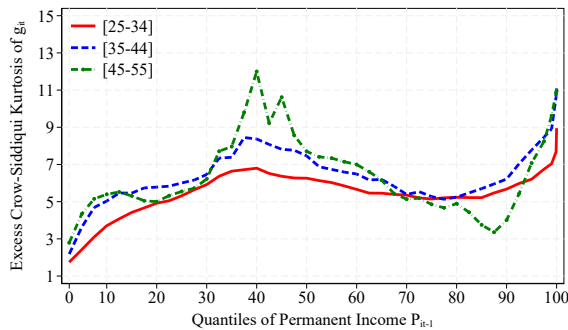


(C) Women

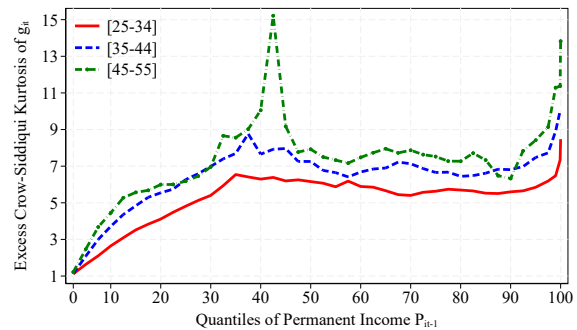


(D) Men

Crow-Siddiqui Excess Kurtosis



(E) Women



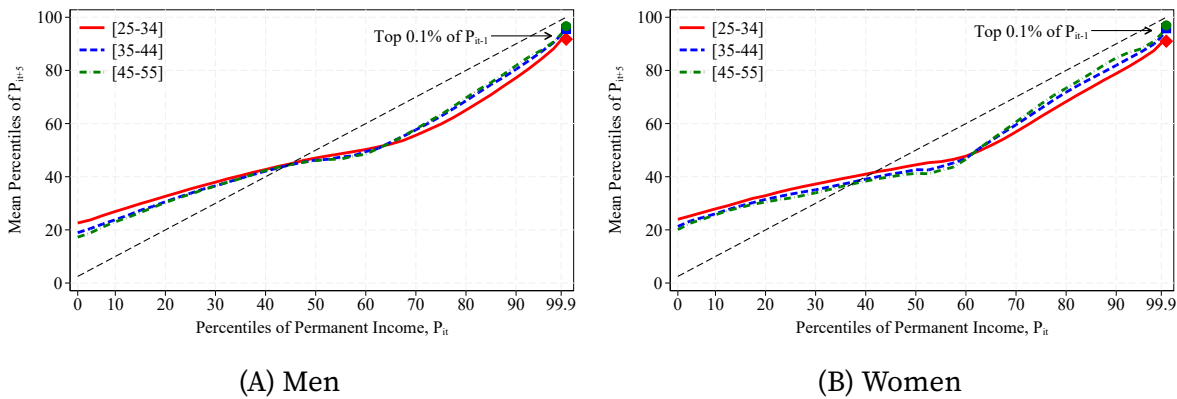
(F) Men

**Notes:** Panels present moments of the distribution of the one-year forward change in residualized log-earnings by quantile of the permanent income  $P_{i,t-1}$  and for individuals aged 25–34 (solid red line), 35–44 (dashed blue line), and 45–55 (dot-dashed green line). Permanent income is computed as the three-year average of earnings. All results are for workers aged 25 to 55. Panels A and B plot the P90–P10 differential for women and men, respectively. Panels C and D plot the Kelley skewness measure for women and men, respectively. Panels E and F plot the Crow-Siddiqui excess kurtosis measure for women and men, respectively. **Source:** authors' calculations using PILA 2009–2024.

### 4.3. Life-cycle dynamics of earnings mobility

Lastly, we examine earnings mobility. Figure 8 plots workers’ current permanent income percentile against their mean percentile five years later, broken down by age. The 45-degree line indicates the no-mobility benchmark where individuals remain at the same percentile of permanent income across time. Mobility follows an S-shaped curve for all age groups and for women and men alike: future ranks are above the 45-degree line at the bottom of the current permanent income distribution and below it at the top, with flattening at the middle. This curve is consistent with a “regression-to-the-mean” pattern. Individuals below the median are more likely to “move up the percentile ladder” (while staying below the median), and individuals in the upper half move down (while staying above the median). Mobility is lowest at the median and at the very top of the earnings distribution, consistent with the positively skewed nature of five-year shocks shown in Figure A.1, which highlights the role of large upward shocks as a key driver of dispersion.

FIGURE 8. Evolution of 5-Year Mobility Over the Life-Cycle, by Gender



**Notes:** The figure plots the average percentile of permanent income in  $t + 5$ ,  $P_{i,t+5}$ , conditional on the current percentile of permanent income,  $P_{i,t}$ , for individuals aged 25–34 (solid red line), 35–44 (dashed blue line), and 45–55 (dot-dashed green line). Permanent income is computed as the three-year average of earnings. Panels report results for men and women, respectively. All results are for workers aged 25 to 55. **Source:** authors’ calculations using PILA 2009–2024.

## 5. Labor market trajectories and formal attachment

One chief challenge in understanding earnings dynamics in markets with high informality is that workers are not permanently in formal or informal employment. Instead, they transition between employment relationships in the formal and the informal sector, as well as through unemployment spells and time out of the labor force. So, while we know that close to 50 percent of the Colombian working population is in a formal employment relationship at any given time, this statistic is only a static and incomplete snapshot of the underlying dynamic labor market trajectories that shape income dynamics in the formal sector.

These transitions into and out of formal work relationships raise fundamental questions regarding workers' *formal attachment* (that is, how consistently individuals hold formal employment over time) and its impact on the distribution and volatility of earnings.<sup>16</sup> In particular, the behavior of earnings can differ substantially between workers who persistently hold formal employment and those who are seldom formally employed, as well as between workers who become more attached to the formal labor market over time and those who instead detach from the formal sector (either having informal employment relationships or dropping out of the labor force). Beyond earnings dynamics, characterizing the labor market by the types of trajectories workers follow provides a key missing piece of information for policymakers who are mostly informed by cross-sectional reports on the state of the labor market.

In this section, we use the PILA to follow individuals through their prime working years to document trends in workers' formal attachment and how these trends shape

---

<sup>16</sup>There are many reasons for some workers to be more consistently employed in the formal labor market and others not. While some of those reflect personal preferences, the word *attachment* should not be understood as a desire of the worker to be formal. It is, for instance, possible for workers who are not in a formal employment relationship to want to be in one, but to be unable to be in one due to labor market frictions. It is also possible for workers to elect to be in an informal employment relationship or to choose to be out of the labor market entirely.

earnings dynamics (as in the analysis in Section 4 following GRID guidelines). The longitudinal dimension of the data is key to this, as it overcomes the limitations faced by analysis based on information from household surveys, which capture only short-run transitions, thereby limiting the ability of researchers and policymakers to study labor market trajectories over years.<sup>17</sup>

### 5.1. Labor market trajectories of formal attachment

We provide a new characterization of workers' trajectories in the formal labor market by identifying the main patterns of formal attachment across individuals' work lives. For this, we focus on the cohort of workers in their prime working years during our sample. We use the longitudinal dimension of our data to study the work histories of the cohort of individuals within the GRID sample who were 25 to 30 years old in 2009 and had at least one formal employment spell between 2009 and 2024. We therefore observe these individuals' labor market trajectories up until they are 40 to 45 years old. There are 3,551,608 such individuals.<sup>18</sup> Critically, our data makes it possible to follow individuals over 16 years and directly capture their patterns of formal attachment over a decade and a half. Our focus on the cohort of workers aged 25 to 30 in 2009 also ensures that our analysis is balanced with respect to age, so that we are not mixing patterns of high- or low-formality between younger and older workers in any given year.

We measure individual  $i$ 's formal attachment for calendar year  $t$ ,  $f_{i,t}$ , as the share of months in  $t$  in which they contribute to the social security system (seen as entries in

---

<sup>17</sup>The longitudinal data from the PILA have their own limitations. Most notably, it does not follow individuals out of formal employment relationships, so it is not possible to know when a worker is in an informal employment relationship. It also does not follow individuals who have migrated out of the country. The data from the official household survey have their own limitations, as they rely on short-run retrospective questions that do not directly link workers to their formal status for computing transition statistics.

<sup>18</sup>In order to belong to our sample, individuals must have reported earnings to the PILA in at least one month between 2009 and 2024 and satisfy the sample selection criteria laid out in Section 3 for at least one year in order to construct a balanced panel. Focusing on the group of young (25–30) individuals in 2009 highlights the life cycle's role and the effects of aggregate shocks to the Colombian labor market, such as the 2020 pandemic.

the PILA). Crucial to our purposes, our sample for this exercise is a balanced panel. An individual is included in the sample even if they do not have any formal employment throughout the whole calendar year. Formal attachment for that individual is zero in that year,  $f_{i,t} = 0$ .<sup>19</sup> Similarly, if in year  $t$  an individual holds a formal job for 3 months, their formal attachment would be  $f_{i,t} = 0.25$ . An individual who is employed formally throughout the year has  $f_{i,t} = 1$ .

Our objects of interest are the trajectories of formal attachment. For each individual  $i$  we construct a vector of length  $T = 16$  with each element indicating individual  $i$ 's formal attachment in year  $t$ ,  $F_i \equiv [f_{i,2009}, \dots, f_{i,2024}]$ . Importantly, both the order of the elements in the vector and their values provide information relevant to characterize an individual's formal attachment history. For example, consider three people who were formally employed for 16 months over the 16-year period between 2009 and 2024. Person 1 works formally for 16 consecutive months when they are 25 years old; person 2 works 16 consecutive months when they are 40 years old; and person 3 works one month per year in each year between 2009 and 2024. In this example, we would have  $F_1 = (1, 1/3, 0, \dots, 0)$ ,  $F_2 = (0, \dots, 0, 1/3, 1)$ , and  $F_3 = (1/12, \dots, 1/12)$ . These three histories paint very different pictures of formal attachment, despite all having the same number of months contributing to the formal sector.

Workers in the cohort we study exhibited an annual average formal attachment of 39 percent, spending 4.6 months per year in formal employment over the 16 years of

---

<sup>19</sup>The GRID project's sample criteria specify a minimum annual labor income threshold equivalent to earning 260 hours at the hourly minimum wage, which we adapt to the Colombian context by imposing a threshold of one monthly minimum wage. Therefore, individuals with  $f_{i,t} = 0$  would be excluded from the GRID sample in year  $t$  since annual formal earnings would be zero. Because understanding the inflows and outflows to the formal sector is a key component of the labor market trajectories analysis, we apply the minimum earnings threshold at the individual level, instead of the individual-year level. That is, we construct the sample in two steps. First, we identify all the individuals who meet the minimum earnings threshold criteria at least once and who are 25–30 years old in 2009. Second, we construct a balanced panel of their formal employment and earnings from 2009 through 2024.

the sample.<sup>20</sup> Formal attachment increases over time. Specifically, individuals aged 25 to 35 were formally employed for approximately one-third of the year (4 months per year). Attachment rises to just over 40 percent (5 months) for workers over the age of 35. These averages mask substantial heterogeneity in workers' formal attachment. We next turn to characterize this heterogeneity and to analyze its implications for the Colombian labor market and formal earnings dynamics.<sup>21</sup>

## 5.2. Finding patterns of formal attachment in labor market trajectories

Our objective now is to dissect the cohort, recovering the most salient patterns of formal attachment through individuals' work lives. For this, we use *agglomerative hierarchical clustering* to flexibly and non-parametrically characterize the most relevant labor trajectories in the data. The clustering works by grouping individuals with similar labor market trajectories of formal attachment, that is, with similar vectors  $\{F_i\} \in \mathbb{R}_+^{16}$  as determined by the standard Euclidean distance.

Crucially, by comparing full labor market trajectories, we do not restrict which elements in an individual's history matter for the groupings, nor do we collapse histories into a single summary statistic. Moreover, because labor market trajectories are ordered vectors, the timing of when workers hold a formal job matters for distinguishing them, as we discussed above. So, the groupings take into account the full heterogeneity in formal participation across individuals and its variation over time.

---

<sup>20</sup>Our sample, and the analyses and statistics that follow, refer to the set of workers with at least one formal employment spell between 2009 and 2024. In this sense, formal attachment excludes individuals with zero formal attachment in every year during the analysis period.

<sup>21</sup>It is important to clarify that a low value of formal attachment does not imply high "informal attachment." Our data allow us to observe spells of formal employment, but we are not able to determine whether the remaining time is spent in informal employment, unemployment, or out of the labor force. However, it is likely that many non-formal spells are spent in informality, given Colombia's high levels of labor market participation for individuals between the ages of 25 and 40.

**Agglomerative hierarchical clustering.** One key advantage of the algorithm is that it is entirely transparent in how it groups workers. The algorithm works by recursively pairing the two observations with the closest (most similar) labor market trajectory vectors. Formally, the algorithm starts from  $G = N$  groups (with a single worker in each group) and proceeds by merging (or agglomerating) workers with the closest vectors. These two workers form a new group, represented by their average formal attachment vector. There are now  $G = N - 1$  groups. The algorithm continues performing the same steps, agglomerating the two groups with the closest vectors, reducing the number of groups by one in each iteration. This algorithm ends when there is only one group with the average labor market trajectory of the sample.

The algorithm proceeds by repeatedly grouping the most similar workers into larger clusters. The order of these merges defines the hierarchy: workers grouped first have the most similar labor trajectories, while those grouped later are increasingly different. Moreover, groups are nested across levels of the hierarchy by construction. In other words, any group formed at a later step is the union of two groups from earlier steps, and therefore fully contains both of them. This nesting has an important practical implication: choosing the number of clusters to analyze is not an arbitrary decision. Different choices yield consistent groupings of workers, with finer partitions (that is, more clusters) splitting existing groups into smaller ones, rather than reallocating workers across them. We discuss our preferred number of clusters below.<sup>22</sup>

We operationalize the algorithm via *Ward's method* and use the total within-cluster variance as the dissimilarity metric. Equation (2) presents the minimization problem:

$$\operatorname{argmin}_{g, g' \in G, g \neq g'} \sqrt{\frac{2N_g N_{g'}}{N_g + N_{g'}}} \times \left\| \bar{\mathbf{F}}^g - \bar{\mathbf{F}}^{g'} \right\|_2, \quad (2)$$

where  $g$  and  $g'$  are disjoint groups,  $N_g$  denotes the number of observations in group

---

<sup>22</sup>We provide a step-by-step numerical example of how the algorithm works in Appendix B.

$g$ , and  $\bar{F}^g$  is the average (or centroid) value for the outcome of interest across the observations in group  $g$ . In our case, this outcome is a vector of element-wise averages of formal attachment histories  $F_i$  for the members of the group. We implement the algorithm on a 5 percent sample of our cohort.<sup>23</sup>

**Selecting groups.** A crucial feature of our approach is that it does not pre-specify the number of groups to use when categorizing individuals' histories. Instead, as explained above, the algorithm outputs a hierarchy of groupings that spans all possible numbers of groups  $G = \{1, \dots, N\}$ . When increasing the number of groups from  $g$  to  $g + 1$ , the algorithm splits the group with the highest variance in formal attachment trajectories among its members, leaving the other  $g - 1$  groups unchanged. This implies a nested (or hierarchical) structure.

We compute the share of the total variation in work histories explained by the clusters at each level of the hierarchy, the  $R^2$ , and then select the minimum number of clusters so as to explain the majority (60 percent) of the variation in formal attachment. That is, for a partition  $\mathcal{G}_G = \{g_i\}_{i=1}^N$ ,

$$R^2 = 1 - \frac{\sum_{i,t} (f_{i,t} - \bar{F}_t^{g(i)})^2}{\sum_{i,t} (f_{i,t} - \bar{F})^2}, \quad (3)$$

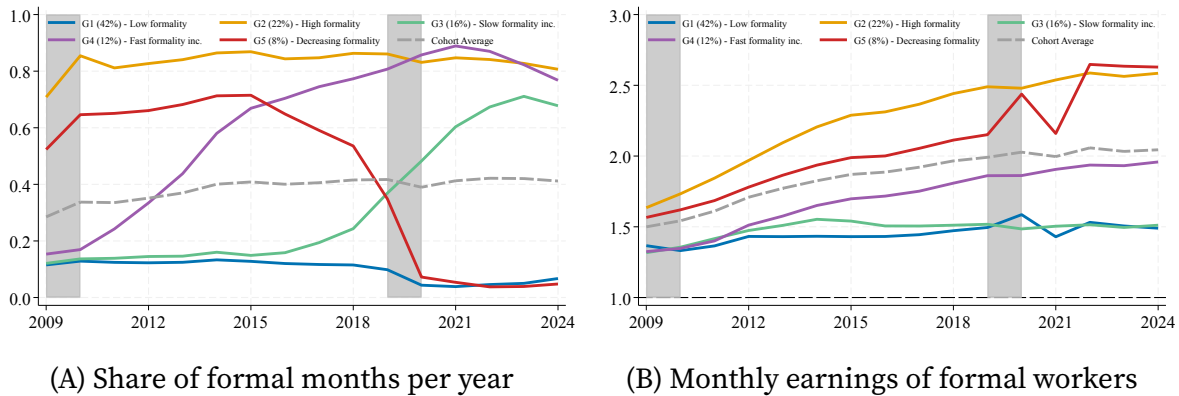
where  $f_{i,t}$  is the share of months worker  $i$  contributed to the formal social security system in year  $t$ ,  $\bar{F}_t^{g(i)}$  is the average for the worker  $i$ 's group, and  $\bar{F}$  is the average in the sample (over workers and time).

We select  $G = 5$  groups that explain 60 percent of the total variation in the sample and over 80 percent of the between-group variation in histories of formal attachment.

---

<sup>23</sup>Administrative restrictions imposed by the data provider on the use of the PILA prevent us from using our full cohort universe of individuals. However, we re-estimated all results using different random samples (5 and 10% of the full cohort). Our results are robust to different random samples.

FIGURE 9. Labor Market Trajectories of Formal Attachment



**Notes:** Panel A plots the average share of formal months for the workers of each formal attachment group (solid lines) and in the sample cohort (dashed gray line). Panel B plots the average monthly earnings of formal workers of each formal attachment group (solid lines) and in the sample cohort (dashed gray line). All results are for the cohort of individuals aged 25 to 30 in 2009 who have ever reported earnings to the PILA. Shaded areas correspond to recession years as defined by Colombia's business cycle chronology (Arango and Ospina 2025). **Source:** Authors' calculations using PILA 2009–2024.

Increasing the number of groups yields very small marginal gains in the share of explained variation. Appendix C discusses how changes in the number of groups determine the total share of variation explained by our clustering exercise, as well as the share of between- and within-group variation in histories.

### 5.3. Groups of formal attachment

**Formal attachment.** The labor market trajectories of the five groups we identify paint a bleak picture of the Colombian formal labor market, as shown in Figure 9A. The workers in the largest group (G1, which represents 42 percent of the cohort) participate in the formal market only 9.8 percent of the year on average (just over a month), well below the cohort's average of 39 percent. This low formal attachment is persistent throughout their work lives. The second largest group (G2, 22 percent of the cohort) counteracts this with a high and steady average formal attachment of 83 percent—close to 10 months per year. Taken together, these two groups provide evidence of a segmented labor market; 64 percent of individuals have either persistently low or persistently high formal attachment.

Among the three remaining groups, two (G3 and G4, respectively, 16 and 12 percent of the cohort) show an upward trend in formal attachment that differs in its timing over the life cycle (either early or late in the cohort's working life). We return to the differences between these groups in the analysis below. The last group (G5, 8 percent of the cohort) shows a higher-than-average formal attachment of 65.5 percent (8 months per year) until 2015, when formal attachment starts to fall and then plummets after the pandemic to 5 percent.

There is, of course, substantial heterogeneity in formal attachment within each group. Two particularly informative margins are the share of group members who, in a given year, miss the formal sector entirely ( $f_{i,t} = 0$ ) and those who remain formally employed every month of the year ( $f_{i,t} = 1$ ). The share of workers with no formal contribution in a year is especially high for the low-formality group (G1), averaging 73 percent over the sample period. It is, on average, high among G3, at 46 percent. The average masks the gradual decline in the share of workers with no formal attachment, which starts at 70 percent and falls below 20 percent by the end of the sample. Meanwhile, full-year formal employment remains limited even among the high-formality group (G2): only 43 percent of its members are formally employed for all twelve months of the year. Full-year formality is also low in several other groups: as low as 1.3 percent for the low-formality group (G1), and 12 and 9.3 percent for G3 and G5, respectively.<sup>24</sup>

**Formal earnings.** While formal attachment is correlated with earnings, the two diverge for some of the groups as their formal attachment changes. For instance, the two groups with the lowest and highest formal attachment (G1 and G2) have, correspondingly, the lowest and highest average earnings over the sample. By 2024, the high-formality

---

<sup>24</sup>Figure D.1 in the Appendix shows the within-group dynamics also mirror the broader trajectory patterns. For instance, as G5's formality falls, its share of workers with no formal contribution converges toward that of G1. Conversely, as G4's formal attachment rises, its share of full-year contributors approaches that of G2.

group earns, on average, 74 percent more per month of formal employment (up from 20 percent more in 2009). The earnings trends of the groups with changing formal dynamics reveal more interesting patterns. We see this in Figure 9B, which presents average monthly earnings series for each group relative to each year's minimum wage.<sup>25</sup>

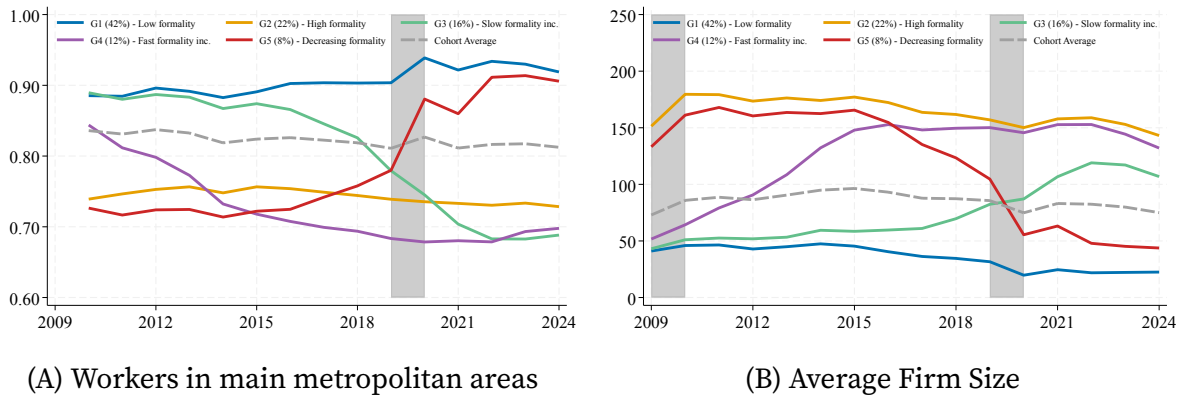
Starting with group G3, average earnings track those of the low-formality group (G1) throughout the sample, even though they only have similar formal attachment until 2015. For this group of the population, the sharp increase in formality (from having a formal job 15 percent of the year to 70 percent) is not reflected in higher monthly earnings (although their annual formal earnings naturally rise). Their story is consistent with landing more stable jobs rather than higher-paying ones. Group G4 is quite different. Their formality increases steadily and matches that of the high-formality group (G2) by 2019. At the same time, their earnings increase, but their rise does not get them to match the high earnings level of workers in group G2.

Finally, group G5 shows an interesting divergence between formality and earnings. This group starts with high formality that then decreases later in life and drops during the pandemic. Their initial earnings are lower than those of the high-formality group, but markedly higher than those of other groups (about 15 percent higher). The initial decrease in their formal attachment in 2015 coincides with a slowdown in the growth of their average earnings, but this is reversed once their formal attachment collapses. This points to a clear selection within this decreasing-formality group, with those who remain formally employed being the highest-paid workers.

---

<sup>25</sup>We first compute average monthly earnings for each individual as the sum of formal labor earnings reported in PILA across all jobs and months in the year, divided by the number of months with at least one formal job. We then express each individual's average monthly earnings in multiples of the annual national minimum monthly wage. Finally, we average these ratios across all members in each group.

FIGURE 10. Employers of Formal Workers



**Notes:** Panel A plots the share of workers who live in Colombia’s main metropolitan areas in each formal attachment group (solid lines) and in the sample cohort (dashed gray line). Panel B plots the average number of employees at the firms of formal workers of each formal attachment group (solid lines) and in the sample cohort (dashed gray line). All results are for the cohort of individuals aged 25 to 30 in 2009 who have ever reported earnings to the PILA. Shaded areas correspond to recession years as defined by Colombia’s business cycle chronology (Arango and Ospina 2025). **Source:** Authors’ calculations using PILA 2009–2024.

**Location, firms and formal attachment.** We now look at where formal workers work.<sup>26</sup> Somewhat surprisingly, workers with higher formal attachment are less likely to be located in one of Colombia’s largest 13 metropolitan areas—the standard definition of the urban sector in the country—despite our sample being heavily skewed towards urban dwellers (Figure 10A).<sup>27</sup> The differences are meaningful. Over 90 percent of workers who have a formal job for about one month a year are located in a main metropolitan area, while this share can be as low as 70 percent for workers with higher formal attachment. Changes in the location of workers also track changes in formal attachment, with workers becoming less urban as their formal attachment increases. As we discuss below, part of this difference is likely due to the role of government jobs in rural locations. These jobs are the main source of formal employment outside of large urban centers.

<sup>26</sup>We identify the location and firm of a worker in a given year as the modal location and firm of the worker among the formal employment relationships they held during the year. We break ties based on primary employment, defined as that with the highest monthly wage.

<sup>27</sup>Our sample contains workers from the cohort who have at some point held a formal job. This means we do not see workers who remain informal between 2009 and 2024. The higher degree of permanent informality outside the main urban centers means that our sample skews towards urban dwellers. Over 80 percent of workers are located in one of the main metropolitan areas when they hold a formal job.

Figure 10B reports the size of the firms that employed formal workers in each group (excluding self-employed individuals). The connection between formal attachment (Figure 9A) and firm size is remarkable: High formal attachment is closely linked to working in large firms. This suggests an important role for job ladders tied to firm size in the movement of workers toward formal jobs. As we show in Section 6, these jobs are characterized by higher and more stable earnings.

In terms of the sector and industries of employment for formal workers in each group, most are employed in the service sector, but the share of workers in this sector varies by group and over time. Notably, the share of government and social services in the service sector increases consistently, together with formal attachment. Workers in groups with low (or decreasing) formal attachment switch between services and agriculture (despite agriculture being the smallest sector in the economy).<sup>28</sup> By contrast, the share of services and manufacturing is relatively stable among the groups with high formal attachment (like G2) or during periods of high attachment for the group with increasing formality rates (like group G4 after 2019). We provide a more detailed discussion of the sectoral composition of formal jobs and its industry-level decomposition in Appendix D.

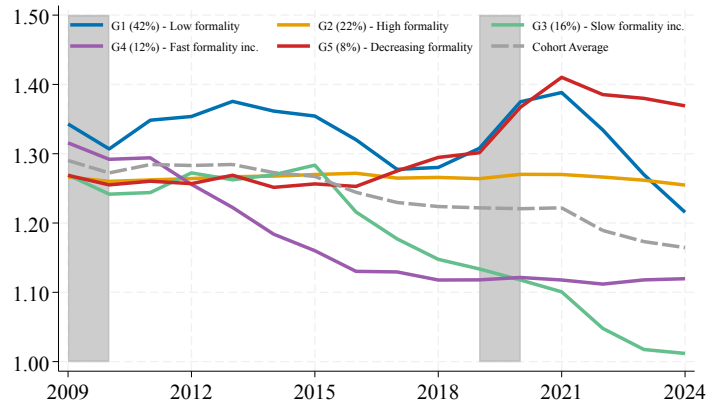
***Formal attachment of male and female workers.*** Finally, Figure 11 shows the male-to-female ratio of individuals who are formally employed in each year.<sup>29</sup> Formal employment in Colombia is disproportionately male, but increases in formal attachment over time are primarily driven by women. This is particularly noticeable in groups G3 and G4, which experience a sustained decrease in male-to-female ratios as formal attachment increases, and in the sustained differences in composition between groups G1 and G2 with the lowest and highest formality throughout the sample.

---

<sup>28</sup>Formal jobs in agriculture are mostly linked to industrial processing of agricultural produce and to large-scale agricultural activities such as sugar cane, banana, and palm oil plantations.

<sup>29</sup>The gender composition of each group is, of course, constant over time.

FIGURE 11. Gender Composition of Formal Workers



(A) Male-to-female ratio

**Notes:** The figure plots the male-to-female ratio of formal workers in each formal attachment group (solid lines) and in the sample cohort (dashed gray line). All results are for the cohort of individuals aged 25 to 30 in 2009 who have ever reported earnings to the PILA. Shaded areas correspond to recession years as defined by Colombia's business cycle chronology (Arango and Ospina 2025). **Source:** Authors' calculations using PILA 2009–2024.

Conversely, group G5 experiences the largest decline in formal attachment after the onset of the pandemic and concurrently exhibits a spike in the male-to-female ratio. In this sense, and consistent with evidence from developed countries (see, for instance, Albanesi and Kim 2021; Alon, Coskun, Doepke, Koll, and Tertilt 2022), the collapse in formality in group G5 during the pandemic is predominantly a female phenomenon, with the majority of the remaining formal workers in this group being men—the male-to-female ratio spikes from just below 1.25 in 2015 to over 1.6 in 2020.

#### 5.4. Taking stock of the dynamics of formality

The dynamics of formal attachment, characterized by our clustering exercise, can be summarized as follows. The two groups with the largest shares of workers are highly segmented, with the largest, G1, having the lowest attachment (9.8 percent) and the second largest, G2, having the highest average attachment to formality (83 percent). In the next section, we show that the differences in formal attachment across these groups are an important driver of the dispersion in labor income. In particular, workers

in group G2 have more stable labor market trajectories, along with higher earnings growth, while those in group G1 have low and volatile earnings when comparing their short formal spells across years.

Next, two groups, G3 and G4, exhibit clear increases in formal attachment and earnings over time. Part of the increase in formal attachment in these two groups reflects the life cycle dynamics of participation in Colombia: workers reach the peak of participation in the labor market around age 30 ([Grupo de Análisis del Mercado Laboral 2025](#)). The public and private service industries become more prevalent in these two groups as their formal attachment rises. Finally, there is a group, G5, comprising 8 percent of the cohort, that experienced catastrophic and long-lasting effects from the pandemic. Their formal attachment began to fall after 2015 but plummeted after 2020, never to recover. It is plausible that firm closures in these industries explain the reduction in formality in G5.<sup>30</sup>

## **6. Earnings dispersion and earnings dynamics across groups**

We now revisit the patterns of earnings dispersion and earnings dynamics described in Section 4 using our formal-attachment groups. We find that both the dispersion and volatility of earnings are driven by groups with low formal attachment. Earnings inequality is lower among individuals with high formal attachment, who also experience greater earnings stability over the course of their work lives.

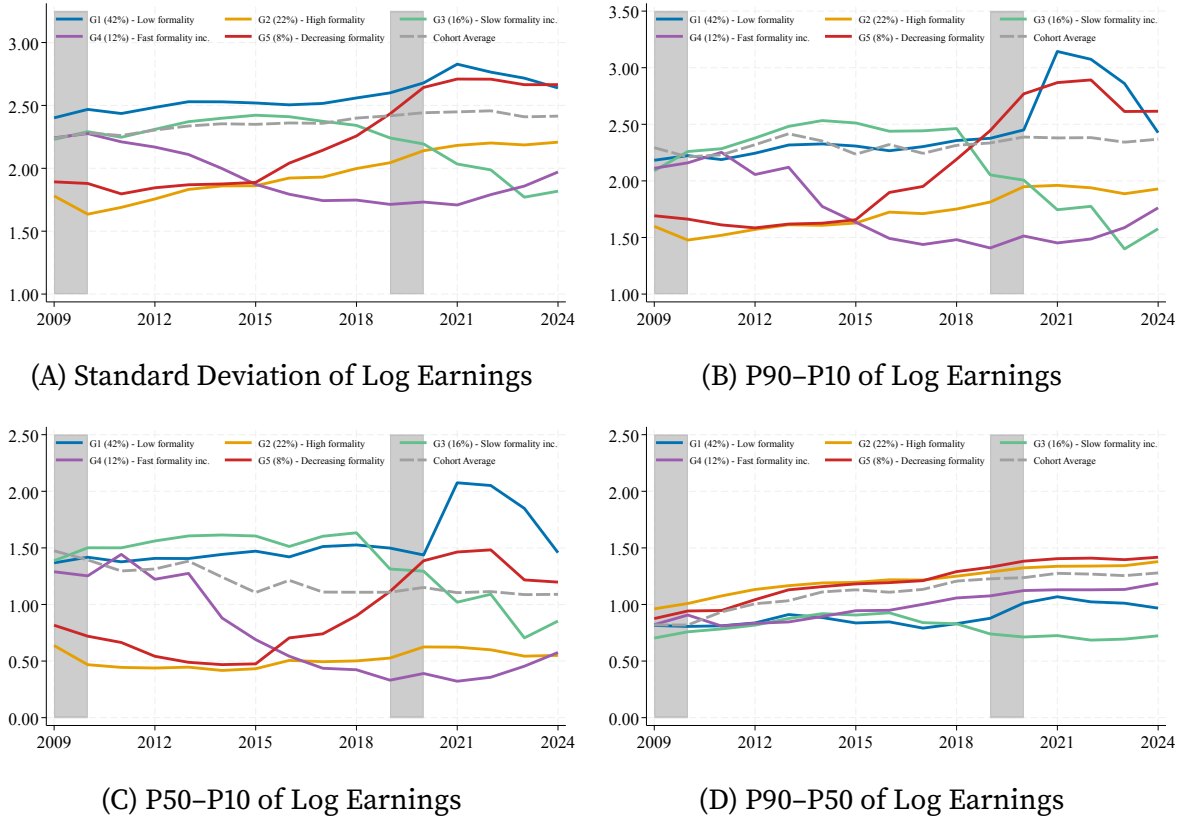
Furthermore, our results reveal an important relationship between within-group earnings dispersion and the dynamics of formal attachment. Groups with increasing attachment (G3 and G4) also have clear reductions in earnings dispersion and volatility. Conversely, group G5 experiences a clear reduction in formal attachment accompanied

---

<sup>30</sup>These sectors were substantially exposed to lockdowns during the COVID-19 crisis, resulting in negative consequences ([Morales-Zurita, Bonilla-Mejía, Pulido, Flórez, Hermida, Pulido-Mahecha, and Lasso-Valderrama 2022](#); [Cortés Cortés, Posso, and Villamizar-Villegas 2022](#)).

by an increase in the dispersion of formal earnings among its members (particularly in the lower half of their earnings distribution). This highlights the prominent role of transitions in and out of the formal labor market (whether through informality or unemployment) in driving earnings dynamics.

FIGURE 12. Earnings Dispersion Across Formal Attachment Groups



**Notes:** Panel A plots the standard deviation of log-earnings for formal workers in each formal attachment group (solid lines) and in the sample cohort (dashed gray line). Panel B plots the P90–P10 differential of log-earnings for formal workers in each formal attachment group (solid lines) and in the sample cohort (dashed gray line). Panels C and D plot the corresponding P50–P10 and P90–P50 differentials. All results are for the cohort of individuals aged 25 to 30 in 2009 who have ever reported earnings to the PILA. Shaded areas correspond to recession years as defined by Colombia’s business cycle chronology (Arango and Ospina 2025). **Source:** Authors’ calculations using PILA 2009–2024.

## 6.1. Earnings inequality

We begin by looking at log-earnings dispersion across formal attachment groups. Panels A and B of Figure 12 present the standard deviation and the P90–P10 differential of log-earnings for the cohort as a whole and for each group. Both measures are relatively

stable over the cohort's working life, at levels similar to those of Colombian formal workers, as shown in Figure 5.

Aggregate log-earnings dispersion is mostly driven by the dispersion in earnings among individuals with low formal attachment, even as the monthly earnings differ markedly across groups (see Figure 9). Group G1, with consistently low formal attachment, has the highest dispersion. This dispersion in log-earnings is matched by group G3 until 2015, when it had similarly low levels of formal attachment. After 2015, the group's formal attachment starts rising and the dispersion in log-earnings starts falling. Group G5 starts with higher formal attachment and with a level of log-earnings volatility and dispersion comparable to that of G2, the group with the highest formal attachment throughout the sample. Nevertheless, as formal attachment falls for group G5, log-earnings volatility and dispersion increase, ultimately matching the levels of the low-formality group, G1. Conversely, the volatility and dispersion of log-earnings decrease in groups G3 and G4, coinciding with the timing of their increases in formal attachment.

Panels C and D of Figure 12 show that the patterns we just described are driven mostly by the lower half of the log-earnings distribution. Individuals in groups with lower formal attachment exhibit greater dispersion among their low earners (see the P50–P10 log-differential in Panel C). Earnings dispersion above the median is much more similar across all groups, although higher for the high formality groups (see the P90–P50 log-differential in Panel D). Notably, it is the lower half of the distribution that compresses as individuals in a group increase their formal attachment (see groups G3 and G4) and increases as formal attachment decreases (as for group G5), while the dispersion in the upper half of the distribution grows at a more constant rate over time.

## 6.2. Earnings dynamics

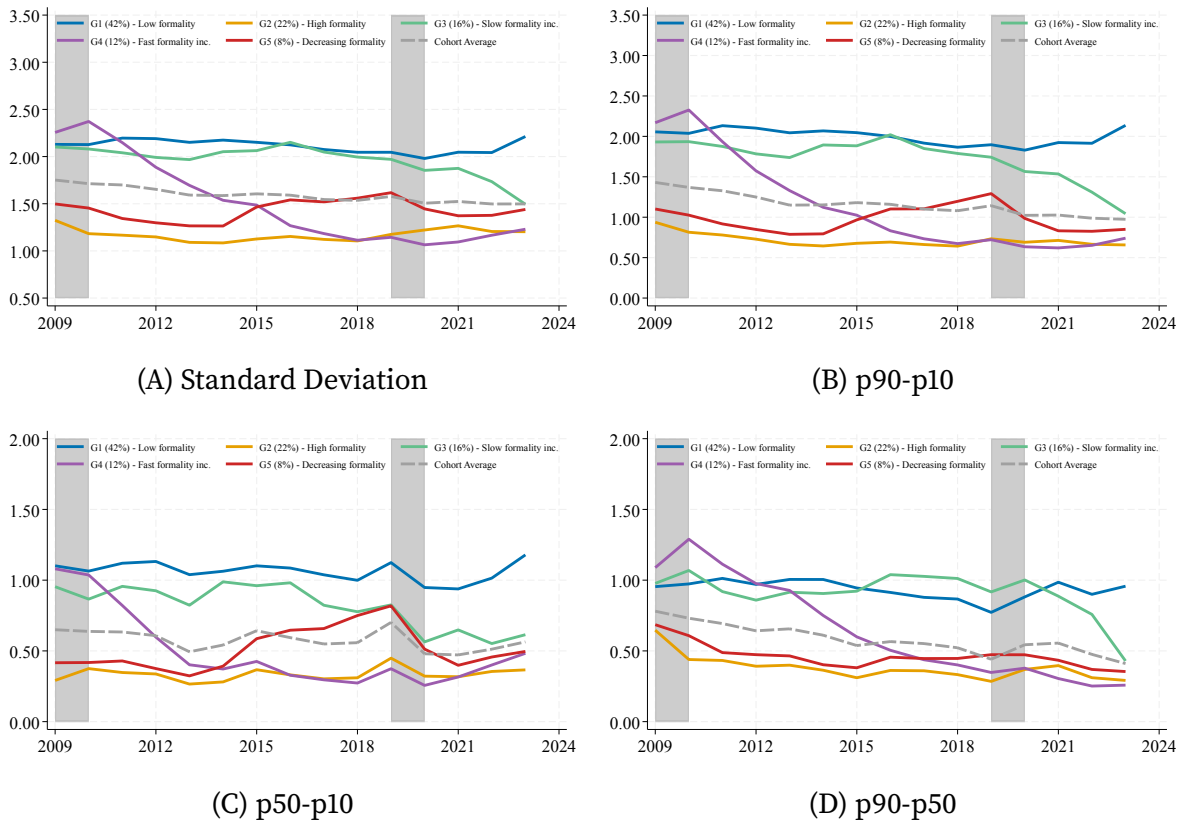
The volatility of annual log-earnings growth displays patterns similar to those of log-earnings inequality. Groups with low formal attachment have the highest volatility in earnings growth, as we show in Panels A and B of Figure 13. The role of formal attachment in reducing log-earnings growth volatility is particularly noticeable for group G4 (in purple). This is the group that exhibits the largest and most sustained rise in formality. The volatility of its log-earnings growth decreases sharply as formal attachment increases, and log-earnings dispersion within the group decreases. Meanwhile, the group with the lowest formal attachment (G1) has the highest log-earnings volatility throughout the sample. The main difference from our previous results is that the dispersion of log-earnings growth of group G5 (with decreasing formality) does not increase as the dispersion of log-earnings levels does.

Panels C and D of Figure 13 show the dispersion of log-earnings growth for the upper and lower halves of the log-earnings distribution. Unlike the dispersion in log-earnings levels, the dispersion of log-earnings growth follows the same patterns above and below the median. Again, attachment to the formal sector is negatively correlated with log-earnings growth dispersion.

## 7. Conclusions

We documented the main features of earnings dynamics and their relationship to individuals' attachment to the formal sector across their work lives, using administrative data on the universe of formally employed workers in Colombia between 2009 and 2024. In doing so, we provided a new characterization of participation in formal labor markets, grounded in the long-run patterns of individuals' *attachment*—the fraction of time that an individual is employed in the formal sector. The longitudinal dimension of the

FIGURE 13. Dispersion of One-Year Log-Earnings Changes Across Formal Attachment Groups



**Notes:** Panel A plots the standard deviation of the one-year forward change in residualized log-earnings for formal workers in each formal attachment group (solid lines) and in the sample cohort (dashed gray line). Panel B plots the P90–P10 differential of the one-year forward change in residualized log-earnings for the workers of each formal attachment group (solid lines) and in the sample cohort (dashed gray line). Panels C and D plot the corresponding P50–P10 and P90–P50 differentials. All results are for the cohort of individuals aged 25 to 30 in 2009 who have ever reported earnings to the PILA. Shaded areas correspond to recession years as defined by Colombia’s business cycle chronology (Arango and Ospina 2025). **Source:** Authors’ calculations using PILA 2009–2024.

Colombian administrative data is crucial for this. Our exercise allows us to understand how low-frequency trends in formal attachment shape the high-frequency changes in earnings widely studied in the income dynamics literature.

During the 16-year period we study (2009–2024), individuals employed in Colombia’s formal sector experienced consistent earnings growth, with a right shift in the aggregate earnings distribution. This overall trend implied modest decreases in earnings concentration. Earnings dispersion among formally employed workers was comparable to that of the U.S., Brazil, and Argentina, placing Colombia near the

median level of dispersion among countries in the *Global Repository of Income Dynamics*. Consistent with international patterns, dispersion among the bottom half of the earnings distribution is markedly and persistently greater than among the upper half. Earnings volatility among workers who are in formal jobs is markedly lower than in other countries. For instance, the P90–P10 differential of the one-year change in residualized earnings is 0.3 log points lower than in the U.S. and 0.4 log points lower than in France.

Our analysis of individuals' work histories reveals that most of the dispersion in earnings levels and growth rates is concentrated among individuals with persistently low formal attachment. These individuals have a formal job for 1–3 months a year and tend to earn income close to the minimum legal monthly wage, in contrast to groups of individuals who hold formal jobs for 8–10 months a year and tend to experience consistent earnings growth over their work lives. In this way, the Colombian labor market is segmented (almost in half) between groups that enjoy high or increasing formality, paired with growing and stable earnings, and groups that switch in and out of formal jobs, with correspondingly low and volatile earnings.

## References

- Ahn, Hie Joo, Bart Hobijn, and Ayşegül Şahin. 2023. “The dual U.S. labor market uncovered.” Finance and Economics Discussion Series 2023–031, Board of Governors of the Federal Reserve System. URL <https://doi.org/10.17016/FEDS.2023.031>.
- Albanesi, Stefania and Jiyeon Kim. 2021. “Effects of the COVID-19 recession on the US labor market: Occupation, family, and gender.” *Journal of Economic Perspectives* 35 (3):3–24. URL <https://doi.org/10.1257/jep.35.3.3>.
- Alon, Titan, Sena Coskun, Matthias Doepke, David Koll, and Michèle Tertilt. 2022. “From mancession to ahcession: Women’s employment in regular and pandemic recessions.” *NBER Macroeconomics Annual* 36:83–151. URL <https://doi.org/10.1086/718660>.
- Arango, Luis E and Luz A Flórez. 2021. “Regional labour informality in Colombia and a proposal for a differential minimum wage.” *The Journal of Development Studies* 57 (6):1016–1037. URL <https://doi.org/10.1080/00220388.2020.1841170>.
- Arango, Luis Eduardo and Juan José Ospina, editors. 2025. *Características cuantitativas de los ciclos económicos en Colombia*, vol. 110. Ensayos sobre Política Económica (ESPE). URL <https://repositorio.banrep.gov.co/items/744319b1-6d8c-40b5-aa6a-7829e822b1c7>.
- Astorga-Rojas, Diego, Nicolás Guida-Johnson, and Leonardo Fabio Morales-Zurita. 2026. “Evaluating Unemployment Insurance in a Developing Economy: Evidence from Colombia.” Working Paper.
- Audoly, Richard, Rory McGee, Sergio Ocampo, and Gonzalo Paz Pardo. 2025. “The life-cycle dynamics of wealth mobility.” NY FED Staff Report 1097, Federal Reserve Bank of New York. URL <https://doi.org/10.59576/sr.1097>.
- Banco de la República. 2016. “El mercado laboral colombiano: contraste de la GEIH y la PILA con otras fuentes de información.” Reportes del Emisor 205, Banco de la República de Colombia, Bogotá, D.C. URL <https://www.banrep.gov.co/es/publicaciones/reportes-emisor-205>. Editora: María Teresa Ramírez Giraldo. ISSN 0124-0625.
- Becerra-Camargo, Oscar Reinaldo and Leonardo Fabio Morales-Zurita. 2025. “Labor demand responses to payroll taxes in an economy with wage rigidity: Evidence from colombia.” Borradores de Economía 1297, Banco de la República de Colombia. URL <https://repositorio.banrep.gov.co/server/api/core/bitstreams/bfe0f425-9c47-4f2b-bc33-f1bef26d9a2f/content>.
- Bernal, Raquel, Marcela Eslava, Marcela Meléndez, and Alvaro Pinzón. 2017. “Switching from Payroll Taxes to Corporate Income Taxes: Firms’ Employment and Wages after the 2012 Colombian Tax Reform.” *Economía* 18 (1):41–74. URL <https://doi.org/10.31389/eco.51>.
- Bonilla-Mejía, Leonardo, Luz A Florez, Didier Hermida, Francisco Lasso-Valderrama, Leonardo Fabio Morales, and José Pulido. 2025. “Payroll Subsidies for SMEs in Informal Labor Markets.” Tech. rep., Banco de la Republica de Colombia.
- Bosch, Mariano and William F. Maloney. 2008. “Cyclical Movements in Unemployment and Informality in Developing Countries.” Policy Research Working Paper 4648, The World Bank, Washington, DC. URL <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/911311468049138854>.
- Carranza, Rafael, Joaquín Prieto, and Kirsten Sehnbruch. 2025. “Job loss and earnings inequality:

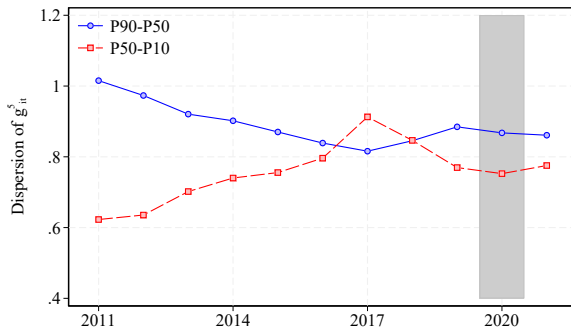
- Distributional effects of formal re-employment in Chile.” *Economic Analysis and Policy* 86:1020–1036. URL <https://doi.org/10.1016/j.eap.2025.02.021>.
- Cortés Cortés, Darwin, Christian Posso, and Mauricio Villamizar-Villegas. 2022. *COVID19: Consecuencias y desafíos en la economía colombiana. Una mirada desde las universidades*. Editorial Universidad del Rosario. URL <https://doi.org/10.12804/urosario9789587848496>.
- Donovan, Kevin, Will Jianyu Lu, and Todd Schoellman. 2023. “Labor market dynamics and development.” *The Quarterly Journal of Economics* 138 (4):2287–2325. URL <https://doi.org/10.1093/qje/qjad019>.
- Eslava, Marcela, John C. Haltiwanger, and Álvaro Pinzón. 2022. “Job Creation in Colombia Versus the USA: “Up-or-out Dynamics” Meet “The Life Cycle of Plants”.” *Economica* 89 (355):511–539. URL <https://doi.org/10.1111/ecca.12418>.
- Fernández, Cristina and Leonardo Villar. 2017. “The impact of lowering the payroll tax on informality in Colombia.” *Economía* 18 (1):125–155. URL <https://doi.org/10.31389/eco.53>.
- Flórez, Luz A, Leonardo Fabio Morales-Zurita, Daniel Medina, and José Lobo. 2021. “Labor flows across firm size, age, and economic sector in Colombia vs. the United States.” *Small Business Economics* 57 (3):1569–1600. URL <https://doi.org/10.1007/s11187-020-00362-8>.
- Gregory, Victoria, Guido Menzio, and David Wiczer. 2025. “The alpha beta gamma of the labor market.” *Journal of Monetary Economics* 150:103695. URL <https://doi.org/10.1016/j.jmoneco.2024.103695>.
- Grupo de Análisis del Mercado Laboral. 2024. “Pérdida de dinamismo en la ocupación y el impacto del programa de certificación de habilidades del SENA.” Reportes del Mercado Laboral 29, Banco de la República de Colombia. URL <https://doi.org/10.32468/rml.29>. Editor, Leonardo Fabio Morales-Zurita.
- . 2025. “Crecimiento del empleo liderado por el área rural y dinámica de la participación laboral.” Reportes del Mercado Laboral 35, Banco de la República de Colombia. URL <https://doi.org/10.32468/rml.35>. Editor, Leonardo Fabio Morales-Zurita.
- Güvenen, Fatih, Serdar Ozkan, and Jae Song. 2014. “The Nature of Countercyclical Income Risk.” *Journal of Political Economy* 122 (3):621–660. URL <https://doi.org/10.1086/675535>.
- Güvenen, Fatih, Luigi Pistaferri, and Giovanni L. Violante. 2022a. “The Global Repository of Income Dynamics.” <https://www.grid-database.org>.
- . 2022b. “Global trends in income inequality and income dynamics: New insights from GRID.” *Quantitative Economics* 13 (4):1321–1360. URL <https://doi.org/10.3982/QE2260>.
- Hastie, Trevor, Robert Tibshirani, and Jerome H Friedman. 2009. *The elements of statistical learning: Data mining, inference, and prediction*, vol. 2. New York, NY: Springer.
- Herreño, Juan and Sergio Ocampo. 2023. “The macroeconomic consequences of subsistence self-employment.” *Journal of Monetary Economics* 136:91–106. URL <https://doi.org/10.1016/j.jmoneco.2023.02.002>.
- Humphries, John Eric. 2022. “The causes and consequences of self-employment over the life cycle.” URL [https://johnerichumphries.com/Humphries\\_SelfEmployment.pdf](https://johnerichumphries.com/Humphries_SelfEmployment.pdf). Yale.
- Jolly, Nicholas A. 2013. “Job Displacement and the Inter-Temporal Movement of Workers Through the Earnings and Income Distributions.” *Contemporary Economic Policy* 31 (2):392–406. URL <https://doi.org/10.1111/j.1465-7287.2011.00311.x>.

- Kugler, Adriana, Maurice Kugler, and Luis Omar Herrera Prada. 2017. "Do payroll tax breaks stimulate formality? Evidence from Colombia's reform." *Economía* 18 (1):3–40. URL <https://doi.org/10.31389/eco.50>.
- Kugler, Adriana D., Maurice Kugler, Laura Ripani, and Rodimiro Rodrigo. 2020. "U.S. Robots and Their Impacts in the Tropics: Evidence from Colombian Labor Markets." NBER Working Paper 28034, National Bureau of Economic Research. URL <https://www.nber.org/papers/w28034>.
- Lagakos, David, Stelios Michalopoulos, and Hans-Joachim Voth. 2025. "American life histories." Working Paper 33373, National Bureau of Economic Research. URL <https://doi.org/10.3386/w33373>.
- Menezes-Filho, Naercio and Renata Narita. 2025. "Labor market turnover and inequality in Latin America." *Oxford Open Economics* 4 (1):349–375. URL <https://doi.org/10.1093/ooec/odae027>.
- Morales, Leonardo Fabio and Carlos Medina. 2017. "Assessing the effect of payroll taxes on formal employment: The case of the 2012 tax reform in Colombia." *Economía* 18 (1):75–124. URL <https://doi.org/10.31389/eco.52>.
- Morales-Zurita, Leonardo Fabio, Leonardo Bonilla-Mejía, Jose Pulido, Luz A Flórez, Didier Hermida, Karen L Pulido-Mahecha, and Francisco Lasso-Valderrama. 2022. "Effects of the COVID-19 pandemic on the Colombian labour market: Disentangling the effect of sector-specific mobility restrictions." *Canadian Journal of Economics* 55:308–357. URL <https://doi.org/10.1111/caje.12549>.
- Otero-Cortés, Andrea, editor. 2025. *Nueva evidencia sobre la informalidad laboral y empresarial en Colombia*, vol. 108. Ensayos sobre Política Económica (ESPE). URL <https://repositorio.banrep.gov.co/server/api/core/bitstreams/6db78584-da96-4f8b-a0c9-fb5a306cdffc/content>.
- Puggioni, Daniela, Mariana Calderón, Alfonso Cebreros Zurita, León Fernández Bujanda, José Antonio Inguanzo González, and David Jaume. 2022. "Inequality, income dynamics, and worker transitions: The case of Mexico." *Quantitative Economics* 13 (4):1669–1705. URL <https://doi.org/10.3982/QE1863>.
- Samaniego de la Parra, Brenda and León Fernández Bujanda. 2024. "Increasing the Cost of Informal Employment: Evidence from Mexico." *American Economic Journal: Applied Economics* 16 (1):377–411. URL <https://www.aeaweb.org/articles?id=10.1257/app.20200763>.
- Samaniego de la Parra, Brenda, Andrea Otero-Cortés, and Leonardo F. Morales-Zurita. 2024. "The labor market effects of facilitating social security contributions under part-time employment contracts: Evidence from Colombia." *Labour Economics* 91:102630. URL <https://doi.org/10.1016/j.labeco.2024.102630>.
- Schoar, Antoinette. 2010. "The divide between subsistence and transformational entrepreneurship." *Innovation Policy and the Economy* 10:57–81. URL <https://doi.org/10.1086/605853>.
- Ulyssea, Gabriel. 2020. "Informality: Causes and consequences for development." *Annual Review of Economics* 12:525–546. URL <https://doi.org/10.1146/annurev-economics-082119-121914>.

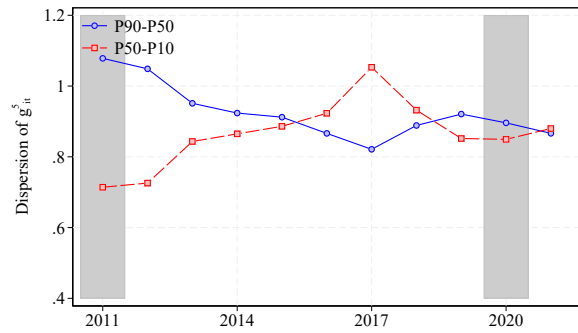
## **Appendix A. Five-year changes in earnings**

Turning to the dispersion of earnings growth driven by more permanent shocks, Figure [A.1](#) presents analogous statistics to those in Section [4.2](#) for the five-year changes in earnings. The dispersion of “downside shocks” is again initially below that of “upside shocks,” although the gap in dispersion between the bottom and the upper half of the distribution is now significantly larger at the beginning of the sample period. The upper and lower spreads for earnings changes start converging after the recovery from the 2009–2011 recession. While five-year earnings growth was initially positively skewed (with above-median earnings changes contributing more to overall dispersion), this asymmetry declined over time and stabilized around the onset of the pandemic. The skewness of five-year earnings growth thus declined from large positive values (around 0.25) and hovered around zero after 2017.

FIGURE A.1. Distribution of 5-Year Residual Log-Earnings Growth  
Dispersion Above and Below Median

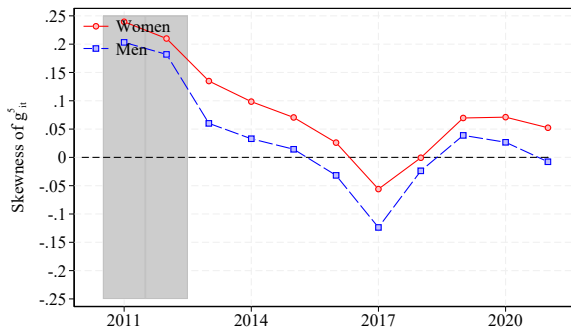


(A) Women

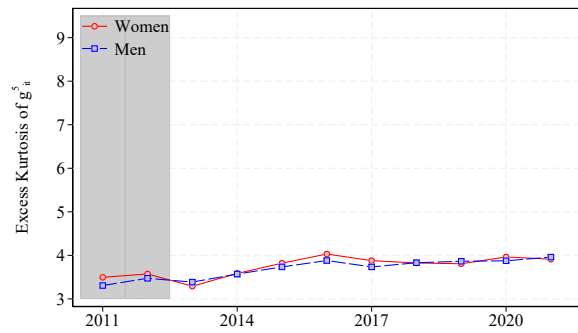


(B) Men

Kelley Skewness and Crow-Siddiqui Excess Kurtosis



(C) Skewness

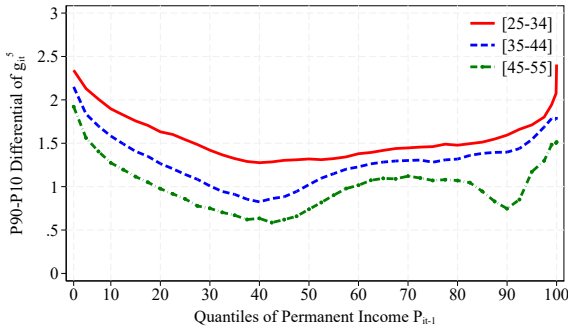


(D) Kurtosis

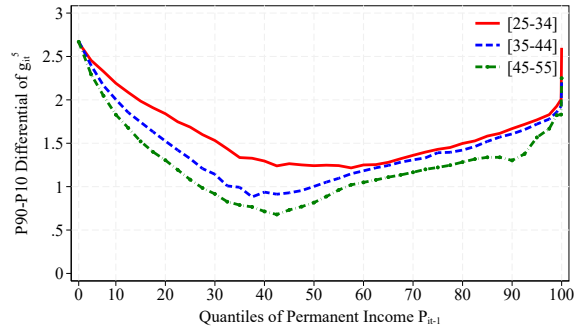
**Notes:** Panels A and B plot the P90–P50 differential of the five-year change in residualized log-earnings (blue circles) along with the corresponding P50–P10 differential (red squares) for formal workers in each year. Panel C plots the Kelley Skewness measure of the five-year change in residualized log-earnings for women (red circles) and men (blue squares). Panel D plots the Crow-Siddiqui Excess Kurtosis measure of the five-year change in residualized log-earnings for women (red circles) and men (blue squares). All results are for workers aged 25 to 55. Shaded areas correspond to recession years as defined by Colombia's business cycle chronology (Arango and Ospina 2025). **Source:** authors' calculations using PILA 2009–2024.

FIGURE A.2. Dispersion of Residualized Log-Earnings Growth by Permanent Income

P90–P10 Differential

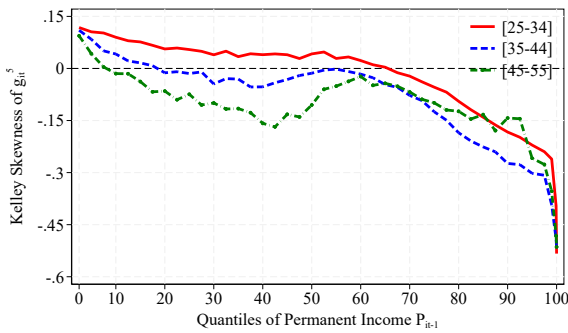


(A) Women

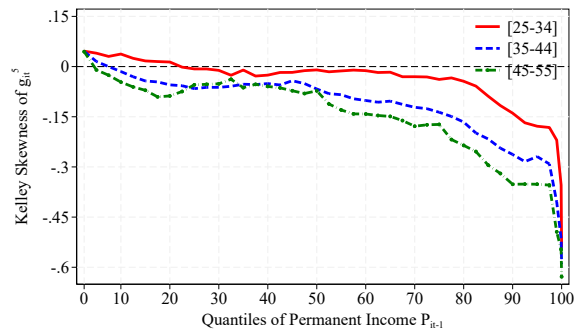


(B) Men

Kelley Skewness

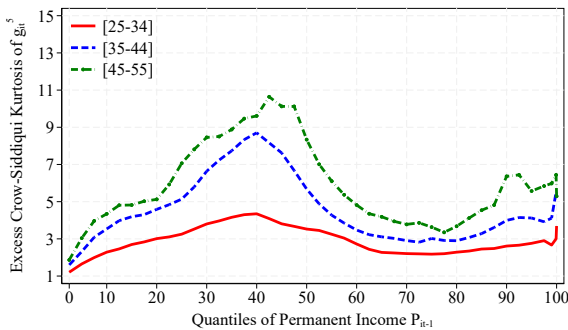


(C) Women

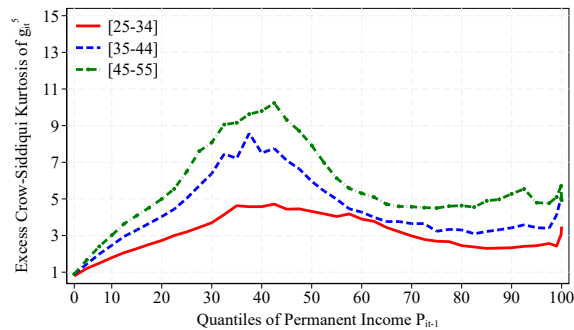


(D) Men

Crow-Siddiqui Excess Kurtosis



(E) Women



(F) Men

**Notes:** Panels present moments of the distribution of the five-year forward change in residualized log-earnings by quantile of the permanent income  $P_{i,t-1}$  and for individuals aged 25–34 (solid red line), 35–44 (dashed blue line), and 45–55 (dot-dashed green line). Permanent income is computed as the three-year average of earnings. All results are for workers aged 25 to 55. Panels A and B plot the P90–P10 differential for women and men, respectively. Panels C and D plot the Kelley skewness measure for women and men, respectively. Panels E and F plot the Crow-Siddiqui excess kurtosis measure for women and men, respectively. **Source:** authors' calculations using PILA 2009–2024.

## Appendix B. Example of agglomerative hierarchical clustering

This appendix provides a tractable example of the agglomerative hierarchical clustering algorithm described in Section 5. To make the mechanics transparent, suppose that we observe six workers for three periods (our empirical application uses sixteen-year labor market trajectories of formal attachment for millions of workers). Each trajectory  $F_i$  records the share of months in each period in which worker  $i$  is formally employed.

| Trajectory | $t = 1$ | $t = 2$ | $t = 3$ | Description                             |
|------------|---------|---------|---------|---|
| $F_1$      | [0.00   | 0.10    | 0.20]   | Low formal attachment, slight increase  |
| $F_2$      | [0.75   | 0.80    | 0.85]   | High formal attachment                  |
| $F_3$      | [0.20   | 0.50    | 0.85]   | Increasing formal attachment            |
| $F_4$      | [0.05   | 0.10    | 0.20]   | Low formal attachment, close to $F_1$   |
| $F_5$      | [0.85   | 0.80    | 0.70]   | High formal attachment, slight decrease |
| $F_6$      | [0.85   | 0.45    | 0.10]   | Decreasing formal attachment            |

The distance between two singleton trajectories is the Euclidean distance between the two vectors. For example,  $F_1$  and  $F_4$  are close to each other because they both describe workers with persistently low formal attachment:

$$\|F_1 - F_4\|_2 = \sqrt{(0.00 - 0.05)^2 + (0.10 - 0.10)^2 + (0.20 - 0.20)^2} = 0.050.$$

By contrast,  $F_1$  and  $F_2$  are far apart because one trajectory has low attachment throughout, while the other has high attachment:

$$\|F_1 - F_2\|_2 = \sqrt{(0.00 - 0.75)^2 + (0.10 - 0.80)^2 + (0.20 - 0.85)^2} = 1.214.$$

The lower triangular matrix below reports all pairwise distances among the six initial trajectories:

|       | $F_1$ | $F_2$ | $F_3$ | $F_4$ | $F_5$ | $F_6$ |
|-------|-------|-------|-------|-------|-------|-------|
| $F_1$ | -     |       |       |       |       |       |
| $F_2$ | 1.214 | -     |       |       |       |       |
| $F_3$ | 0.789 | 0.626 | -     |       |       |       |
| $F_4$ | 0.050 | 1.184 | 0.778 | -     |       |       |
| $F_5$ | 1.209 | 0.180 | 0.731 | 1.175 | -     |       |
| $F_6$ | 0.925 | 0.834 | 0.994 | 0.879 | 0.695 | -     |

The closest pair is  $F_1$  and  $F_4$ , with a distance of 0.050. The first step of the algorithm, therefore, merges these two trajectories. The new cluster is represented by its centroid, the element-by-element average of the merged trajectories:

$$F_{14} = \frac{F_1 + F_4}{2} = (0.025, 0.100, 0.200).$$

Once two trajectories have been merged, the distance between clusters is computed using the Ward criterion from equation (2). The matrix below shows the updated distances after replacing  $F_1$  and  $F_4$  with the merged cluster  $F_{14}$ :

|          | $F_{14}$ | $F_2$ | $F_3$ | $F_5$ | $F_6$ |
|----------|----------|-------|-------|-------|-------|
| $F_{14}$ | -        |       |       |       |       |
| $F_2$    | 1.385    | -     |       |       |       |
| $F_3$    | 0.904    | 0.626 | -     |       |       |
| $F_5$    | 1.376    | 0.180 | 0.731 | -     |       |
| $F_6$    | 1.041    | 0.834 | 0.994 | 0.695 | -     |

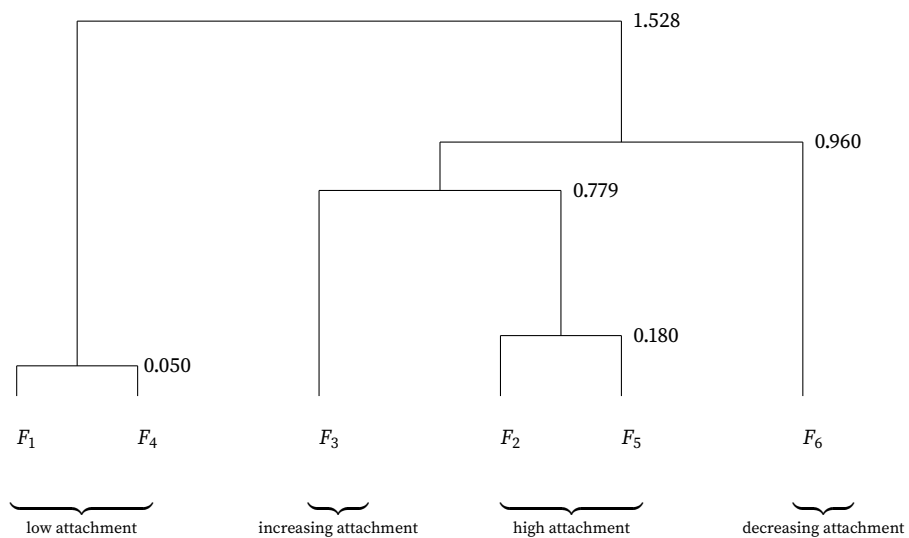
At this stage, the closest pair is  $F_2$  and  $F_5$ , with distance 0.180. So, the algorithm merges them next.

The process continues by repeatedly merging the two clusters with the lowest Ward distance and updating the centroids. In this example, the full merge order is:

| Step | Merged clusters                             | Ward distance |
|------|---|---------------|
| 1    | $F_1$ and $F_4$                             | 0.050         |
| 2    | $F_2$ and $F_5$                             | 0.180         |
| 3    | $F_3$ and $\{F_2, F_5\}$                    | 0.779         |
| 4    | $F_6$ and $\{F_3, F_2, F_5\}$               | 0.960         |
| 5    | $\{F_1, F_4\}$ and $\{F_6, F_3, F_2, F_5\}$ | 1.528         |

Figure B.1 displays the corresponding clustering tree. The leaves are the original trajectories, and the horizontal connections show the order in which groups are merged. The vertical scale is schematic, but the labels report the Ward distance at each merge.

FIGURE B.1. Agglomerative Hierarchical Clustering Tree in the Toy Example



**Notes:** The figure shows the merge sequence for the toy trajectories in the example. Merge labels report the Ward distance at which the corresponding clusters are joined.

## Appendix C. Selecting groups of formal attachment

Figure C.1 shows the variation explained for the first 12 levels of the hierarchy. We select  $G = 5$  groups that explain 60 percent of the total variation in the sample. Increasing the number of groups yields very small marginal gains in the share of explained variation, while adding significantly to the cognitive burden of the exercise. Moreover, any increase in the number of groups is obtained by dividing one of the 5 main groups, which already capture the main patterns of formal attachment.

Figure C.1 also shows measures of the share of the *between* and *within* group variation explained at different levels of the clustering hierarchy. The between  $R^2$  is an indicator of whether the groups are different from each other. It captures how dissimilar the typical trajectories are across groups as measured by the cross-sectional variation in formal attachment explained by the typical trajectories of each group (after averaging over the longitudinal dimension of the panel). Specifically, the between  $R^2$  is

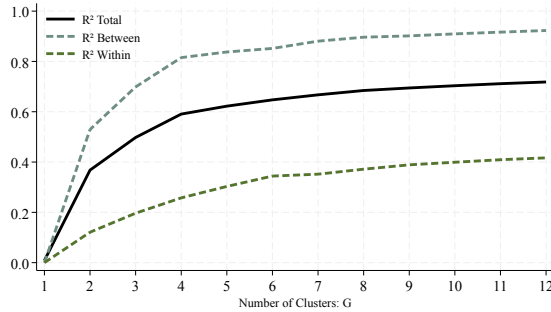
$$R_{between}^2 = 1 - \frac{\sum_i (\tilde{f}_i - \bar{F}^{g(i)})^2}{\sum_i (\tilde{f}_i - \bar{f}_i)^2}, \quad (C.1)$$

where  $\tilde{f}_i = \sum_t f_{i,t}/T$  is the within-worker average formality between 2009 and 2024,  $\bar{f}_i$  its average across individuals, and  $\bar{F}_i^{g(i)}$  its average for cluster  $g(i)$ . The between  $R^2$  for  $G = 5$  groups is over 0.80, indicating that the number of groups we select already captures most of the differences in typical formal attachment histories.

The within  $R^2$  measures within-group heterogeneity. It is the share of the variation in formal attachment along the longitudinal dimension of the panel explained by the clusters' typical trajectories and is computed as

$$R_{within}^2 = 1 - \frac{\sum_i \sum_t ((f_{i,t} - \tilde{f}_i) - (\bar{F}_t^{g(i)} - \bar{f}_i^{g(i)}))^2}{\sum_i \sum_t ((f_{i,t} - \tilde{f}_i) - (\bar{f} - \bar{f}_i))^2}. \quad (C.2)$$

FIGURE C.1. Selecting Main Groups of Formal Attachment



(A) Share of total variation explained by groups

**Notes:** The figure plots the share of the total variation in formal attachment explained at different levels of the hierarchy of clusters obtained from (2). The total  $R^2$  corresponds to equation (3). The between and within  $R^2$  compare the variation across and within groups. See definitions in equations (C.1) and (C.2), respectively.

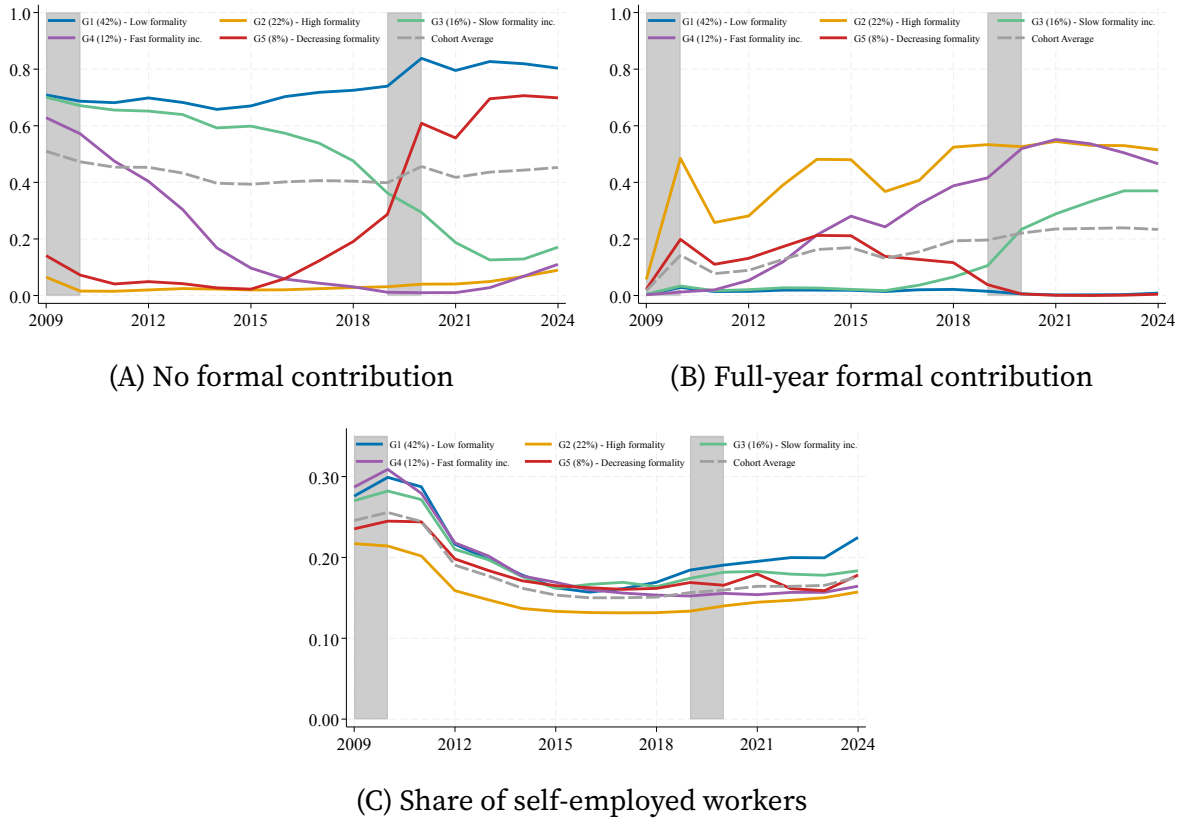
We define the deviation of an individual’s formal attachment in  $t$  relative to their average attachment over time,  $f_{i,t} - \tilde{f}_i$ , and contrast it with the population-wide average deviation in formality,  $\bar{f} - \tilde{f}_i$ . The cluster’s deviation,  $\bar{f}_t^{g(i)} - \bar{F}^{g(i)}$ , measures the within variation explained by the clusters’ typical trajectories. The within  $R^2$  is remarkably high for  $G = 5$  at 0.25, given the large variation in individual histories of formal attachment. This is reassuring for the representativeness of the five groups we focus on in the main text.

## Appendix D. Formal attachment: Additional results

**Self-employment and formal attachment.** Self-employment is prevalent across all groups. About 18 percent of formal workers are self-employed, with higher shares at the beginning and end of the sample (Figure D.1C). Although the differences between groups in self-employment rates are small, self-employment is higher throughout for the group with the lowest formal attachment (G1) and lowest for the group with the highest formal attachment (G2).

After 2020, the two groups with low formal attachment (G1 and G5, 50 percent of the cohort) exhibit an increase in self-employment. This, together with the earnings

FIGURE D.1. Extensive Margins of Formal Attachment

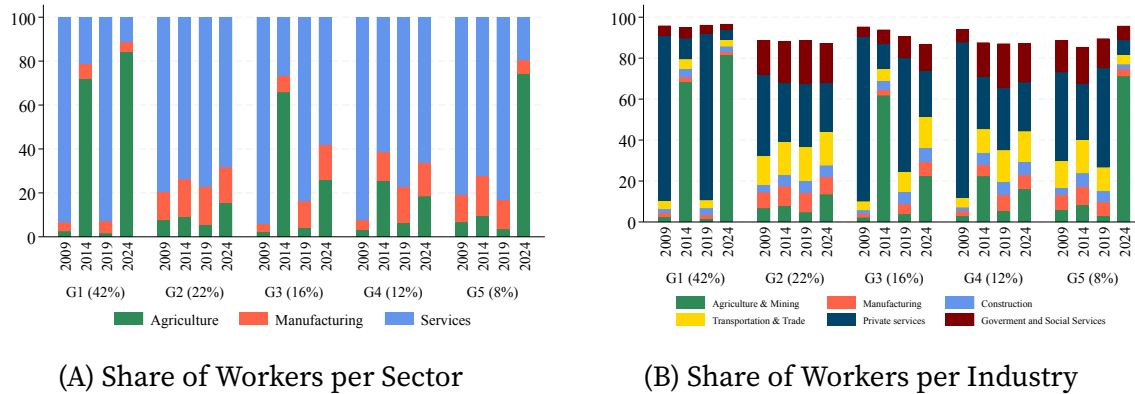


**Notes:** Panel A plots the share of workers in each formal attachment group with no formal contribution in a year,  $f_{i,t} = 0$ . Panel B plots the share of workers in each formal attachment group who contribute during the full year,  $f_{i,t} = 1$ . Panel C plots the share of self-employed workers among formal workers in each formal attachment group (solid lines) and in the sample cohort (dashed gray line). All results are for the cohort of individuals aged 25 to 30 in 2009 who have ever reported earnings to the PILA. Shaded areas correspond to recession years as defined by Colombia's business cycle chronology (Arango and Ospina 2025). **Source:** Authors' calculations using PILA 2009–2024.

patterns in Figure 9B, suggests that precarious forms of self-employment linked to low wage-earning activities are common among formal workers, particularly after the pandemic.<sup>31</sup>

<sup>31</sup>The literature has highlighted the role of *subsistence*, as opposed to *transformational*, self-employment in the labor markets of developing countries (see, among others, Schoar 2010; Donovan, Lu, and Schoellman 2023). Self-employment is concentrated among the lowest-earning individuals in developing countries as a response to frictional labor markets, particularly in Colombia, as shown by Herreño and Ocampo (2023).

FIGURE D.2. Sectors and Industries of Employers of Formal Workers



**Notes:** Panel A reports the sectoral composition of formal workers of each formal attachment group in 2009, 2014, 2019 and 2024. Panel B displays the industry composition of formal workers of each formal attachment group in 2009, 2014, 2019 and 2024. All results are for the cohort of individuals aged 25 to 30 in 2009 who have ever reported earnings to the PILA. **Source:** Authors' calculations using PILA 2009–2024.

### D.1. Firm sectors

Figure D.2 reports the sectors and industries of employment for formal workers in each group. While formal workers are employed primarily in the service sector, the share of workers in every group in this sector varies widely over time. Workers in groups with low (or decreasing) formal attachment switch between services and agriculture (despite agriculture being the smallest sector in the economy). Agriculture plays a prominent role in 2014 and 2024 for the groups with the lowest formality (G1 and G5). In contrast, the share of services and manufacturing is relatively stable among the groups with high formal attachment (e.g., G2) or during periods of high attachment for the group with increasing formality rates (see group G4 in 2019 and 2024).

Panel B breaks manufacturing into construction and other manufacturing, and services into transportation, private services, and government (including social services). The increasing role of government jobs for groups with increasing formal attachment (G3 and G4) is notable.

## Appendix E. Formal workers in administrative and survey data

The PILA and Colombia's official household survey, the *Gran Encuesta Integrada de Hogares* (GEIH), are complementary data sources for studying the formal labor market. The GEIH is a monthly household sample administered by DANE, in which respondents self-report their employment status, sector of activity, and labor income. The PILA, by contrast, is the administrative census of all social security contributions filed by firms and self-employed workers. Consistent with existing evidence for Colombia, our series of formal workers from the PILA between 2009 and 2024 closely tracks the GEIH-based series over the same period: formal employment in the GEIH averages 8.9 million workers during our study period, and 9.6 million in PILA. Both data sources also show common cyclical movements, including the 2020 contraction and the subsequent recovery, and exhibit the same long-run upward trend.

Figure E.1 compares the distributions of log monthly earnings across both data sources: the PILA distribution is similar to the distribution of formally employed individuals in the GEIH, with a large mass near the minimum wage and a long upper tail. Meanwhile, the distribution that pools formal and informal workers in the GEIH has a lower mean and higher dispersion, with a long left tail composed almost entirely of informally employed individuals. The earnings for these informal matches are absent from the PILA by construction. It is crucial to highlight, however, that the longitudinal dimension of the PILA, which allows us to track individuals across all their formal spells, substantially mitigates this concern for our purposes. This is due to the transient nature of informality in Colombia.<sup>32</sup>

While approximately half the workforce holds a formal job at any given point in time, a much larger share of the population is formally employed at least once over their

---

<sup>32</sup>[Samaniago de la Parra and Fernández Bujanda \(2024\)](#) document high rates of transition between the formal and informal sectors in Mexico. [Bosch and Maloney \(2008\)](#) confirm such job-to-job cross-formality dynamics in Brazil, too.

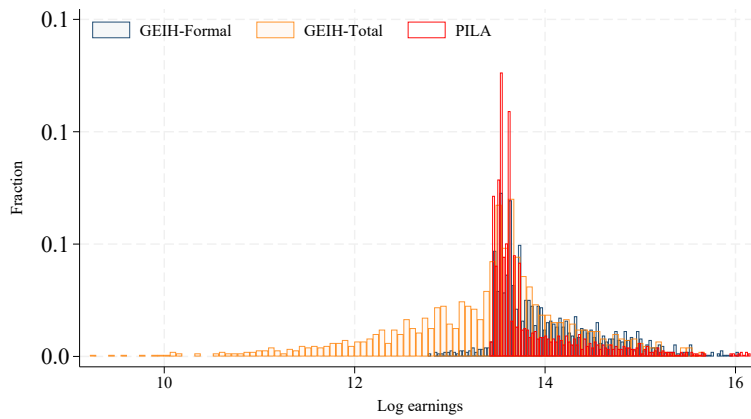
working lives and, as a result, our sample covers a much larger share of the population. For example, according to the 2018 Colombian census and population projections, the population aged 25 to 30 in 2009 was 4.2 million individuals.<sup>33</sup> Our cohort sample (workers aged 25 to 30 in 2009 with at least one formal employment spell between 2009 and 2024 and who meet the GRID project minimum annual income threshold in at least one formal spell) contains roughly 3.55 million individuals. Thus, the vast majority of the cohort, 85 percent, participate in the formal sector at least once during their prime working years. Therefore, the central object of our analysis, the trajectory of formal attachment over the work life, is observed for nearly the entire cohort rather than for the narrow slice of workers who happen to be formally employed in any given year.

While the PILA sample recovers a larger share of the population than is implied by the stock of formally employed workers at any given point in time, there are two caveats regarding the earnings dynamics we recover. First, by construction, PILA excludes income earned outside formal employment spells, so the residualized earnings, growth rates, and dispersion measures describe the dynamics of *formal* earnings rather than total labor income. The direction in which this missing information affects our estimates relative to a hypothetical measure that included informal income is not obvious. On the one hand, informal earnings could provide a partial cushion during formal non-employment spells, in which case incorporating them would compress the dispersion of earnings growth and potentially reduce measured volatility, particularly in the low-attachment groups. On the other hand, informal earnings tend to be lower and more variable than formal earnings, so including them could instead increase the dispersion of earnings levels and increase measured volatility, especially in the lower tail of the distribution. Unfortunately, no Colombian dataset combines the longitudinal coverage of PILA with reliable measurements of informal earnings, so a direct comparison is not

---

<sup>33</sup>Source: 2018 Population Census Colombia <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion>.

FIGURE E.1. Log Monthly Earnings in PILA and GEIH



*Note:* The figure shows histograms of the log of real monthly earnings for three different worker populations. First, the PILA population of formal workers. Second, the population of formal workers in GEIH. Third, the entire employed population in GEIH.

feasible. We therefore treat our results as describing the dynamics of formal earnings.