

Earnings inequality and dynamics in the presence of informality: The case of Brazil

NIKLAS ENGBOM

Stern School of Business, New York University, CEPR, NBER, and UCLS

GUSTAVO GONZAGA

Department of Economics, PUC-Rio

CHRISTIAN MOSER

Graduate School of Business, Columbia University and CEPR

ROBERTA OLIVIERI

Department of Economics, Cornell University

Using rich administrative and household survey data spanning 34 years from 1985 to 2018, we document a series of new facts on earnings inequality and dynamics in a developing country with a large informal sector: Brazil. Since the mid-1990s, both inequality and volatility of earnings have declined significantly in Brazil's formal sector. Higher-order moments of the distribution of earnings changes show cyclical movements in Brazil that are similar to those in developed countries like the US. Relative to the formal sector, the informal sector is associated with a significant earnings penalty and higher earnings volatility for identical workers. Earnings changes of workers who switch from formal to informal (from informal to formal) employment are relatively negative (positive) and large in magnitude, dispersed, negatively (positively) skewed, and less leptokurtic. Our results suggest that informal employment is an imperfect insurance mechanism.

KEYWORDS. Earnings inequality, earnings dynamics, mobility, higher-order moments, informality, sectoral transitions, administrative data, survey data.

JEL CLASSIFICATION. J31, J46, J62.

Niklas Engbom: nengbom@stern.nyu.edu

Gustavo Gonzaga: gonzaga@econ.puc-rio.br

Christian Moser: c.moser@columbia.edu

Roberta Olivieri: rso29@cornell.edu

We thank seminar participants at Columbia University and the Global Repository of Income Dynamics Conferences at Stanford SITE in 2019 and held virtually in 2020 for helpful comments and suggestions. We also thank the Brazilian Ministry of the Economy and the Brazilian Institute of Geography and Statistics for providing us with data access and the Data Zoom initiative by the Department of Economics at PUC-Rio for sharing its data cleaning procedures. Special thanks go to Serdar Ozkan and Sergio Salgado for providing their code and support. Engbom and Moser are grateful to the Federal Reserve Bank of Minneapolis and the Heller-Hurwicz Economics Institute at the University of Minnesota for their generous hospitality during a significant share of the period of work on this project. Any errors are our own.

1. INTRODUCTION

A salient feature of many developing and even some developed countries is a large informal sector, in which jobs evade government oversight in the labor market. The informal sector can be thought of as serving a dual role in labor market dynamics. On one hand, it offers workers readily available employment in case a worker is laid off or decides to quit owing to unfavorable pay or business conditions in his or her previous job. Through this channel, the informal sector provides insurance against labor income risk. On the other hand, it allows workers and firms to avoid costly labor regulations and income taxes, thereby enhancing the efficiency of hiring, firing, and production. This increased efficiency also means that workers are not covered by labor regulations such as employment protection laws, the minimum wage, social security contributions, and other benefits offered by formal jobs. Through this channel, the informal sector increases labor income risk.

As part of the Global Repository of Income Dynamics (GRID) Project, we study earnings inequality, volatility, and mobility in Brazil. As such, the statistics contained in this paper as well as many additional ones computed on the data will be publicly available through the GRID online database together with harmonized statistics for the other countries in this special issue. Brazil is an upper middle income country with a large informal sector. We find that among Brazilian metropolitan regions in 2004, 42% of all jobs were informal (i.e., without a formal work permit). At the same time, earnings inequality and informality rates significantly declined between the early 2000s and the late 2010s. This makes Brazil a particularly interesting setting to study for our purposes.

To dissect the distribution of earnings levels and earnings changes, we use a combination of rich administrative and household survey data from Brazil covering 34 years over the period from 1985 to 2018. The administrative records cover nearly the entire universe of formal sector workers in Brazil over those years. We complement these administrative records with detailed household survey data that follow individuals within households in Brazil's six largest metropolitan regions in a 16-month rotating panel structure from 2002 to 2015. The advantage of the household survey data is twofold. First, the data let us validate our findings on labor market outcomes in Brazil's formal sector, based on the administrative records. Second, they allow us to compare earnings levels and earnings changes between workers in Brazil's formal and informal sectors as well as for workers switching sectors between survey waves. In this way, we uncover a set of new facts for workers within and between the formal and informal sectors of Brazil.

In the first part of the paper, we compute a set of standardized statistics on earnings inequality, volatility, and mobility in Brazil's formal sector based on administrative data covering the period from 1985 to 2018. We start by documenting a remarkable decrease in earnings inequality for both men and women, which starts around 1995 and lasts until the end of our sample.¹ The decrease in overall earnings inequality is associated with relatively greater compression in the left tail of the distribution, which in turn is due to rapid real earnings growth among bottom earnings percentiles. The decrease

¹See Barros, de Carvalho, Franco, and Mendonça (2010) and Firpo and Portella (2019) for an overview of recent inequality trends in Brazil.

in inequality is also widespread among the lower 95% of the distribution. In contrast, the top five earnings percentiles have fanned out because of growth rates in real earnings that were increasing in ranks between the 95th and the 99.99th percentiles of the distribution. Earnings inequality across cohorts entering the labor market also fell over this period, but more so in the upper than in the lower tail of the distribution.

We then turn to earnings dynamics among formal sector workers in Brazil. Overall dispersion in 1-year earnings changes, conditional on gender-specific controls for worker age and educational attainment, rose rapidly during a volatile economic period in Brazil in the late 1980s and early 1990s, which included several inflation stabilization plans and a hyperinflationary period. This rise in dispersion of earnings changes is driven almost entirely by increasing lower-tail dispersion, that is, greater downside earnings risk. Following the macroeconomic stabilization after 1994, the dispersion of earnings changes decreased markedly, first driven by a decrease in the lower tail and later followed by a decrease in the upper tail of the distribution. We also find that the Kelley skewness of earnings changes is strongly procyclical (i.e., it is lower during recessions) but without much of a trend, while the excess Crow–Siddiqui kurtosis of earnings changes increased secularly from 1985 onward.

Although inequality in current earnings in Brazil has fallen dramatically over our period of study, this does not mean that inequality in more permanent earnings has followed the same trend. The relation between current and more permanent inequality is commonly summarized through measures of earnings mobility (Shorrocks (1978)). We find comparable levels of earnings mobility in Brazil, compared with those in concurrent studies for the US (McKinney, Abowd, and Janicki (2022)) and Canada (Bowlus, Gouin-Bonenfant, Liu, Lochner, and Park (2022)). Moreover, the extent of earnings mobility has not changed much over time, despite lower earnings volatility. That is, individuals now move across the earnings distribution to the same extent as in the past while the dispersion of earnings changes has declined. This suggests that the magnitude of earnings changes associated with a move between any two quantiles of the earnings distribution is smaller in more recent years, consistent with the underlying earnings distribution having been compressed over time (Alvarez, Benguria, Engbom, and Moser (2018)).

In the second part of the paper, we complement our analysis of Brazil's formal sector based on administrative records with longitudinal household survey data for 2002–2015 for the six largest metropolitan regions. We make the two data sets as comparable as possible and use them to validate our findings on earnings inequality and volatility across data sets. Although there remain important differences between the two data sets, the evolution of earnings inequality lines up quite closely between administrative and household survey data. Earnings volatility shows somewhat more diverging trends across the two data sources: the volatility of earnings changes is flat in the administrative records but decreasing in the household survey data over the 2002–2015 period. These differences are plausibly due to discrepancies in the coverages, income definitions, and the response rates across data sources.²

²In this manner, we contribute to an emerging literature that compares administrative and household survey data in other lower-income countries, such as Argentina (Blanco, Díaz de Astarloa, Drenik, Moser, and Trupkin (2022)) and Mexico (Puggioni, Calderón, Cebreros Zurita, Fernández Bujanda, Inguanzo González, and Jaime (2022)).

We proceed to exploit the longitudinal household survey data to study earnings inequality and dynamics within and between the formal and informal sectors. We draw four conclusions. First, mean 1-year residual earnings changes are similar in formal and informal jobs, but informal changes are significantly more dispersed, with greater probability mass in both tails of the distribution. Second, workers who switch between sectors have highly asymmetric earnings changes: workers transitioning from the informal to the formal sector tend to make earnings gains, whereas workers making the opposite transition on average lose earnings. Third, there was a pronounced decrease in the dispersion of earnings changes in the overall economy (the formal sector pooled with the informal sector) during the early 2000s, which was followed by a period of stabilization from 2006 onwards. Fourth and finally, holding everything else fixed, the large employment shift toward the less volatile formal sector on its own results in a fall in the volatility of earnings; the fall corresponds to 50% of the total decline since 2002. In other words, the process of labor market formalization appears to have played an important role in the decline in earnings volatility over this period. Together, these facts paint a rich picture of earnings inequality and dynamics in Brazil—a developing country with a large informal sector.

Our work combines two separate strands of the literature on informality and income dynamics. The first strand of the literature is concerned with labor market informality; see [Ulyssea \(2020\)](#) for an excellent review. [Meghir, Narita, and Robin \(2015\)](#) use a subperiod of the same household survey data that we use for the second part of our analysis. They show that both the distribution of wages and that of firm productivity substantially overlap between Brazil's formal and informal sectors. [Ulyssea \(2018\)](#) uses linked employer–employee survey of informal establishments to document facts about the distribution of (in)formal employment across the firm size distribution. Among the drivers behind high levels of informality in developing countries are high labor regulation costs ([Almeida and Carneiro \(2012\)](#)), weak enforcement ([Seminario-Amez \(2021\)](#)), payroll taxes ([Haanwinckel and Soares \(2021\)](#)), and the incidence of social policies like the minimum wage and conditional cash transfer programs such as Bolsa Família in Brazil ([Fairris and Jonasson \(2020\)](#)). We complement this body of research by studying earnings dynamics within and between the two sectors, highlighting the importance of the informal sector in particular.

Related work by [Dix-Carneiro and Kovak \(2019\)](#), [Ponczek and Ulyssea \(2021\)](#), and [Dix-Carneiro, Goldberg, Meghir, and Ulyssea \(2021\)](#) also highlights the role of the informal sector as an insurance mechanism against negative shocks to Brazilian local labor markets in the context of international trade. Building on their insights, we characterize earnings inequality and dynamics within and between the formal and informal sectors.

We also contribute to a growing literature on the causes of the pronounced decrease in earnings inequality in Brazil since the mid-1990s. [Firpo and Portella \(2019\)](#) provide an excellent review of recent studies that quantify the importance of falling returns to education and experience ([Ferreira, Firpo, and Messina \(2017\)](#)), falling returns to firm productivity ([Alvarez et al. \(2018\)](#)), trade liberalization ([Gonzaga, Filho, and Terra \(2006\)](#)), [Ferreira, Leite, and Wai-Poi \(2007\)](#), [Dix-Carneiro and Kovak \(2015\)](#)), and the rapid rise of

the minimum wage (Engbom and Moser (forthcoming))—among other factors—toward this decrease in earnings inequality.³

The second strand of the literature is concerned with income dynamics. Earnings dynamics have been studied in administrative and household survey data in many developed countries (see, e.g., Moffitt and Gottschalk (1995) and Sabelhaus and Song (2010)). A seminal contribution in this area is that by Guvenen, Ozkan, and Song (2014), who use 34 years of Social Security records to document new facts on the cyclical properties of higher-order moments of earnings changes in the US. Recent work has shed further light on the nature of earnings dynamics over the life cycle (Guvenen, Karahan, Ozkan, and Song (2021)) and over time (Bloom, Guvenen, Pistaferri, Sabelhaus, Salgado, and Song (2017)) in the US context. Hoffmann and Malacrino (2019) show that unemployment insurance reduces some of the cyclicity in skewness of earnings changes in Italy. We contribute to this literature by presenting a set of new empirical facts on earnings dynamics in a developing country with a large informal sector.

A recent study by Gomes, Iachan, and Santos (2020) also studies earnings dynamics in Brazil's formal and informal sectors. Their analysis is based on different survey data that are nationally representative over the period from 2012 to 2018. Our analysis complements theirs in several ways. First, we leverage a combination of administrative and household survey data, which we cross-validate with one another. Second, our focus lies on both earnings inequality and dynamics. Third, we use a longer panel from 2002 to 2015 in our household survey data and from 1985 to 2018 in our administrative data in order to document secular and cyclical movements of higher-order moments of the distribution of earnings changes. In this way, our study provides a holistic picture of Brazil's formal and informal labor markets.

The rest of the paper is structured as follows. Section 2 describes Brazil's institutional background from 1985 to 2018 and introduces the administrative and household survey data on which we base our analysis. Section 3 presents a set of standardized statistics pertaining to earnings inequality, volatility, and mobility, before validating findings between the administrative and household data. Section 4 dissects the role of (in)formality in Brazil's labor market. Finally, Section 5 concludes.

2. BACKGROUND AND DATA

2.1 *Brazil's macroeconomy from 1985 to 2018*

Between 1985 and 2018, Brazil underwent a transformative and volatile macroeconomic period. The period was characterized by rapid growth spurts interlaced with severe economic recessions, with negative GDP per capita growth recorded during the high-inflation period of the late 1980s and early 1990s, the financial crisis of the late 1990s, the global financial crisis around 2008, and the commodity price bust and political turmoil from 2014 to 2016; see panel A of Figure 1.

With rapid growth came other fundamental economic changes for Brazil. Over the period from 1985 to 2018, the services sector grew from 47% to 74% of total GDP, while

³We here restrict attention to monetary earnings; see Morchio and Moser (2020) for a study of nonpecuniary job amenity values in the context of Brazil.

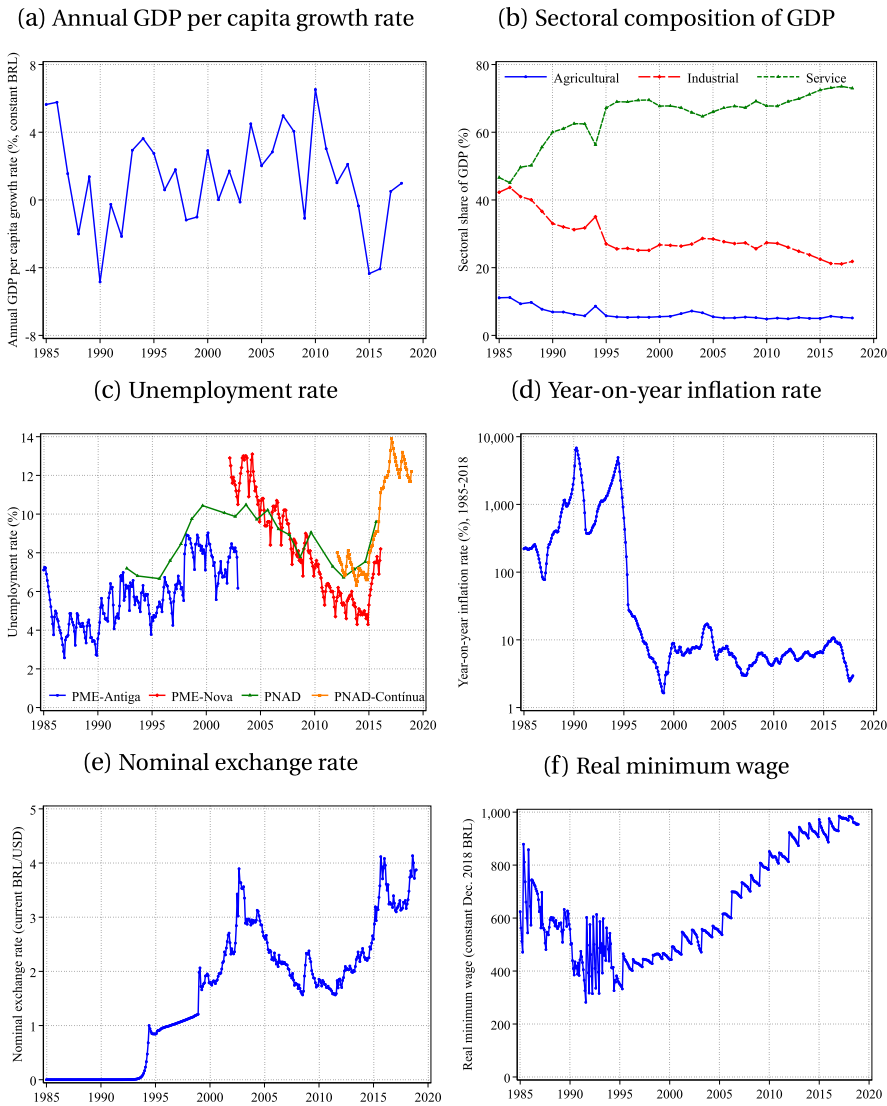


FIGURE 1. Macroeconomic indicators for Brazil, 1985–2018. *Note:* Panel A shows annual GDP per capita growth in constant USD. Panel B shows shares of total GDP by three sectors: agriculture (circles), industrial (diamonds), and services (triangles). Panel C combines data from the PME-Antiga (circles, January 1985–December 2002), the PME-Nova (diamonds, March 2002–February 2016), PNAD (triangles, September 1992–2014 except 1994, 2000, and 2010), and PNAD-Continua (squares, February 2012–December 2018). PME-Antiga and PME-Nova cover Brazil’s six largest metropolitan areas: Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, and Porto Alegre. PNAD and PNAD-Continua are representative of Brazil as a whole. See footnote 10 for details. Panel D shows year-on-year inflation using the IPCA price index on a log scale. Panel E shows the nominal exchange rate between Brazilian Reais (BRL) and US Dollars (USD). Panel F shows the real minimum wage in constant December 2018 BRL. *Source:* Panel A plots data from World Bank. Panels B, D, and F plot data from IPEA. Panel C plots data from IPEA and IBGE. Panel E plots data from Federal Reserve Economic Data (FRED).

the industrial sector shrank from 42% to 21% and the agricultural sector shrank from 11% to 5%, as illustrated by panel B.⁴

Although this statistic does not fully reflect labor market slack in the presence of a large informal sector, Brazil's unemployment rate fluctuated between 3% in the late 1980s and 13% in the early 2000s; see panel C.

A particularly scarring event in Brazil's recent macroeconomic history was a prolonged episode of high inflation in the first part of our sample period, from 1985 to 1994. Our preferred measure of inflation is based on the Extended National Consumer Price Index *Índice Nacional de Preços ao Consumidor Amplo* (IPCA) used by the country's central bank since 1999, when an inflation-target system was implemented. After fast-rising inflation during the early 1980s, Brazil eventually suffered from hyperinflation, with annual inflation rates above 6500% and several different national currencies. After several stabilization attempts, inflation was eventually brought under control with the implementation of the Plano Real in 1994 and was relatively stable thereafter (Ayres, Garcia, Guillén, Kehoe, Scheinkman, and Ter-Minassian (2021)); see the panel description.

Over the same period, Brazil's currency fluctuated significantly in real terms, first depreciating heavily from the early 1990s until 2002, then appreciating quickly for a decade in the wake of a commodity supercycle, and eventually depreciating again with its reversal and the deterioration of fiscal accounts (Ayres et al. (2021)) and (Benguria, Saffie, and Urzúa (2021)); see panel E.

Finally, Brazil implemented several economic policy changes between 1985 and 2018. Among the most salient changes is the rapid rise of the minimum wage starting in the early 2000s; its increase coincided with the election of the left-leaning Workers' Party. Over the subsequent decade and a half, Brazil's minimum wage increased by over 100% in real terms; see panel F.⁵

2.2 Data

In this section, we describe the two data sets used in our empirical analysis and our sample selection criteria. Our administrative data source is the *Relação Anual de Informações Sociais* (RAIS), a linked employer–employee longitudinal data set that covers nearly all formal jobs in Brazil. We complement our empirical analysis with microdata from the Brazilian monthly labor force survey *Pesquisa Mensal de Emprego* (PME) to validate our findings based on administrative data and to investigate whether income dynamics differ between formal and informal workers.

2.2.1 Administrative data

2.2.1.1 Data description Our main data source is RAIS, which contains administrative records from Brazil's Labor Statistics Dissemination Program (*Programa de Disseminação das Estatísticas do Trabalho*) within the Brazilian Ministry of the Economy

⁴The figures during Brazil's high-inflation period from 1985–1995 should be interpreted with caution due to the substantial volatility of nominal variables during these years.

⁵Engbom and Moser (forthcoming) and Haanwinckel (2020) show that the rise of the minimum wage had a pronounced effect on the earnings distribution over this period.

(*Ministério da Economia*), formerly the Ministry of Labor (*Ministério do Trabalho*). RAIS covers nearly the entire universe of workers in tax-registered firms. It excludes informal workers and firms, firm owners and shareholders unless they are (self-)employed, and certain less-populated regions in the years before 1994. The RAIS data are based on legally required annual reports made by firms that transmit information to the Brazilian government on all employees who were on the payroll in the previous year.

Compliance with filling in RAIS is high because of large penalties for late, incomplete, or inaccurate data. Since the main purpose of RAIS is to administer a federal wage bonus to formal employees, there are incentives for truthful reporting. RAIS is also used by ministries for administering an array of social programs related to the monitoring of formal jobs.

Each observation in RAIS is a worker-establishment match, or job, in a given year. For each job, the data set includes worker-related variables (e.g., gender, age, education, and unique worker identifier), firm-related variables (e.g., sector of activity, establishment size, municipality, and unique establishment and firm identifiers), and job-related variables (e.g., mean monthly earnings during the current year, contractual weekly hours, tenure, occupation, months of hiring and separation, and reason for separation). Our income variable corresponds to mean monthly earnings during the current year and includes wage, salary, and (holiday and performance) bonus payments before taxes.

Each worker has a unique identification number in RAIS, which allows us to recover the full formal work history of all individuals in the database. We use data from 1985 to 2018. RAIS is very large, with an average of around 40 million observations per year, which sums to approximately 1.2 billion job records for the 1985–2018 period. For the year 1996, this corresponds to around 42.5% of the labor force, 46.0% of employment, and close to all formal employment (Engbom and Moser (forthcoming)).

2.2.1.2 Sample selection We apply some standard filters to the administrative data. First, we drop all workers without valid identification numbers or with zero earnings. We then restrict the data to workers in the 25–55 age range. Earnings data in RAIS are censored (i.e., reported as missing) above 120 times the national minimum wage. A Pareto tail imputation exercise suggests that censored observations correspond to a very small proportion, approximately 0.01% of the sample. To focus on workers with a meaningful attachment to the labor market, we drop those with total annual nominal earnings—defined in the next paragraph—below earnings associated with part-time employment at the minimum wage for 3 months.⁶

2.2.1.3 Variable construction Our interest lies in features of the distribution of earnings or wages.⁷ Since the period from 1985 to 1994 was characterized by high inflation

⁶We drop any observation of an individual i in year t with total annual nominal earnings y_{it} if $y_{it} < \frac{1}{2} \times 40 \frac{\text{hours}}{\text{week}} \times 4 \frac{\text{weeks}}{\text{month}} \times 3 \text{ months} \times \overline{\text{MW}}_t$, where $\overline{\text{MW}}_t$ is the mean prevailing minimum wage over the individual i 's period of employment in year t .

⁷Throughout the paper, we use the terms *earnings* and *wages* to interchangeably refer to total annual real earnings from all (formal) employment.

and multiple currency switches, we first focus on monthly earnings reported in multiples of the prevailing minimum wage. We multiply this multiple by the mean nominal minimum wage in current Brazilian Reais during the months of a given job spell. Since we know for each job spell the hiring and separation dates, we construct total annual nominal earnings for each individual by summing their mean monthly earnings over all months of employment and all jobs recorded in RAIS in a given calendar year. Finally, we obtain total annual real earnings by deflating total annual nominal earnings by the mean IPCA, normalized in December 2018, during the months of employment. By measuring earnings in this way, we minimize measurement error related to volatility in nominal variables. For consistency, we apply the same method each year throughout the whole period 1985–2018.

Using the administrative data, we construct the following five variables for an individual i of gender $G(i) \in \{\text{male, female}\}$ and age group $A(i, t) \in \{25, 26, \dots, 55\}$ in year $t \in \{1985, 1986, \dots, 2018\}$:

1. Log total annual real earnings, or *log earnings*, $\ln y_{it}$, computed as the sum of all of an individual's earnings across employment spells in a given year;
2. Residual log earnings conditional on gender-year-specific age dummies, or *residual earnings*,

$$\varepsilon_{it} = \ln y_{it} - \sum_{G', t', A'} \alpha_{G' t' A'} \mathbf{1}[G(i) = G', t = t', A(i, t) = A'], \tag{1}$$

where $\alpha_{G' t' A'}$ is a gender-year-age-specific coefficient on an indicator for the combination of gender G' , year t' , and age A' , denoted $\mathbf{1}[G(i) = G', t = t', A(i, t) = A']$;⁸

3. One-year-forward changes in residual earnings based on equation (1), or *1-year earnings changes*,

$$g_{it}^1 = \varepsilon_{i,t+1} - \varepsilon_{i,t};$$

4. Five-years-forward changes in residual earnings based on equation (1), or *5-year earnings changes*,

$$g_{it}^5 = \varepsilon_{i,t+5} - \varepsilon_{i,t};$$

5. Residual log mean earnings over the previous 3 years, conditional on gender-year-specific age dummies, or *permanent earnings*,

$$P_{i,t} = \ln \left(\frac{y_{i,t-2} + y_{i,t-1} + y_{i,t}}{3} \right) - \sum_{G', A'} \gamma_{G' t' A'} \mathbf{1}[G(i) = G', t = t', A(i, t) = A'],$$

⁸In Appendix B of the Online Supplemental Material (Engbom, Gonzaga, Moser, and Olivieri (2022)), we present results of an alternative definition of residual earnings that also conditions on the education group $E(i) \in \{\text{primary, middle, high school, college}\}$,

$$\varepsilon_{it} = \ln y_{it} - \sum_{G', t', A'} \alpha_{G' t' A'} \mathbf{1}[G(i) = G', t = t', A(i, t) = A'] - \sum_{G', t', E'} \beta_{G' t' E'} \mathbf{1}[G(i) = G', t = t', E(i) = E'].$$

TABLE 1. Summary statistics on formal sector workers in administrative data.

	1985		1996		2007		2018	
	Men	Women	Men	Women	Men	Women	Men	Women
Share	0.693	0.307	0.634	0.366	0.600	0.400	0.557	0.443
Age:								
Mean	34.288	33.917	35.548	35.731	35.890	36.067	36.999	37.118
Std. dev.	7.658	7.407	8.079	7.956	8.413	8.444	8.370	8.348
Education shares:								
Middle school	0.180	0.161	0.212	0.172	0.219	0.153	0.190	0.119
High school	0.177	0.303	0.232	0.352	0.386	0.476	0.536	0.590
College	0.102	0.212	0.097	0.217	0.103	0.235	0.110	0.223
Earnings:								
Mean	30,687	21,820	29,297	23,342	28,263	24,620	33,674	29,709
Std. dev.	41,580	25,672	42,168	31,850	44,689	36,625	46,626	38,139
Log earnings:								
Mean	9.770	9.532	9.655	9.480	9.698	9.584	9.969	9.871
Std. dev.	1.069	0.966	1.133	1.080	1.000	0.979	0.917	0.889
Obs. (mm)	10.765	4.778	13.845	7.985	20.255	13.527	23.455	18.631

Note: The table shows summary statistics for select years separately by gender. The omitted education category is primary school. Observations are in millions. Earnings are in 2018 BRL. Source: RAIS, 1985–2018.

where $\alpha_{G't'A'}$ is a gender-year-age-specific coefficient on an indicator for the combination of gender G' , year t' , and age A' , denoted $\mathbf{1}[G(i) = G', t = t', A(i, t) = A']$.

We will frequently refer to percentiles of the distribution of (residual) log earnings or (residual) log earnings changes. To this end, we use the notation Pp , for $p \in [0, 100]$, to denote the p th percentile in logs. For example, $P50$ denotes the median while $P99.9$ denotes the top 0.1 percentile. Similarly, $Pp_1 - Pp_2$, for $p_1, p_2 \in [0, 100]$, denotes the difference between Pp_1 and Pp_2 , or the log ratio between the p_1 th percentile and the p_2 th percentile.

2.2.1.4 Summary statistics Table 1 presents basic summary statistics for selected years between 1985 and 2018 on the gender and age composition, the earnings distribution, and sample sizes, based on our sample from RAIS data.⁹ Among the noteworthy features of Brazil's formal sector over this period are the pronounced increases in female labor force participation, high school completion, and overall employment, with a concurrent decline in the standard deviation of log earnings.

2.2.2 Household survey data

2.2.2.1 Data description To study earnings inequality dynamics for both formal and informal workers in Brazil, we use microdata from PME. The survey was conducted by the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística*, or IBGE) in Brazil's six largest metropolitan areas: São Paulo, Rio de Janeiro,

⁹Table A.1 in Appendix A.1 of the Online Supplemental Material presents detailed summary statistics on the distribution of earnings for each year between 1985 and 2018. Appendix Tables A.2 and A.3 of the Online Supplemental Material show the same statistics separately for men and for women.

Belo Horizonte, Porto Alegre, Salvador, and Recife. It was administered from the early 1980s until 2016, when it was terminated and replaced with a different survey. While it was active, the survey was used to compute official unemployment statistics for Brazil.

The rotating panel design of PME is such that the surveys are representative at the metropolitan-area level in each month. We use data from the so-called New PME (*PME-Nova*), covering 2002–2015. We do not make use of the Old PME (*PME-Antiga*) survey data covering the period before 2002, since they do not contain the necessary information on employment and income to allow us to replicate all of our analysis for the earlier years it covers.¹⁰ By the end of the period, the sample covered around 34,000 households and 95,000 individuals in each month. The pooled data for the 2002–15 period feature approximately 7.3 million observations, or around 500,000 per year on average. The main variables we use are the worker ID, gender, age, schooling, monthly earnings, labor market status (employed, self-employed, unemployed, or out of the labor force), and information on whether the individual holds a formal work permit, as explained below. Monthly earnings are in 2018 BRL and include wage, salary, and (holiday and performance) bonus payments before taxes.

Formal employees in Brazil are hired under the Brazilian labor codes *Consolidação das Leis do Trabalho*, CLT. CLT states that each employer has to fill in and sign the employee's working card (*Carteira de Trabalho*) when formally hiring a worker in Brazil.¹¹ After asking if he or she is employed, PME elicits whether the worker possesses a signed working card. Since RAIS covers only workers hired under CLT, workers with a working card correspond to those in the administrative data.

It is important to note that all household surveys run by IBGE are anonymous. IBGE has long had reputation of never granting outsiders access to any personally identifiable information of respondents.

PME surveys have a rotating panel structure similar to that of the Current Population Survey (CPS) in the US. Households are surveyed for two spells of four consecutive months; the spells are 8 months apart from each other. This means that households complete four monthly interviews, followed by an 8-month pause, and then by another four monthly interviews. This rotating panel structure means that the months the individual is interviewed are the same in any 2 consecutive years. Interviews are spread evenly within a month and households are always interviewed in the same week of the month.

Households are correctly identified throughout all eight interviews. However, PME does not assign the same identification number to each individual in the household

¹⁰The Old PME started in 1982 and ended in December 2002. The New PME, which started in March 2002, comprised major revisions that include a significantly larger questionnaire and more extensive checks on measures of labor market participation and other variables. The New PME, in turn, was discontinued after February 2016, since in the first quarter of 2012 IBGE had introduced the Continuous PNAD (*PNAD-Contínua*), a new household survey with 5-quarter longitudinal information and nationally representative coverage.

¹¹The working card is a booklet with information on an individual's complete formal labor market history, including all details of each job held by a worker—the hiring and separation dates, paid vacation periods, leaves of absence, etc.

across interviews. To reduce attrition, we use an algorithm developed by Ribas and Soares (2008), which identifies the same individual in each household across interviews using a fuzzy merge based on the combination of reported dates of birth and genders.¹²

2.2.2.2 Sample selection We make several choices to make the information in the two data sets—RAIS and PME—as comparable as possible for formal workers. First, we use only data for workers in the 25–55 age range. For the cross-sectional exercises, we construct comparable measures of annual earnings, y_{it} . Since the panel structure of PME allows us to follow workers for only 1 year, we also compute 1-year-forward residualized log annual earnings changes.

For comparability reasons, we drop all business owners who contribute to social security and domestic employees in PME, since they are not measured in the administrative data. We consider as formal (informal) workers only those individuals who show up as formal (informal) employees in all monthly observations within a calendar year. We drop individuals who work in both sectors within the same year. We further restrict attention to individuals' primary job.¹³

In addition, we apply the following selection criteria. We drop individuals with year-on-year survey attrition or without positive earnings from any (formal or informal) employment during any of the survey waves during a year. For the longitudinal statistics, we restrict ourselves to individuals who have a full 8 months of nonmissing responses in the 2 consecutive years. To mimic the topcoding in RAIS, we drop monthly earnings above 120 times the minimum wage. The reason for doing this is that the original RAIS data have dropped, and not winsorized, earnings above this threshold. Finally, we trim observations with annualized incomes below the equivalent of 1.5 months of full-time work at the prevailing minimum wage, which is the equivalent of the bottom threshold we used in our baseline analysis of RAIS.

The final sample is thus composed of formal workers (i.e., employees with a valid working card) and informal workers (i.e., employees without a valid working card and self-employed individuals who do not contribute to social security). Note that this sample includes workers employed in the public sector.

2.2.2.3 Variable construction We construct variables in PME to be analogous to those in the administrative data whenever possible. Since the household survey follows a rotating panel format, we create seasonal dummy variables that identify each 4-month period in the calendar year in which the individual is interviewed (which we refer to as a survey wave).¹⁴

¹²Standardized cleaning procedures and the panel linkage method are available from Data Zoom by PUC-Rio at <http://www.econ.puc-rio.br/datazoom/english/index.html>.

¹³In Appendix D of the Online Supplemental Material, we show that extending the analysis to workers with multiple concurrent jobs, who make up less than 3% of all workers, does not substantially alter our conclusions.

¹⁴There are twelve seasonal dummies for the cross-section statistics, based on PME's 4–8–4 rotating panel system: (January–April), (February–May), (March–June), (April–July), (May–August), (June–September), (July–October), (August–November), (September–December), (October–December; January), (November–December; January–February), (December; January–March). For the longitudinal exercises, only a subset consisting of nine seasons is relevant: (January–April), (February–May), (March–June), (April–July), (May–August), (June–September), (July–October), (August–November), and (September–December).

Using the household survey data, we construct the following three variables for an individual i of gender $G(i) \in \{\text{male, female}\}$, age group $A(i, t) \in \{25, 26, \dots, 55\}$, and season group $S(i, t) \in \{(\text{January–April}), (\text{February–May}), \dots, (\text{December; January–March})\}$ in year $t \in \{2002, 2003, \dots, 2015\}$:

1. Log total annual real earnings, or *log earnings*, $\ln y_{it}$, based on summing earnings during the (first or second set of) four consecutive interview months, multiplied by three;
2. Residual log earnings conditional on gender-year-specific age and season dummies, or *residual earnings*,

$$\begin{aligned} \varepsilon_{it} = \ln y_{it} &- \sum_{G', t', A'} \delta_{G' t' A'} \mathbf{1}[G(i) = G', t = t', A(i, t) = A'] \\ &- \sum_{G', t', S'} \eta_{G' t' S'} \mathbf{1}[G(i) = G', t = t', S(i, t) = S'], \end{aligned} \tag{2}$$

where $\delta_{G' t' A'}$ is a gender-year-age-specific coefficient on an indicator for the combination of gender G' , year t' , and age A' , denoted $\mathbf{1}[G(i) = G', t = t', A(i, t) = A']$, and $\eta_{G' t' S'}$ is a gender-year-season-specific coefficient on an indicator for the combination of gender G' , year t' , and season S' , denoted $\mathbf{1}[G(i) = G', t = t', S(i, t) = S']$;

3. One-year-forward changes in residual earnings based on equation (2), or *1-year earnings changes*,

$$g_{it}^1 = \varepsilon_{i,t+1} - \varepsilon_{i,t}.$$

2.2.2.4 Summary statistics One may suspect that the workers employed in Brazil’s formal sector are not identical to those in the informal sector. Indeed there are some important differences in worker composition across sectors, as shown in Table 2. Informal workers have earnings that are 83.7 log points lower on average and their distribution is more dispersed than that of formal workers, confirming the findings from our visual analysis. Informal workers are slightly more likely to be female, less likely to be white and more likely to be black, less likely to have a high school degree, older, have slightly more children in the household, are less likely to work in manufacturing or services (i.e., more likely to work in agriculture and residual sectors), less likely to work in white collar occupations, and less likely to work at firms with 6–10 or more than 10 persons. Informal workers also have significantly lower tenure in their current job, by around 4.5 years, and work significantly fewer hours, by around 8.5 hours per week.

3. CORE STATISTICS ON EARNINGS INEQUALITY AND DYNAMICS IN BRAZIL’S FORMAL SECTOR

This section documents patterns of earnings inequality and earnings dynamics in the formal sector in Brazil using the administrative RAIS matched employer–employee data. In the next section, we turn to the informal sector and compare administrative and survey data.

TABLE 2. Summary statistics on workers in survey data, overall and by sector.

	Overall	Formal	Informal
Mean log earnings	9.653	9.934	9.097
Std. dev. of log earnings	0.922	0.796	0.902
Share female	0.432	0.430	0.437
Share Nonwhite	0.418	0.398	0.459
Share with at least high school education	0.564	0.651	0.393
Share age 35–44	0.344	0.343	0.345
Share age 45–55	0.285	0.258	0.338
Share with children in household	0.356	0.352	0.365
Share in manufacturing	0.192	0.212	0.153
Share in services	0.529	0.580	0.428
Share white collar	0.383	0.457	0.237
Share employed at firm with 6–10 employees	0.037	0.039	0.034
Share employed at firm with > 10 employees	0.637	0.915	0.175
Mean tenure in years	7.659	7.810	7.360
Std. dev. of tenure in years	7.286	7.223	7.401
Mean weekly hours worked	36.899	39.747	31.265
Std. dev. of weekly hours worked	12.690	9.778	15.579
Obs. (weighted)	311,895,692	207,175,606	104,720,086
Obs. (unweighted)	743,975	489,555	254,420

Note: All statistics are for workers aged 25–55. Earnings are in 2018 BRL. Residual categories for which no share is reported are men, white, less than high school education, age 25–34, no children in household, agriculture, blue collar, and firms with 1–5 persons. The weighted number of observations corresponds to the sum of survey sampling weights, while the unweighted number of observations corresponds to the number of data points in the sample. *Source:* PME, 2002–2015.

3.1 Earnings inequality

3.1.1 Evolution of earnings percentiles Most percentiles of the earnings distribution in Brazil experienced significant cumulative real wage growth over the 34 years in our data, as illustrated by Figure 2. Wage growth was not, however, monotone. In particular, real wages fell consistently during the high inflationary years of the late 1980s and early 1990s. Since the extensive macroeconomic reforms undertaken in the early 1990s, most workers across the earnings distribution have seen real wage growth. The patterns are quite similar for men and women, as well as when pooling both genders; see Appendix B of the Online Supplemental Material.

While all parts of the earnings distribution experienced significant real wage growth since the early 1990s, there is also important heterogeneity across the earnings distribution. In particular, since the early 1990s, earnings have grown disproportionately at the bottom of the earnings distribution. For instance, since 1995, earnings at the 75th percentile grew by about 10 log points, while at the 25th percentile, they rose by 60 log points. This pattern is reversed at the very top of the earnings distribution, which experienced widening inequality similar to that of many developed countries. For instance, earnings at the 99th percentile rose by more than at the 90th percentile.

3.1.2 Evolution of earnings inequality Due to the faster real wage growth at the bottom of the distribution, earnings inequality fell dramatically in Brazil starting in the early

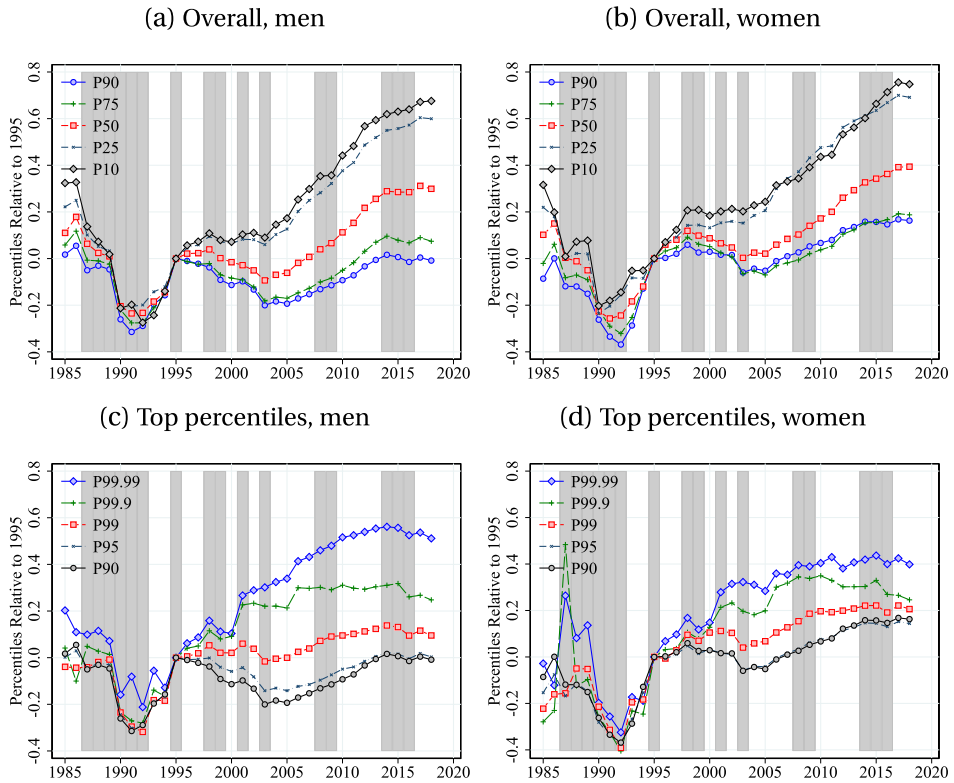


FIGURE 2. Evolution of earnings percentiles, by gender. *Note:* Workers aged 25–55. Percentiles of the distribution of log real annual earnings, normalized to 1995. *Source:* RAIS, 1985–2018.

1990s, as shown in Figure 3. The 90–10 percentile ratio declined from 3 to 2.3. The patterns are again quite similar among men and women, as well as when pooling both genders; see Appendix B of the Online Supplemental Material. Moreover, the fall in inequality was particularly pronounced at the bottom of the earnings distribution, as evidenced by the larger fall in the 50–10 ratio compared with that of the 90–50 percentile ratio. Nevertheless, the 90–50 percentile ratio also fell by a significant amount, driven by fast real median wage growth. This large decrease in inequality is particularly remarkable given that many countries experienced increases in inequality over the same period. That said, Brazil continues to have high levels of inequality.

3.1.3 *The role of entry conditions* Figure 4 plots lower and upper tail inequality among 25-year-olds by gender over time. As for the aggregate trends, from the early 1990s onwards, young workers experienced a large decline in inequality. In other words, the large overall decline in inequality was not solely the result of changes in earnings dynamics after labor market entry. Instead, inequality is lower also among labor market entrants. The compression in the earnings distribution among young workers was again particularly pronounced at the bottom of the earnings distribution, as evidenced by the larger fall in the 50–10 percentile ratio relative to that of the 90–50 ratio.

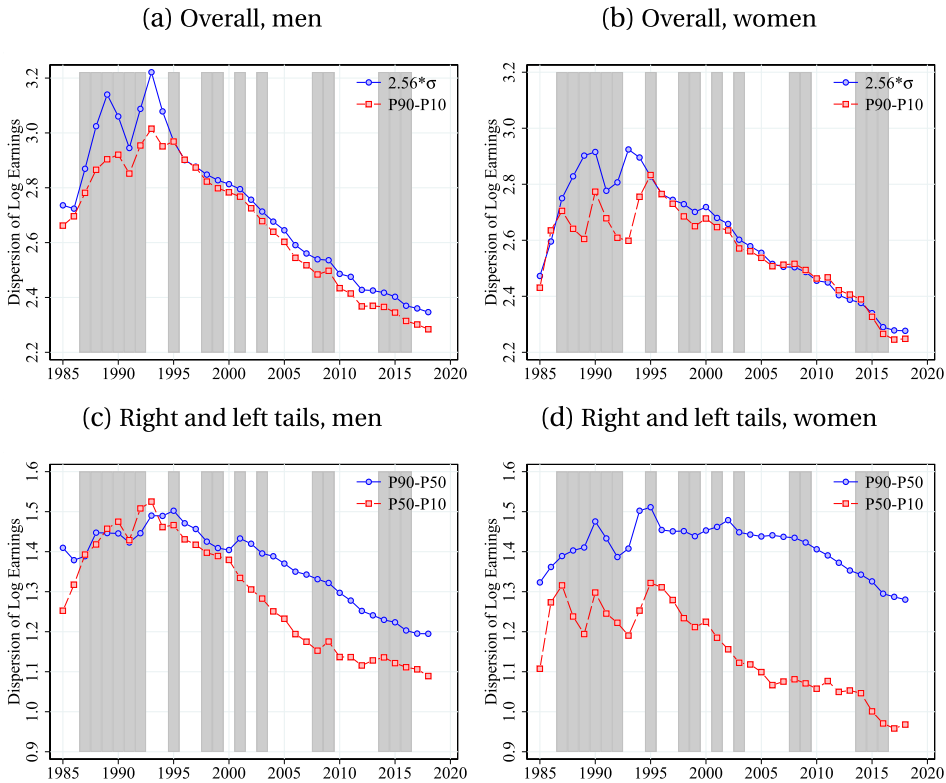


FIGURE 3. Earnings inequality, by gender. *Note:* Workers aged 25–55. σ denotes the standard deviation of log real annual earnings. *Source:* RAIS, 1985–2018.

To further investigate the role of changes in initial conditions versus that of changes in post-entry life-cycle dynamics, in Figure 5 we follow cohorts over time as they age. The earliest cohort to turn 25—those who did so in 1985—saw an initial increase in in-

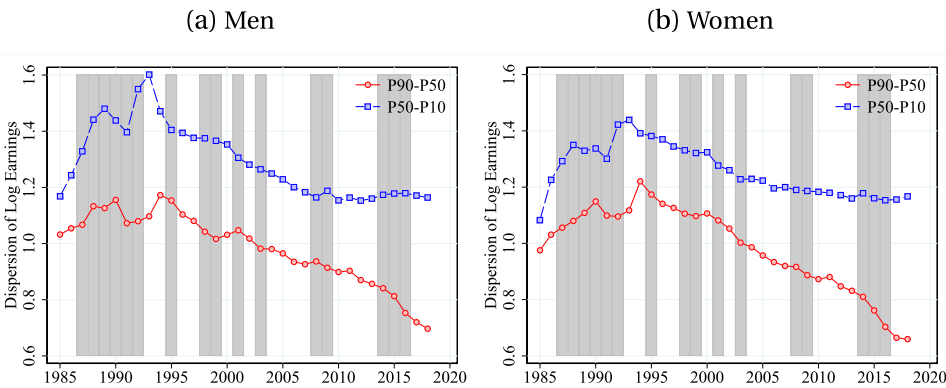


FIGURE 4. Earnings inequality among 25-year-olds, by gender. *Note:* Workers aged 25. *Source:* RAIS, 1985–2018.

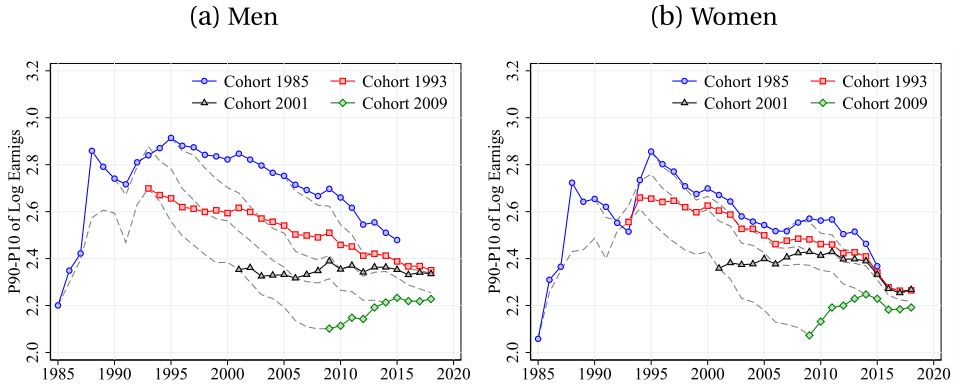


FIGURE 5. Life-cycle inequality across cohorts, by gender. *Note:* Workers aged 25–55. *Source:* RAIS, 1985–2018.

equality during the first 10 years, followed by a subsequent decline. This initial increase, however, may be the result of a time effect associated with the period of high inflation experienced by Brazil over this period. Subsequent cohorts of men and women have seen a gradual flattening and eventual reversal of the profile of within-cohort inequality with age. One possible factor behind this interesting pattern is the rapid increase in the minimum wage over this period. It is well known that a minimum wage tends to disproportionately affect young workers, and it may have contributed to a particular compression in inequality at labor market entry. As older workers are less affected by the minimum wage, inequality has fallen by less at older ages, which by itself has contributed toward a steepening of life-cycle inequality profiles.

3.2 Earnings dynamics

3.2.1 Evolution of dispersion in earnings changes Figure 6 plots upper-tail (i.e., 90–50) and lower-tail (i.e., 50–10) percentile ratios of 1-year residual log earnings changes separately by gender. Overall, men have somewhat more dispersed earnings changes, especially in the lower tail of the distribution. Relatively low (high) earnings changes became more (less) dispersed during the years of high inflation in the late 1980s and early 1990s. Since then, earnings have gradually become less volatile. The lower-tail dispersion of earnings changes is countercyclical—meaning that the dispersion in relatively low earnings changes increases during recessions, while the upper tail of earnings changes is procyclical—meaning that the dispersion in relatively high earnings changes decreases during recessions. It is worth noting that these patterns are qualitatively different from the results documented by Busch, Domeij, Guvenen, and Madera (2022) for Germany, Sweden, and the US.

3.2.2 Evolution of higher-order moments of the distribution of earnings changes Letting P_p denote the p th percentile of the distribution of (log real) earnings changes, we adopt two quantile-based measures of higher-order moments. First, the *Kelley skewness* is defined as $[(P_{90}-P_{50}) - (P_{50}-P_{10})]/(P_{90}-P_{10})$, which is identically zero for a Gaussian

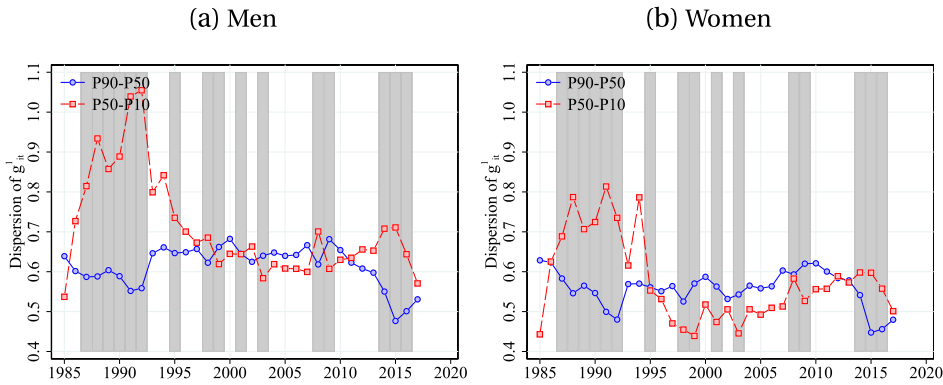


FIGURE 6. Dispersion of 1-year log earnings changes, by gender. *Note:* Workers aged 25–55. *Source:* RAIS, 1985–2018.

distribution. Second, the Crow–Siddiqui kurtosis is defined as $(P97.5 - P2.5) / (P75 - P25)$, which equals 2.91 for a Gaussian distribution. Based on this, we define the *excess Crow–Siddiqui kurtosis* as the Crow–Siddiqui kurtosis minus 2.91, which is thus identically zero for a Gaussian distribution.

Figure 7 plots the Kelley skewness and excess Crow–Siddiqui kurtosis of 1-year residual earnings changes. As suggested by Figure 6, the Kelley skewness of earnings changes is procyclical: relatively low earnings changes become more dispersed during recessions. In contrast, it is difficult to reach a definitive conclusion about the cyclicity of the excess Crow–Siddiqui kurtosis. While the Kelley skewness displays little secular trend over the past 34 years, the excess Crow–Siddiqui kurtosis has gradually risen. In other words, the likelihood of relatively large negative and positive shocks has risen over time. One possible factor behind this pattern is the decline in informality in Brazil over this period. It may be that workers 20 years ago were more likely to leave the formal sector in response to large negative shocks, whereas today they tend to remain formally employed, though at lower earnings. If earnings were to later revert, this would account

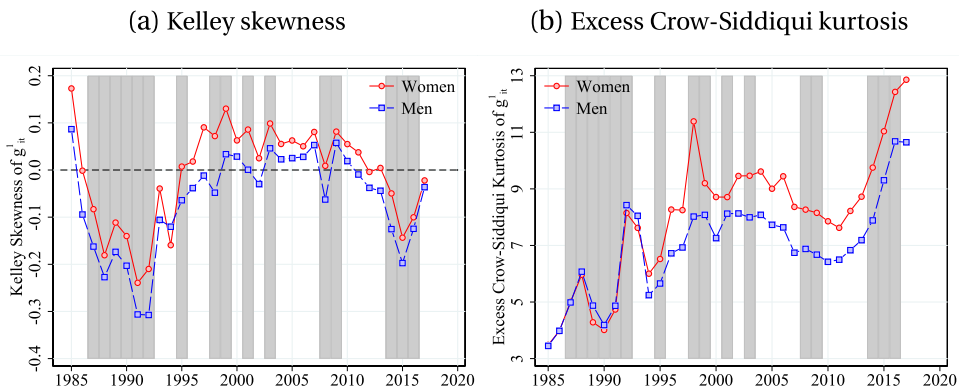


FIGURE 7. Higher-order moments of the distribution of one-year log earnings changes. *Note:* Workers aged 25–55. *Source:* RAIS, 1985–2018.

for an increase in the excess Crow–Siddiqui kurtosis.¹⁵ Another potential explanation behind the rise in the excess Crow–Siddiqui kurtosis over time pertains to the process of trade liberalization that Brazil underwent during the 1990s (Dix-Carneiro and Kovak (2017), Martínez and Mello (2021)).

3.2.3 Life-cycle dynamics Complementing our time-series evidence, Figure 8 studies the distribution of 1-year earnings changes over the life cycle.¹⁶ In particular, it plots the 90–10 percentile ratio, the Kelley skewness, and excess Crow–Siddiqui kurtosis of one-year log earnings changes by age group as a function of permanent earnings. To this end, for reasons of standardization across countries in the special issue, we use 20 years of the RAIS data covering the period from 1999 to 2018. Young workers have more volatile earnings as measured by the 90–10 percentile ratio of earnings changes, as do lower-permanent-earnings workers within all age groups. There is no pronounced systematic pattern for the Kelley skewness across permanent-earnings quantiles, although younger workers have a lower Kelley skewness of earnings changes, especially among women. The ranking of Kelley skewness across age groups is reversed compared to the US (Güvenen et al. (2021)). Interestingly, the pattern for the excess Crow–Siddiqui kurtosis of earnings changes is partly inverted relative to that for the 90–10 percentile ratio. While low-permanent-earnings workers also have the highest excess Crow–Siddiqui kurtosis within all age groups, older workers have higher excess Crow–Siddiqui kurtosis than their younger counterparts. Younger workers are subject to more dispersed but more negatively skewed and less leptokurtic shocks earnings changes compared to older workers. Women are less likely than men to experience negative earnings shocks as evidenced by the higher Kelley skewness of earnings changes, possibly because they are more likely to drop out of the formal sector in response to large negative shocks.

3.2.4 Earnings mobility Figure 9 investigates earnings mobility. In particular, it plots a worker's average rank in the earnings distribution 10 years later as a function of his or her rank in the distribution in the current period, separately by age and gender. Appendix figures B.14–B.15 of the Online Supplemental Material contains similar plots for outcomes 5 years later, with a similar conclusion. The distribution is mean reverting, in the sense that individuals currently at the bottom of the distribution tend to move up the distribution over time and vice versa. Specifically, workers in the bottom 40% of the permanent-income distribution on average gain ranks over the subsequent 10 years, while workers in the upper 60% of the permanent-income distribution on average move down in ranks over the subsequent 10 years. Younger workers are more mobile in the sense that their permanent-income rank is more mean reverting, as evidenced by the flatter solid line in the figure.

Figure 10 conducts the same analysis instead over time: it plots the average rank of workers 10 years later as a function of their rank in the earnings distribution today, separately by year and gender. There is little systematic variation in mobility between 1995,

¹⁵In Section 4.1, we further investigate the distribution of earnings changes within and between the formal versus informal sector as well as the rates of switching sectors.

¹⁶Appendix Figures B.11–B.13 of the Online Supplemental Material present the same statistics for the distribution of 5-year earnings changes, as well as the corresponding second, third, and fourth standardized moments of the distributions of 1-year and 5-year earnings changes.

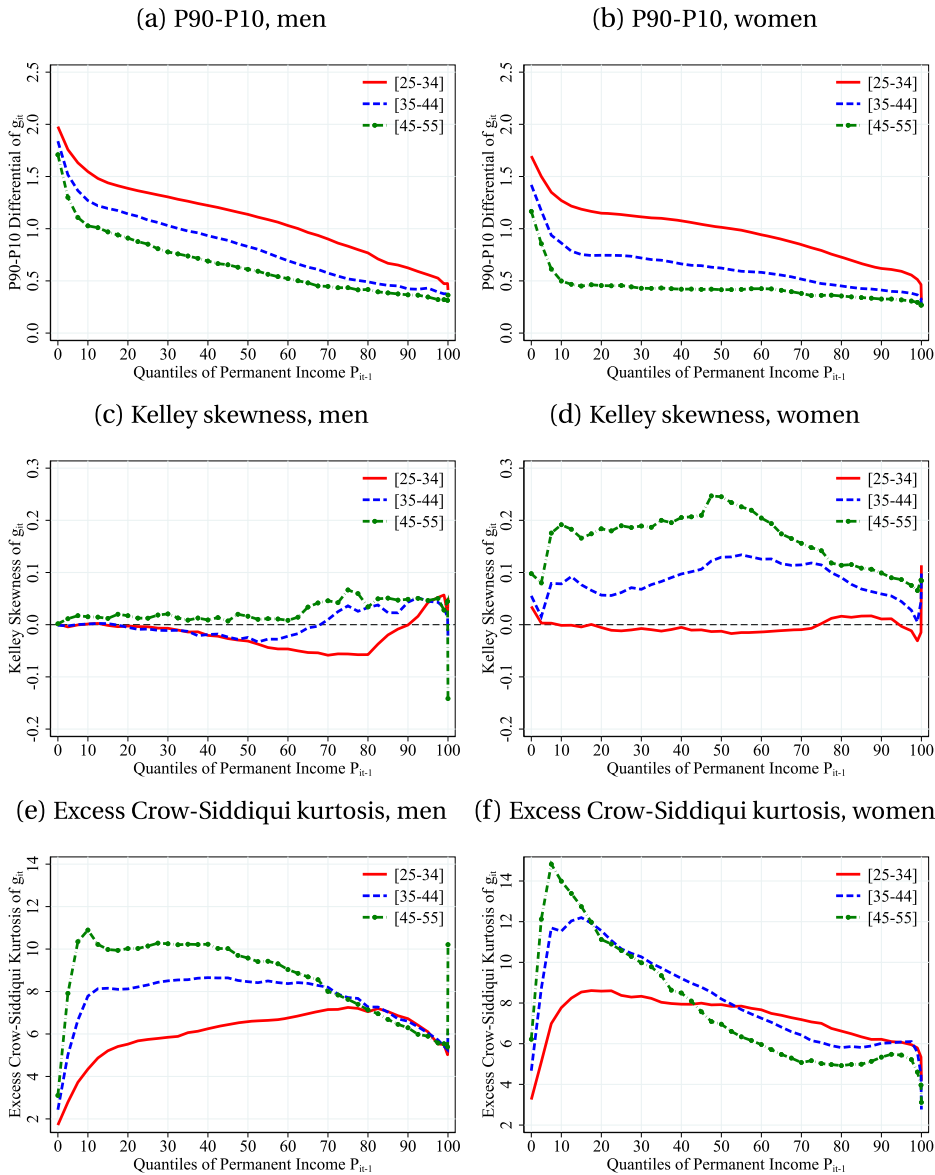


FIGURE 8. Moments of the distribution of 1-year earnings changes, by gender. *Note:* Workers aged 25–55. *Source:* RAIS, 1999–2018.

shown as a solid line, and 2015, shown as a dashed line. This observation is quite interesting in light of the pronounced decline in earnings volatility over this period shown in Figure 6. That is, in spite of lower earnings volatility, individuals move across the earnings distribution to a similar extent now as in the past. Since the underlying earnings distribution is more compressed in later years of the sample period, however, the earnings changes associated with a move between two particular rungs of the earnings distribution is smaller in absolute magnitude.

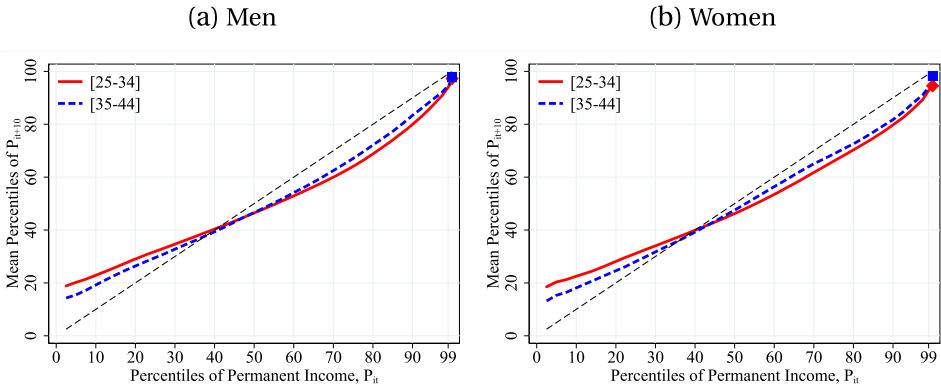


FIGURE 9. Evolution of 10-year mobility over the life cycle, by gender. *Note:* Workers aged 25–55. Square and diamond markers denote the top 0.1% of permanent income P_{it} . *Source:* RAIS, 1985–2018.

3.3 Comparing administrative and household survey data

To what extent do the patterns in administrative data correspond with what households self-report in surveys? To address this question, we compare our results for the formal sector from the RAIS administrative data with the PME household survey. We replicate our exercises for the formal sector from the previous section for the formal sector in the PME. To that end, we restrict the RAIS sample to the subset of the six metropolitan areas covered by the PME and to 2002–2015 to align with the available data from the PME. Since trends for men and women are quite similar, in the interest of space, we pool both genders. We also abstract from an analysis of the very top of the earnings distribution and limit our discussion of higher-order moments of earnings changes, as we believe that the modest sample size of the PME prevents a reliable analysis of these outcomes.

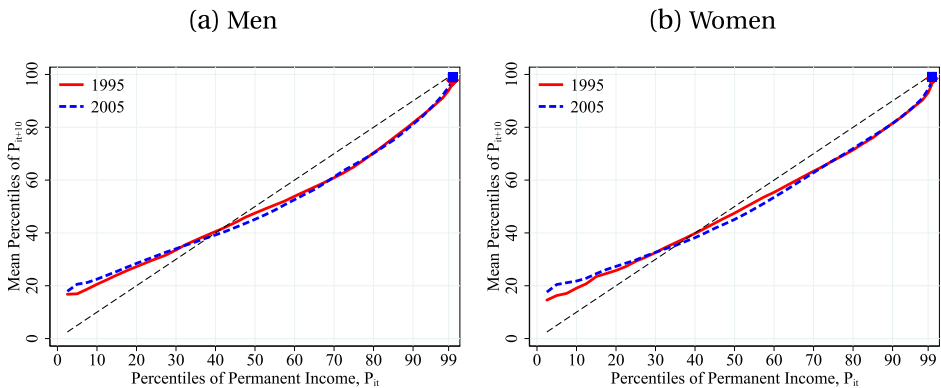


FIGURE 10. Evolution of 10-year mobility over time, by gender. *Note:* Workers aged 25–55. Square and diamond markers denote the top 0.1% of permanent income P_{it} . *Source:* RAIS, 1985–2018.

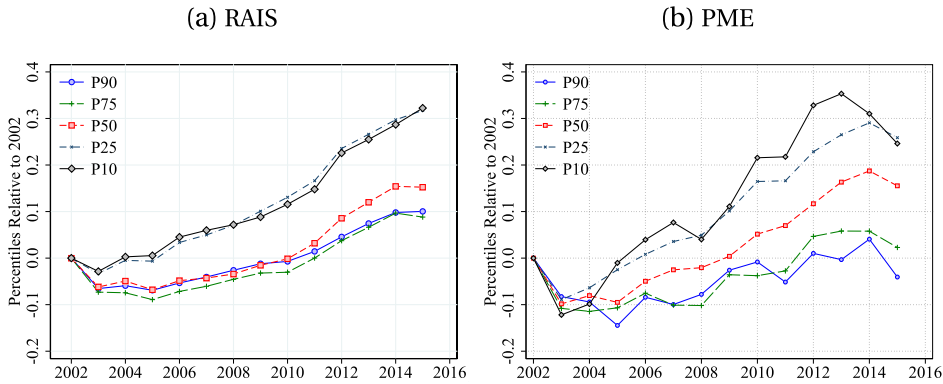


FIGURE 11. Comparison of percentiles of the log real earnings distribution. *Note:* Workers aged 25–55. *Source:* PME and RAIS, 2002–2015.

Figure 11 plots percentiles of the log earnings distribution, normalized to 2002. Note that, in general, the results in panel A differ from those in the previous section, since we here restrict the RAIS data to the subset of the six metropolitan areas covered by the PME. In practice, however, the time trends closely correspond to the trends for the full country in the previous section. Reassuringly, the percentile ratios evolve similarly in the administrative and in the household survey data.

Not surprisingly, given that the percentiles line up closely across the two data sets, Figure 12 finds that measures of inequality follow similar trends in the RAIS and in the PME. That said, the household survey data understate the level of inequality in the administrative data; for instance, the 90–10 percentile ratio is higher by almost 30 log points in the RAIS.

We next turn to a comparison of earnings dynamics in the administrative and survey data. Figure 13 plots dispersion in 1-year earnings changes in the RAIS and the PME separately by gender. The two data sets show broadly similar patterns, although with some important differences, especially during the 2002–2006 period. For instance, ac-

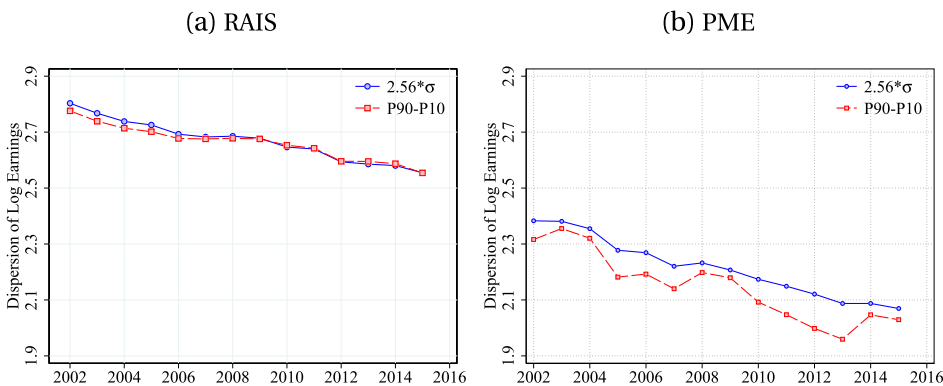


FIGURE 12. Comparison of earnings inequality. *Note:* Workers aged 25–55. *Source:* PME and RAIS, 2002–2015.

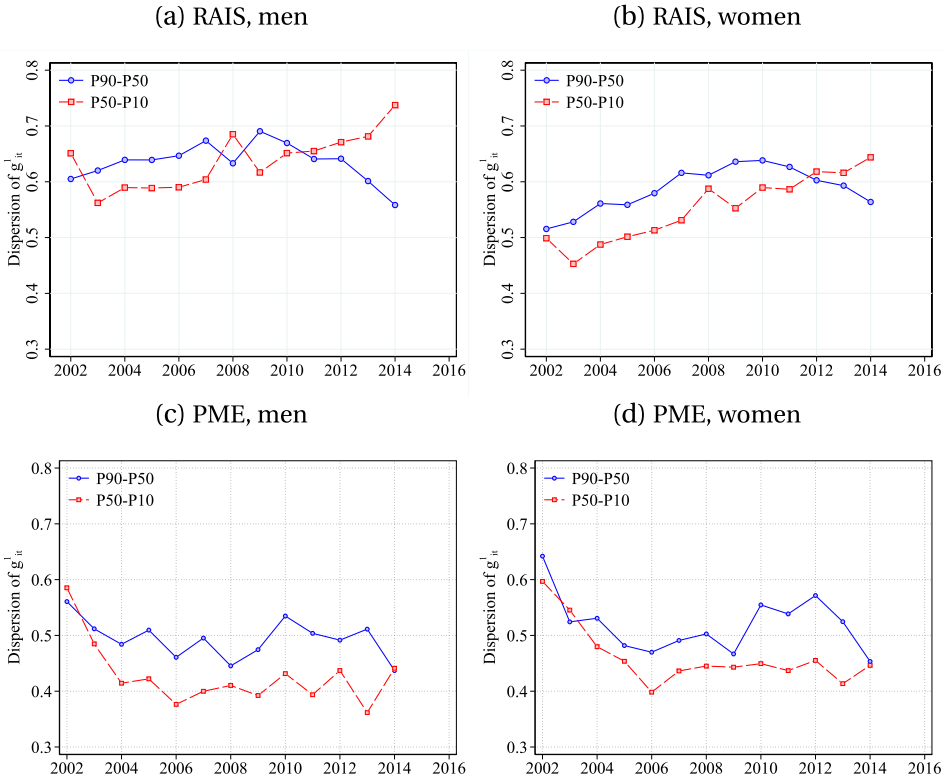


FIGURE 13. Comparison of dispersion of 1-year log earnings changes, by gender. *Note:* Workers aged 25–55. *Source:* RAIS and PME, 2002–2015.

According to PME, dispersion in both the top and the bottom of the distribution of earnings changes for both men and women fell sharply from 2002 to 2004, whereas it was fairly stable in RAIS. One possibility is that the modest sample size of the PME results in a noisy estimate of the underlying population variance of earnings changes. That said, the two surveys give broadly similar results for the period from 2006 onwards, showing a pattern of relatively stable earnings volatility.

4. EARNINGS INEQUALITY AND DYNAMICS IN BRAZIL’S FORMAL AND INFORMAL SECTORS

This section studies earnings inequality in Brazil’s formal and informal sectors as well as earnings dynamics within and across Brazil’s formal and informal sectors.

4.1 Earnings inequality and dynamics within and across sectors

An analysis of informal labor markets is highly relevant in developing countries such as Brazil, which is characterized by a large share of employment in the shadow economy (Ulyssea (2010), Bosch and Esteban-Pretel (2012), Meghir, Narita, and Robin (2015), Ulyssea (2020)). How different is the cross-sectional distribution of earnings in Brazil’s

informal sector relative to that in the formal sector? How do the distributions of earnings changes compare across the two sectors? How have earnings inequality and volatility evolved within each sector? And what are the mechanics by which earnings inequality and volatility in the two sectors differ, both in the cross-section and over time?

To answer these questions, we extend our empirical analysis to Brazil's informal sector, exploiting the power of rich household survey data to complement our analysis of administrative records in the previous sections. We proceed in four steps. First, we study earnings inequality in Brazil's formal and informal sectors. Second, we dissect earnings dynamics within and across the two sectors. Third, we analyze the process of labor market formalization in Brazil over the period from 2002 to 2015. Fourth and finally, we shed light on the large decline in earnings volatility in Brazil over this period.

4.1.1 Earnings inequality in Brazil's formal and informal sectors We begin with an analysis of earnings inequality in Brazil's formal and informal sectors based on the PME household survey data.

4.1.1.1 A first look at the sectoral earnings distributions Figure 14 compares the distribution of (residual) log earnings across the two sectors.¹⁷ We find that mean pay is significantly higher in the formal sector. This confirms the results in Ulyssea (2020), who first estimates and then performs Oaxaca–Blinder decompositions of the informality penalty in Brazil. At the same time, there is significant overlap between the distributions of earnings in each sector, as documented by Meghir, Narita, and Robin (2015). Qualitatively, this pattern is similar for log earnings (panel A) and residual log earnings (panel B).¹⁸

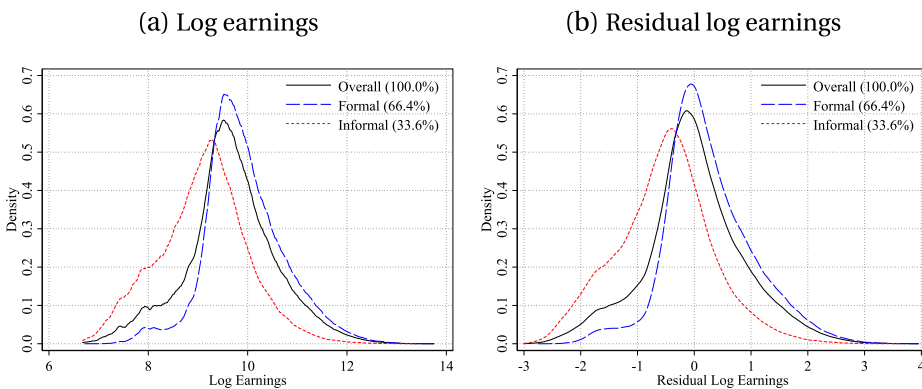


FIGURE 14. Densities of log earnings and residual log earnings, by sector. *Note:* Workers aged 25–55. Kernel densities of log earnings (panel A) and residual log earnings (panel B) in each sector. Residual log earnings are calculated controlling for age and survey wave fixed effects, separately by gender and year. Formal sector includes all employees with a work permit. Informal sector includes all employees without a work permit and the self-employed. *Source:* PME, 2002–2015.

¹⁷Recall that our measure of residual earnings controls for gender-year-specific age dummies.

¹⁸Figures C.3–C.6 in Appendix C of the Online Supplemental Material show that there remains significant overlap across sectors in the distributions of (residual) log earnings by demographic subgroups defined by the intersection of gender, education groups, and age groups in 2002 and 2015.

4.1.1.2 *Mincerian estimates of the informal earnings penalty* Why do workers in Brazil's informal sector earn less on average than those in formal jobs? A second look at Table 2 suggests that there are stark differences in worker composition across sectors. To understand the importance of worker selection for observed differences in mean earnings across sectors, we estimate the informal earnings penalty via a sequence of Mincerian earnings regressions. To this end, we restrict the data to observations that have non-missing information in all relevant control variables. Table 3 shows the result from four specifications with different sets of control variables.

First, we estimate the raw earnings penalty by simply projecting log earnings on an informality indicator. In doing so, we find a 74.4 log points earnings penalty from being informal, which we regard as the baseline value to be explored.

Second, when additionally controlling for year-season-region fixed effects, worker demographics (i.e., gender, race, education, age, number of children), industry, occupation, tenure, and hours worked, the conditional informal earnings penalty is lower by around two-thirds, standing at 26.3 log points. This suggests that worker selection along these dimensions is important for understanding pay across sectors.

Third, adding additional controls for the size of the firm at which a worker is employed, the informal earnings penalty is further cut in half, reaching 12.2 log points. This

TABLE 3. Mincerian estimates of the informal earnings penalty.

	(1)	(2)	(3)	(4)
Informal	-0.744 (0.003)	-0.263 (0.002)	-0.122 (0.003)	-0.116 (0.007)
Year-season-region FE		✓	✓	✓
Gender FE		✓	✓	
Race FE		✓	✓	
Education FE		✓	✓	
Age FE		✓	✓	✓
Number-of-kids FE		✓	✓	✓
Industry FE		✓	✓	✓
Occupation FE		✓	✓	✓
Tenure controls		✓	✓	✓
Hours controls		✓	✓	✓
Firm size FE			✓	✓
Person FE				✓
R^2	0.184	0.699	0.703	0.983
Obs. (weighted)	264,997,721	264,997,721	264,997,721	264,997,721
Obs. (unweighted)	630,021	630,021	630,021	630,021

Note: This table shows estimated regression coefficients, along with standard errors in parentheses. Regressions are estimated on the same subset of observations with nonmissing information for all variables. The dependent variable is log earnings. The independent variable of interest is an indicator for being informally employed. Other control variables include indicators for year-season-region combinations (978 categories), gender (2 categories), race (6 categories), education (4 categories), age (31 categories), number of children living in the household (12 categories), industry (7 categories), occupation (9 categories), log tenure, log hours worked, indicators for firm size (11 categories), and indicators for persons' identity (495,577 categories). The R^2 represents the (unadjusted) coefficient of determination. The weighted number of observations corresponds to the sum of survey sampling weights, while the unweighted number of observations corresponds to the number of data points in the sample. *Source:* PME, 2002–2015.

points toward the importance of firms in understanding informal versus formal labor markets (Ulyssea (2020)).¹⁹

In the fourth and final specification, we add person fixed effects in order to control for (unobserved) permanent individual heterogeneity by using the panel dimension of the data. This effectively identifies the informal earnings penalty of sector switchers, similar to the identification strategy in Alvarez (2020). This most stringent specification yields an informal earnings penalty of 11.6 log points, close to the previous specification that already included a rich set of controls for observable characteristics.

Overall, we conclude that around 84% of the 74.4 log points raw informal earnings penalty is explained by worker selection, with the remainder explained by identical—observationally identical in specifications 1–3, and literally identical in specification 4—workers receiving lower pay in informal employment.

4.1.1.3 Evolution of earnings inequality in both sectors A striking fact documented in Alvarez et al. (2018) is the marked decline in earnings inequality in Brazil's formal sector since the early 2000s. More broadly, there was an equally striking decline in overall earnings inequality across Brazil's formal and informal sectors over this period. Figure 15 shows the evolution of the overall dispersion of (residual) log earnings in Brazil when pooling both sectors. Between 2004 and 2013, overall earnings dispersion declined gradually, both in terms of the scaled standard deviation and the P90-P10 percentile ratio. We find qualitatively similar patterns for log earnings (panel A) and residual log earnings (panel B).

One possible reason behind falling earnings inequality in Brazil is that mean earnings in the two sectors—formal and informal—approached one another over time. In-

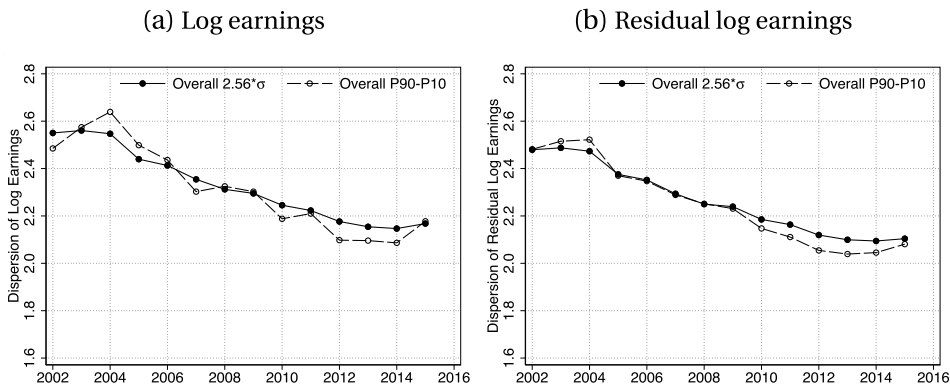


FIGURE 15. Overall dispersion of log earnings and residual log earnings over time. *Note:* Workers aged 25–55. Measures of dispersion of log earnings (panel A) and residual log earnings (panel B). Residual log earnings are calculated controlling for age and survey wave fixed effects, separately by gender and year. Formal sector includes all employees with a work permit. Informal sector includes all employees without a work permit and the self-employed. *Source:* PME, 2002–2015.

¹⁹In this way, our work contributes to recent work highlighting the importance of firm size for worker pay in other contexts (Babina, Ma, Moser, Ouimet, and Zarrutskie (2019), Engbom, Moser, and Sauermaann (forthcoming)).

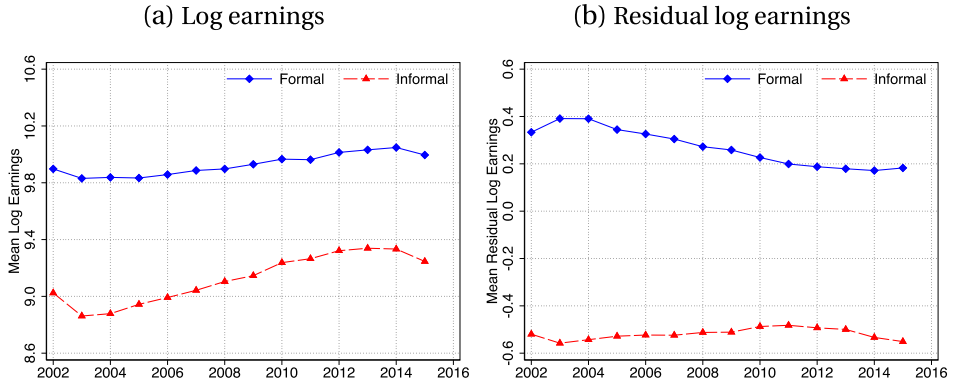


FIGURE 16. Mean (residual) log earnings over time, by sector. *Note:* Workers aged 25–55. Mean log earnings (panel A) and mean residual log earnings (panel B). Residual log earnings are calculated controlling for age and survey wave fixed effects, separately by gender and year. Formal sector includes all employees with a work permit. Informal sector includes all employees without a work permit and the self-employed. *Source:* PME, 2002–2015.

deed, Figure 16 shows that this is the case, both in terms of sectoral means of log earnings (panel A) and residual log earnings (panel B). All else equal, this trend means that a falling between-sector component of the overall variance of earnings has contributed towards the overall decline in earnings inequality.

Another possible reason why overall earnings inequality in Brazil fell is that within-sector earnings dispersion has declined. To investigate this, Figure 17 plots trends in inequality in (residual) log earnings, separately within Brazil’s formal and informal sectors

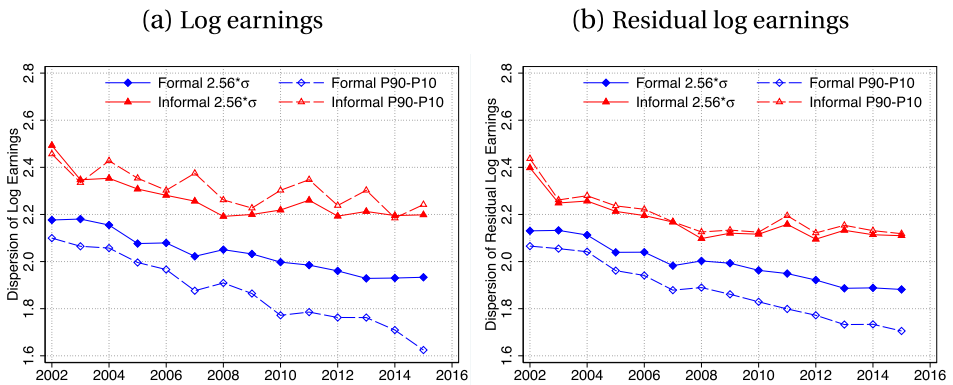


FIGURE 17. Dispersion of (residual) log earnings over time, by sector. *Note:* Workers aged 25–55. Measures of dispersion of log earnings (panel A) and residual log earnings (panel B) in each sector. Residual log earnings are calculated controlling for age and survey wave fixed effects, separately by gender and year. Formal sector includes all employees with a work permit. Informal sector includes all employees without a work permit and the self-employed. *Source:* PME, 2002–2015.

over time.²⁰ As apparent from the analysis above, inequality is higher in Brazil's informal sector relative to its formal sector. Furthermore, inequality has significantly declined in both the formal and informal sector between 2002 and 2015. The fall in inequality is significant in both sectors, although somewhat more pronounced for the formal sector when looking at the P90–10 log percentile ratio. A possible explanation for this is that the rapid increase in Brazil's minimum wage over this period has contributed to a disproportionate reduction in inequality in the formal sector, where it is enforced (Engbom and Moser (forthcoming)). Again, these patterns are qualitatively similar across log earnings (panel A) and residual log earnings (panel B).

4.1.2 Earnings dynamics in Brazil's formal and informal sectors We next assess earnings dynamics in the presence of informality by analyzing the distribution of 1-year residual earnings changes by sectoral transition type. Because workers may change sector across years, we construct four mutually exclusive, collectively exhaustive groups: those who work in the formal sector both in year t and in year $t + 1$ —henceforth *formal-formal* workers; those who work in the informal sector in both years—*informal-informal* workers; and those who work in the formal (informal) sector in year t but in the informal (formal) sector in year $t + 1$ —*formal-informal* and *informal-formal*, respectively.

4.1.2.1 Densities of residual earnings changes Figure 18 shows the estimated densities of residual earnings changes by transition type. Informal-informal workers have more volatile earnings than formal-formal workers, with a higher prevalence of large negative and positive earnings changes. Of course, this pattern should not be interpreted in

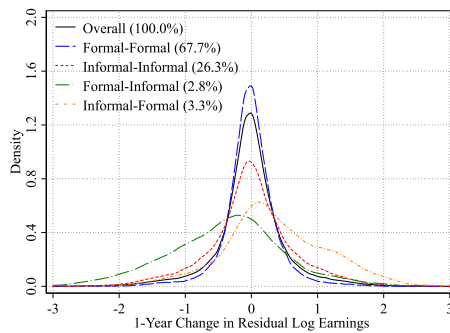


FIGURE 18. Densities of 1-year residual log earnings change, by sectoral transition. *Note:* Workers aged 25–55. Kernel densities of 1-year changes in residual log earnings by worker group. Residuals are calculated controlling for age and survey wave fixed effects, separately by gender and year. Different lines denote different combinations of a worker's current sector of employment and that in the next survey wave (e.g., “Formal-Informal” denotes current employment in the formal sector and employment in the informal sector in the next survey wave). *Source:* PME, 2002–2015.

²⁰Because we apply here the PME sample selection criteria, the pattern for the formal sector differs slightly from that in Figure 12, which applies the RAIS sample selection criteria to the PME household survey data, even though both figures are based on the same data source.

a causal sense, as workers are likely not randomly assigned to sector. Formal-informal workers tend to experience earnings losses, while the opposite is true among informal-formal workers. The fact that a given worker experiences systematic earnings changes when switching sectors suggests that the lower average earnings in the informal sector are not purely due to worker selection, consistent with the evidence presented above and that in Alvarez (2020).²¹

4.1.2.2 Moments of the distribution of residual earnings changes To put our estimates of earnings volatility in Brazil into context, Table 4 summarizes the first four moments of 1-year residual earnings changes in the population and by sector transition. Recalling that year-on-year residual log earnings changes need not sum to zero in an unbalanced panel, we find a small mean change of -0.012 in the overall population. Mean changes are of similarly small magnitudes among formal-formal and informal-informal transitions, ranging from -0.015 to -0.010 . In contrast, sectoral switchers experience residual earnings changes that are large in magnitude and approximately symmetric, with estimated mean changes of -0.388 for formal-informal transitions and 0.361 for informal-formal transitions. At face value, this symmetry speaks against the view that comparative advantage drives worker mobility across sectors because, under such a view, the patterns of earnings gains and losses would be asymmetric.

Turning to the second standardized moment, the overall standard deviation of residual log earnings changes is 0.511 . As suggested by Figure 18, the standard deviation is significantly higher for informal sector stayers (0.652) than for formal sector stayers (0.380). While the overall standard deviation of earnings changes we find in Brazil is similar to that in the US, the standard deviation within the informal sector is closer to that among low-paid workers in the US (Güvenen et al. (2021)). The standard deviation

TABLE 4. Moments of the distribution of 1-year log earnings changes, by sectoral transition.

	Overall	Formal– formal	Informal– informal	Formal– informal	Informal– formal
Mean	-0.012	-0.015	-0.010	-0.388	0.361
Std. dev.	0.511	0.380	0.652	0.880	0.809
Kelley skewness	0.017	0.030	0.013	-0.100	0.112
Excess Crow-Siddiqui kurtosis	2.451	1.683	1.470	0.418	0.448
Obs. (weighted)	140,059,975	94,829,458	36,769,712	3,883,017	4,577,788
Obs. (unweighted)	323,916	216,716	86,212	9786	11,202

Note: All statistics are for workers aged 25–55. The table presents the mean, standard deviation, Kelley skewness, and excess Crow-Siddiqui kurtosis of 1-year residual log earnings changes. Residual earnings flexibly control for age and survey wave fixed effects, separately by gender and year. The weighted number of observations corresponds to the sum of survey sampling weights, while the unweighted number of observations corresponds to the number of data points in the sample. *Source:* PME, 2002–2015.

²¹Figures C.7 and C.8 in Appendix C of the Online Supplemental Material show that a similar pattern to that in Figure 18 obtains for a range of demographic subgroups in 2002 (Figure C.7) and 2015 (Figure C.8). This points in the direction of greater dispersion in residual log earnings changes in the informal sector conditional on various dimensions of worker heterogeneity.

is even higher among workers who switch from formal to informal employment (0.880) and from informal to formal employment (0.809).

In terms of higher-order moments, a few points are worth noting. Starting with the third moment, the overall Kelley skewness and excess Crow–Siddiqui kurtosis of earnings changes are 0.017 and 2.451, respectively, indicating that earnings changes are lightly right-skewed and notably leptokurtic. The Kelley skewness is the most negative for formal-informal transitions (−0.100) and takes on positive values for all other types of transitions, with the highest positive Kelley skewness observed for informal-to-formal transitions (0.112). Moving on to the fourth moment, the excess Crow–Siddiqui kurtosis is 2.451 overall, well above the excess Crow–Siddiqui kurtosis value of zero associated with a Gaussian distribution. While all the distribution of earnings changes is leptokurtic for all transition types, it is highest among formal-formal transitions (1.683) and lowest among formal-informal transitions (0.418).²²

4.1.2.3 Decomposition of earnings volatility Altogether, these results suggest that earnings are significantly more volatile within Brazil's informal sector than within the formal sector, and even more volatile and (positively or negatively) skewed for sector switchers. The higher volatility in Brazil's informal sector may, in turn, be accounted for by a different type of individual working in the informal sector, who may experience more volatile earnings dynamics regardless of sector. Alternatively, earnings volatility may be higher in the informal sector also conditional on workers' type.

To assess the relative importance of differences in worker composition separately from the volatility of earnings for a given type of workers, we consider a shift-share analysis in the spirit of Kitagawa, Oaxaca, and Blinder. Specifically, we first compute 1-year residual earnings changes, controlling for sector interacted with demographic subgroups. Here, we define demographic subgroups based on the intersection of three age groups, gender, two education groups, three industry groups, and two occupation groups.²³ Subsequently, we compute the variance of residual earnings changes in sector $s = \{i, f\}$ specific to demographic group d , $\sigma_{d,s}^2$. Finally, we compute a demographic group's share of employment in sector s , $\omega_{d,s} = N_{d,s}/N_s$, where $N_{d,s}$ is the number of individuals in demographic subgroup d in sector s and N_s is the total number of individuals in sector s , both weighted by the provided survey weights.

Since the mean of the residual earnings changes in sector s for demographic group d is zero by construction, the overall variance of residual earnings changes in sector s , σ_s^2 , can be written as

$$\sigma_s^2 = \frac{1}{N_s} \sum_{i \in s} (\Delta w_i)^2 = \sum_d \frac{N_{d,s}}{N_s} \left(\frac{1}{N_{d,s}} \sum_{i \in d \times s} (\Delta w_i)^2 \right) = \sum_d \omega_{d,s} \sigma_{d,s}^2 \quad (3)$$

²²Since reliably estimating higher-order moments requires large data sets, we caution against attaching too much weight to these statistics, given the modest sample size of the PME household survey data in comparison to the administrative RAIS data.

²³Specifically, the age groups we consider are 25–34, 35–44, 45–55, the education groups are less than high school or high school or more, the industry groups are manufacturing, services and agriculture/other, and the occupation groups are blue or white collar. The size of our data prohibits a reliable estimation of the within-group variance of earnings changes in more detailed subgroups.

We use equation (3) to run a series of counterfactual shift-share decompositions. First, we compute the counterfactual variance of residual earnings changes that would arise if the only difference across sectors were within-group differences in volatility, but demographic composition was the same. To that end, we hold the weights $\omega_{d,s} = \omega_{d,f}$, for $s = \{f, i\}$, fixed at their formal sector values, but let the within-group variances vary as in the data across sectors, $\sigma_{d,s}^2$. We label this the *returns* channel since it isolates the effect of within-group differences in volatility holding composition fixed.

Second, we hold within-group differences in volatility fixed at their formal sector values, $\sigma_{d,s}^2 = \sigma_{d,f}^2$, for $s = \{f, i\}$, but let demographic composition, $\omega_{d,s}$, evolve as in the data. We label the resulting counterfactual variance of residual earnings changes in the informal sector the *composition* channel, since it isolates the effect of differences in composition holding within-group differences in volatility fixed. We also compute the counterfactual variance of residual earnings changes in the informal sector letting only composition adjust in one of the dimensions, holding within-group variances as well as the demographic composition in all other dimensions fixed at their formal sector level, as $\sum_d \tilde{w}_{d,s} \sigma_{d,f}^2$. We compute the counterfactual weights $\tilde{w}_{d,s}$ such that the marginal distribution of the particular demographic characteristic in focus corresponds to that in the informal sector, while the marginal distributions in all other demographic dimensions remain those in the formal sector.

Table 5 presents the results from this shift-share analysis. Little of the higher volatility in the informal sector is accounted for by differences in workforce composition, at least not in the relatively coarse characteristics that we consider. Instead, the great majority of the higher volatility in the informal sector takes place conditional on such worker characteristics, that is, a higher within-group volatility of earnings in the informal sector.

4.1.3 Dissecting the process of labor market formalization in Brazil Between 2002 and 2015, the Brazilian economy underwent a rapid transformation, both politically and economically. A notable aspect of this transformation has been the marked decline in informal employment rates over this period.

4.1.3.1 Evolution of informality shares The evolution of informality shares is illustrated in Figure 19. Panel A shows that the informal share has declined over the past 10–15 years, dropping from 39% in 2002 to 25% in 2015. Panel B dissects the decline across the earnings distribution. In the bottom quartile of the earnings distribution, almost three-quarters of workers were in the informal sector in 2002, while the corresponding figure in the top quartile was 17%. Over time, the decline in informality was particularly pronounced at the bottom of the distribution—the share of the first quartile working in the informal sector fell by 19 percentage points from 2002 to 2015. Yet, the process of labor market formalization was widespread throughout the earnings distribution, with the share of the top quartile working in the informal sector also declining by 7 percentage points. Despite the marked decline in informality overall, informal employment contin-

TABLE 5. Shift-share decomposition of relative earnings volatility across sectors.

	Earnings volatility, σ_s^2	
	Formal	Informal
<i>Panel A. Overall</i>	0.174	0.449
<i>Panel B. Shift-Share Decomposition</i>		
Returns		0.493
Composition		0.178
Age		0.174
Education		0.173
Gender		0.174
Industry		0.177
Occupation		0.175
Minimum obs.	127	46
Maximum obs.	4969	3594
Total obs.	107,972	47,063

Note: Variance of residual earnings changes within the formal and informal sectors, as well as a shift-share decomposition of the differences in the variance of residual earnings changes by sector. *Returns:* Counterfactual variance of residual earnings changes in the informal sector holding fixed demographic composition at its formal sector values, but letting the within-group variance of residual earnings changes vary across sectors as in the data. *Composition:* Counterfactual variance of residual earnings changes in the informal sector holding the within-group variance of residual earnings changes fixed at its formal sector values, but letting the demographic composition vary across sectors as in the data. The first line (“Composition”) corresponds to the impact of changing the demographic composition along all dimensions (age, gender, education, industry, and occupation). *Age/Education/Gender/Industry/Occupation:* Impact of differences in one composition along one demographic attribute (age, education, gender, industry, and occupation) at a time, holding fixed composition in all other dimensions. Min. (max) observations: Number of observations in the sector-demographic subgroup with the fewest (most) observations. All statistics are for workers aged 25–55. *Source:* PME, 2002–2015.

ues to account for over half of employment among workers in the bottom quartile of the earnings distribution.²⁴

4.1.3.2 Relation to sectoral flow rates The decrease in informality in Brazil between 2002 and 2015 is closely related to the evolution of labor market flow rates over this period. Figure 20 provides a high-level overview of sectoral transition rates within and between Brazil’s formal and informal sectors.²⁵

²⁴As an additional robustness check, Appendix D of the Online Supplemental Material shows that multiple job holdings are quantitatively not relevant for understanding formality rates in the cross-section or over time.

²⁵Figures C.1 and C.2 in Appendix C of the Online Supplemental Material show the evolution of the composition of formal-formal and informal-informal transition rates by workers’ educational attainment (Figure C.1) and age (Figure C.2) over time.

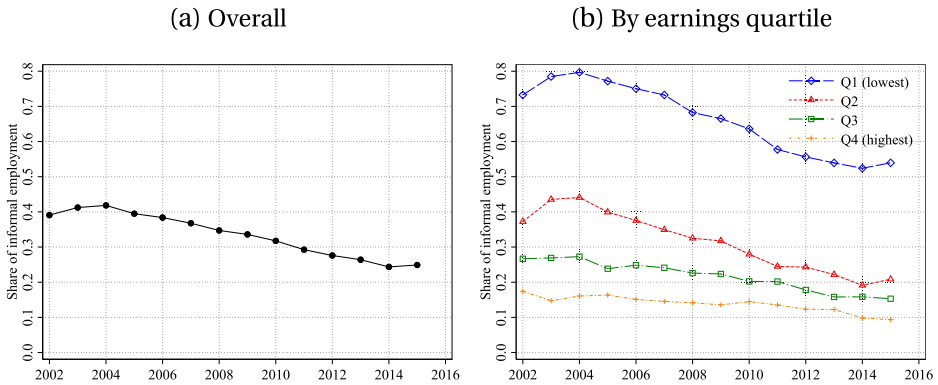


FIGURE 19. Evolution of informal employment shares. *Note:* All statistics are for workers aged 25–55. Residual earnings flexibly control for age and survey wave fixed effects, separately by gender and year. *Source:* PME, 2002–2015.

While a majority of workers remain attached to the same—formal or informal—sector between consecutive years, sector switchers and leavers are of particular interest in assessing the evolution of sectoral shares. To dig deeper into sector switchers and leavers, Figure 21 shows the 1-year-forward transition rates out of the formal sector (panel A) and out of the informal sector (panel B). Sector transitions are categorized as either sector switchers (i.e., from formal to informal in panel A, and from informal to formal in panel B) or sector leavers (i.e., workers moving into nonemployment). For this purpose, we define *nonemployment* as no employment in either the formal or the informal sector (i.e., including unemployed workers and those out of the labor force).

Several results are noteworthy. A striking observation is that transition rates out of the formal sector (panel A) are around four times smaller than transition rates out of

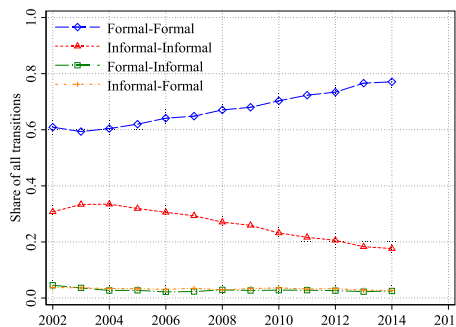


FIGURE 20. Transition rates within and between the formal and informal sectors. *Note:* Workers aged 25–55. “Formal-Formal” corresponds to workers staying employed in the formal sector between consecutive years. Similarly, “Informal-Informal” corresponds to workers staying employed in the informal sector. “Formal-Informal” corresponds to workers switching from formal to informal employment between years. “Informal-Formal” corresponds to workers switching from informal to formal employment between years. *Source:* PME, 2002–2015.

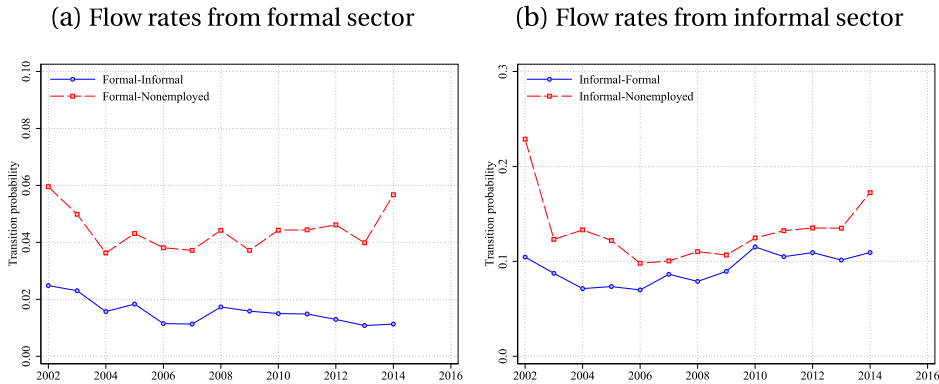


FIGURE 21. Evolution of sectoral flow rates, by origin sector. *Note:* Workers aged 25–55. *Source:* PME, 2002–2015.

the informal sector (panel B). Given that between 2002 and 2015, the formal sector was only between one-and-a-half and three times as large as the informal sector, a balance-flow-equation logic implies that there must have been a net inflow into formality over this period. A second striking observation is that the formal-to-informal transition rate has approximately halved, from around 2.5% to around 1.2%, over this period. At the same time, the informal-to-formal transition rate has slightly increased. Exit rates into nonemployment have been U-shaped in both sectors over this period. We conclude from this analysis that persistently asymmetric flow rates have contributed toward the decline in informal employment in Brazil over this period.

4.1.3.3 Heterogeneity in flow rates by earnings rank While in the aggregate, asymmetric flow rates have contributed toward Brazil's process of formalization, we document significant heterogeneity in sectoral flows across the earnings distribution. Figure 22 plots the mean transition rates as a function of population earnings rank for formal sector workers in panel A and for informal sector workers in panel B.

A few points are noteworthy. First, for both formal and informal workers, the probability of staying in the same sector in consecutive years (the omitted category in each panel of Figure 22) far outweighs that of switching sectors or leaving employment altogether. Second, formal workers are relatively more attached to their sector than informal workers are. Third, there is a marked decrease in exit rates from the formal sector toward informality and nonemployment but an increase in exit rates from the informal sector toward formality and nonemployment toward higher earnings percentiles.

4.1.4 Understanding the decline in earnings volatility in Brazil While it is well known that earnings inequality declined in Brazil since the early 2000s (Alvarez et al. (2018), Firpo and Portella (2019)), less is known about the evolution of earnings volatility in Brazil's formal and informal sectors.

4.1.4.1 Aggregate time series Figure 23 shows that, over the period from 2002 to 2015, there has been a marked decline in earnings volatility in Brazil. Pooling the formal and

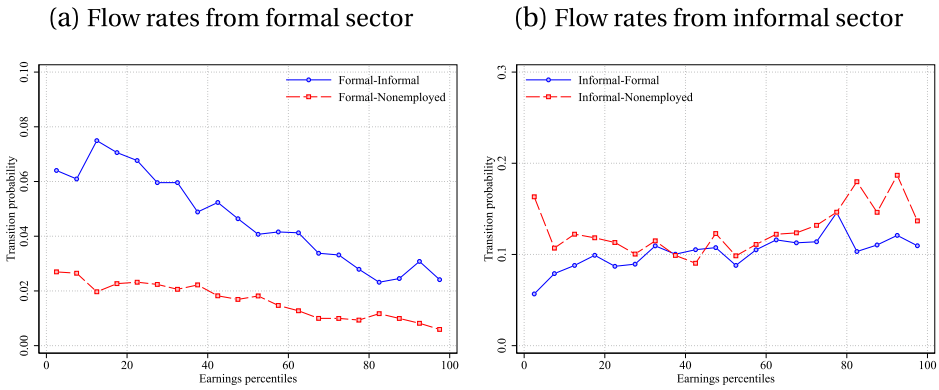


FIGURE 22. Cross-sectional heterogeneity in sectoral flow rates, by origin sector. *Note:* Workers aged 25–55. *Source:* PME, 2002–2015.

informal sectors, the figure shows that earnings volatility fell in the early 2000s and subsequently remained relatively stable from around 2006 onwards.

Besides the evolution of sectoral transition rates documented above, the aggregate decline in earnings volatility in Brazil depends on two additional factors. First, the mean earnings change associated with each type of sectoral transition and, second, the dispersion in earnings changes associated with each type of sectoral transition.

4.1.4.2 *Trends in mean earnings changes by type of sectoral transition* First, to understand the evolution of mean earnings changes by sectoral transition, Figure 24 plots the evolution of mean earnings changes by a worker’s type of sectoral transition. There was a marked decline in the mean change in residual earnings for workers transitioning from the informal to the formal sector between 2002 and 2014. Driven by this decline and relatively stable means for other sectoral transition groups, the between-group dispersion in means was significantly lower in 2014 than it was in 2002.

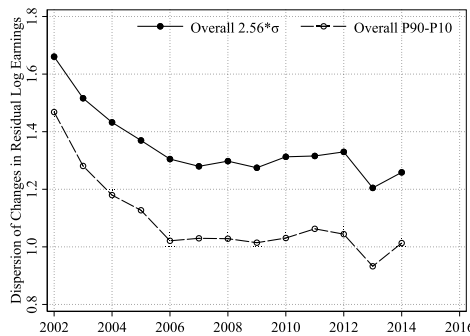


FIGURE 23. Overall dispersion in 1-year residual log earnings changes. *Note:* Workers aged 25–55. Residual earnings control for age and survey wave fixed effects, separately by gender and year. *Source:* PME, 2002–2015.

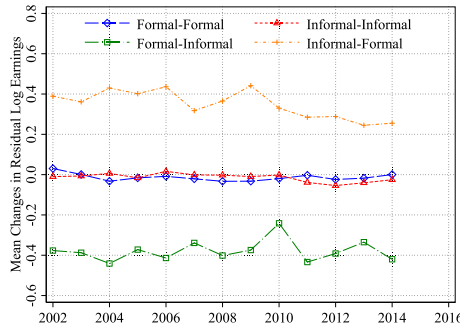


FIGURE 24. Evolution of mean earnings changes by type of sectoral transition. *Note:* Workers aged 25–55. *Source:* PME, 2002–2015.

4.1.4.3 *Trends in dispersion of earnings changes by type of sectoral transition* Second, to understand the role of within-sector changes in earnings volatility, Figure 25 plots the evolution of the scaled standard deviation (panel A) and the P90-P10 percentile ratio of changes in residual log earnings separately by worker transition. As in Figure 20, we distinguish between four transition categories: formal-formal, informal-informal, formal-informal, and informal-formal. The largest decline in earnings volatility occurred within the informal sector stayers. There were less pronounced declines in earnings volatility for formal sector stayers as well as for switchers from the formal into the informal sector. In contrast, the earnings volatility of switchers from the informal to the formal sector actually increased. Qualitatively, these insights are similar across both measures of dispersion of earnings changes.

4.1.4.4 *Decomposition of earnings volatility* Combining our analysis of sectoral transition shares from Section 4.1.3 with the evolution of means and dispersion of earnings volatility documented above, we now conduct a between- versus within-group decomposition of the variance of 1-year changes in residual log earnings by type of sectoral transition groups (i.e., formal-formal, informal-informal, formal-informal, and

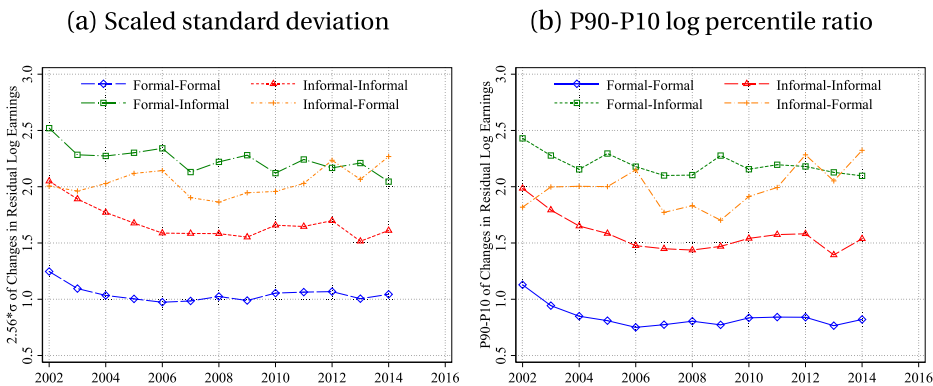


FIGURE 25. Evolution of earnings volatility by type of sectoral transition. *Note:* Workers aged 25–55. *Source:* PME, 2002–2015.

informal-formal transitions). At any point in time, the total variance of residual earnings changes, g_{it} , can be decomposed into two terms, namely a between-group component and a within-group component:

$$\underbrace{\frac{1}{N_t} \sum_i (g_{it} - \bar{g}_t)^2}_{\text{total variance}} = \underbrace{\sum_{s \in S} \frac{N_{st}}{N_t} (\bar{g}_{st} - \bar{g}_t)^2}_{\text{between component}} + \underbrace{\sum_{s \in S} \frac{N_{st}}{N_t} \times \frac{1}{N_{st}} \sum_{i \in S} (g_{it} - \bar{g}_{st})^2}_{\substack{\text{within component} \\ \text{composition channel} \quad \text{returns channel}}}, \quad (4)$$

where $\bar{g}_t = \frac{1}{N_t} \sum_i g_{it}$ is the grand average residual in year t and $\bar{g}_{st} = \frac{1}{N_{st}} \sum_{i \in S} g_{it}$ is the average residual within group s in year t . Note that \bar{g}_t is not identically equal to zero, because our analysis residualizes earnings instead of earnings changes together with the fact that we work with an unbalanced panel. In a balanced panel, if residual earnings have mean zero each year, then earnings changes also have mean zero. But since our panel is unbalanced, the same is not necessarily true in our case, depending on the pattern of household survey responses.²⁶ Note that \bar{g}_{st} is not identically equal to zero, both because the age and education effects do not vary by group and also because the panel is not balanced.

Since different groups are characterized by different volatilities of earnings, two factors in turn contribute to changes in the within component in (4) over time.²⁷ First, changes in the employment weights, $\frac{N_{st}}{N_t}$, of groups in (4) by themselves lead to changes in overall volatility through a composition channel. Second, within-group changes in volatility, $\frac{1}{N_{st}} \sum_{i \in S} (g_{it} - \bar{g}_{st})^2$, lead to changes in overall volatility, holding composition fixed, which we refer to as the returns channel.

Panel A of Figure 26 decomposes the overall change in the variance of earnings changes into its between and within components based on (4). Changes within groups in the volatility of earnings account for the great majority of the fall in earnings volatility over this period.

Panel B of Figure 26 further dissects the within component from the previous decomposition in equation (4) into a composition channel and a returns channel. We isolate the role of these two forces using a shift-share approach (as standard in shift-share analysis, the two components do not add up to the total change). That is, to compute the composition channel, we hold the within-group variances fixed at their initial level and only change the employment weights as in the data. To compute the returns channel,

²⁶That the distribution of earnings changes must have mean zero in a balanced panel, but need not have mean zero in an unbalanced panel, is straightforward to verify. The mean change in residual log earnings between year t and $t + 1$ is

$$\frac{1}{N_t} \sum_i g_{it} = \frac{1}{N_t} \sum_i (w_{it+1} - w_{it}) = \frac{N_{t+1}}{N_t} \frac{1}{N_{t+1}} \sum_i w_{it+1} - \frac{1}{N_t} \sum_i w_{it}.$$

In a balanced panel, this expression equals zero because both $\frac{1}{N_{t+1}} \sum_i w_{it+1} = 0$ and $\frac{1}{N_t} \sum_i w_{it} = 0$ by the nature of their both being the sum of mean-zero residuals in a balanced panel. Since neither summand need to evaluate to zero in an unbalanced panel, the grand average also need not be zero in that case.

²⁷While the same is true also for the between component in (4), we focus on the within component, since that accounts for the great majority of the changes in volatility over this period.

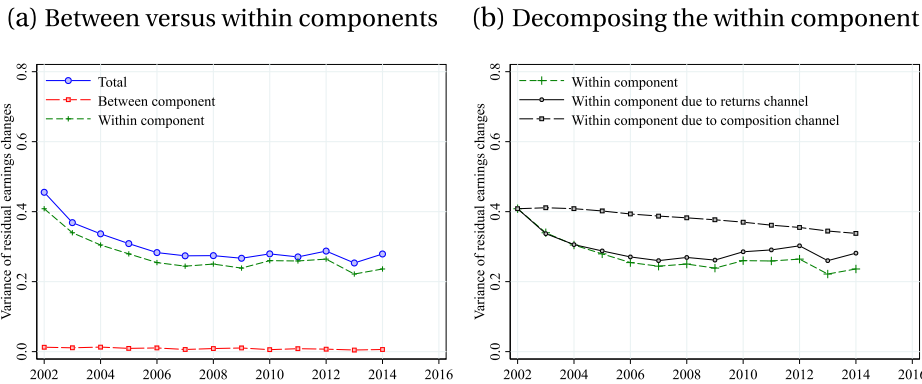


FIGURE 26. Decomposition of the evolution of earnings volatility. *Note:* Workers aged 25–55. Panel A shows a between- versus within-group decomposition of the variance of earnings changes based on (4). Panel B shows a shift-share analysis of the within component of (4). The returns channel holds the sector composition fixed at its initial level and letting the within-group variances evolve as in the data. The composition channel holds the within-group variances fixed at their initial level and letting the sector composition evolve as in the data. *Source:* PME, 2002–2015.

we hold the employment weights fixed at their initial level and only change the within-group variances as in the data. Within-group account for a larger decline in volatility than compositional shifts, although the effect of the latter is also significant. The composition effect arises primarily because employment has gravitated toward formal-formal workers over this period, a shift that is characterized by lower volatility of earnings.²⁸

5. CONCLUSION

In this paper, we document new facts on earnings inequality, dynamics, and mobility in Brazil. Among workers in Brazil's formal sector, there has been a remarkable decrease in earnings inequality, driven by bottom-led growth in real earnings, since the late 1990s. At the same time, the dispersion of earnings changes decreased markedly. Higher-order moments of the distribution of earnings changes differ in levels but show cyclical movements similar to those previously documented in developed countries such as the US. Earnings mobility is comparatively high in Brazil, especially at the bottom of the distribution.

We leverage detailed household survey data to study earnings inequality and dynamics in Brazil's formal and informal sectors. Compared with formal sector workers, there is significantly higher dispersion of earnings changes among informal sector work-

²⁸As an additional robustness check, in Appendix E of the Online Supplemental Material we present a between- versus within-group decomposition of the overall variance of earnings changes by worker education groups. Consistent with the results presented here, we find that most of the decline in the overall variance of earnings changes occurs within, rather than between, education groups.

ers and significantly (positively or negatively) skewed earnings changes among sector switchers. Finally, we find a large decrease in the economy-wide dispersion of earnings changes, which is driven mostly by the within-sector evolution of earnings volatility.

Our work points to two interesting avenues for future research. The first avenue is to shed further light onto the microeconomic sources of Brazil's decline in earnings inequality and volatility that we document. The second avenue is to assess the macroeconomic consequences of this decline as well as its implications for social insurance and other policies.

REFERENCES

- Almeida, Rita and Pedro Carneiro (2012), "Enforcement of labor regulation and informality." *American Economic Journal: Applied Economics*, 4, 64–89. [1408]
- Alvarez, Jorge, Felipe Benguria, Niklas Engbom, and Christian Moser (2018), "Firms and the decline of earnings inequality in Brazil." *American Economic Journal: Macroeconomics*, 10, 149–189. [1407, 1408, 1430, 1438]
- Alvarez, Jorge A. (2020), "The agricultural wage gap: Evidence from Brazilian microdata." *American Economic Journal: Macroeconomics*, 12, 153–173. [1430, 1433]
- Ayres, João, Márcio Garcia, Diogo Guillén, Patrick Kehoe, José A. Scheinkman, and Teresa Ter-Minassian (2021), "The history of Brazil." In *A Monetary and Fiscal History of Latin America, 1960–2017*, 133–196, University of Minnesota Press. [1411]
- Babina, Tania, Wenting Ma, Christian Moser, Paige Ouimet, and Rebecca Zarrutskie (2019), "Pay, employment, and dynamics of young firms." Working Paper. [1430]
- Barros, Ricardo, Mirela de Carvalho, Samuel Franco, and Rosane Mendonça (2010), "Markets, the state, and the dynamics of inequality in Brazil." In *Declining Inequality in Latin America: A Decade of Progress?* (Luis F. López-Calva and Nora Lustig, eds.), 134–174, Brookings Institution Press, Washington, D.C., Chapter 6. [1406]
- Benguria, Felipe, Felipe Saffie, and Sergio Urzúa (2021), "The transmission of commodity price super-cycles." NBER Working Paper No. 24560. [1411]
- Blanco, Andres, Bernardo Diaz de Astarloa, Andres Drenik, Christian Moser, and Danilo Trupkin (2022), "The evolution of the earnings distribution in a volatile economy: Evidence from Argentina." *Quantitative Economics*, 13, 1361–1403. [1407]
- Bloom, Nicholas, Fatih Guvenen, Luigi Pistaferri, John Sabelhaus, Sergio Salgado, and Jae Song (2017), "The great micro moderation." Working Paper. [1409]
- Bosch, Mariano and Julen Esteban-Pretel (2012), "Job creation and job destruction in the presence of informal markets." *Journal of Development Economics*, 98, 270–286. [1427]
- Bowlus, Audra, Émilien Gouin-Bonenfant, Huju Liu, Lance Lochner, and Youngmin Park (2022), "Four decades of Canadian earnings inequality and dynamics across workers and firms." *Quantitative Economics*, 13, 1447–1491. [1407]

Busch, Christopher, David Domeij, Fatih Guvenen, and Rocio Madera (2022), “Skewed idiosyncratic income risk over the business cycle: Sources and insurance.” *American Economic Journal: Macroeconomics*, 14, 207–242. [1421]

Dix-Carneiro, Rafael, Pinelopi Koujianou Goldberg, Costas Meghir, and Gabriel Ulyssea (2021), “Trade and informality in the presence of labor market frictions and regulations.” Working Paper. [1408]

Dix-Carneiro, Rafael and Brian K. Kovak (2015), “Trade liberalization and the skill premium: A local labor markets approach.” *American Economic Review*, 105, 551–557. [1408]

Dix-Carneiro, Rafael and Brian K. Kovak (2017), “Trade liberalization and regional dynamics.” *American Economic Review*, 107, 2908–2946. [1423]

Dix-Carneiro, Rafael and Brian K. Kovak (2019), “Margins of labor market adjustment to trade.” *Journal of International Economics*, 117, 125–142. [1408]

Engbom, Niklas, Gustavo Gonzaga, Christian Moser, and Roberta Olivieri (2022), “Supplement to ‘Earnings inequality and dynamics in the presence of informality: The case of Brazil.’” *Quantitative Economics Supplemental Material*, 13, <https://doi.org/10.3982/QE1855>. [1413]

Engbom, Niklas and Christian Moser (forthcoming), “Earnings inequality and the minimum wage: Evidence from Brazil.” *American Economic Review*. [1409, 1411, 1412, 1432]

Engbom, Niklas, Christian Moser, and Jan Sauermann (forthcoming), “Firm pay dynamics.” *Journal of Econometrics*. [1430]

Fairris, David and Erik Jonasson (2020), “Determinants of changing informal employment in Brazil, 2000–2010.” *Journal of Development Economics*, 1, 1–43. [1408]

Ferreira, Francisco H. G., Sergio P. Firpo, and Julián Messina (2017), “Ageing poorly?: Accounting for the decline in earnings inequality in Brazil, 1995–2012.” IDB Working Paper Series No. 792. [1408]

Ferreira, Francisco H. G., Phillippe G. Leite, and Matthew Wai-Poi (2007), “Trade liberalization, employment flows, and wage inequality in Brazil.” United Nations University Working Paper. [1408]

Firpo, Sergio and Alysson Portella (2019), “Decline in wage inequality in Brazil: A survey.” Policy Research Working Paper 9096, The World Bank. [1406, 1408, 1438]

Gomes, Diego B. P., Felipe S. Iachan, and Cezar Santos (2020), “Labor earnings dynamics in a developing economy with a large informal sector.” *Journal of Economic Dynamics and Control*, 113. [1409]

Gonzaga, Gustavo, Naércio Menezes Filho, and Cristina Terra (2006), “Trade liberalization and the evolution of skill earnings differentials in Brazil.” *Journal of International Economics*, 68, 345–367. [1408]

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song (2021), “What do data on millions of U.S. workers reveal about lifecycle earnings dynamics?” *Econometrica*, 89, 2303–2339. [1409, 1423, 1433]

Guvenen, Fatih, Serdar Ozkan, and Jae Song (2014), “The nature of countercyclical income risk.” *Journal of Political Economy*, 122, 621–660. [1409]

Haanwinckel, Daniel (2020), “Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution.” Working Paper. [1411]

Haanwinckel, Daniel and Rodrigo R. Soares (2021), “Workforce composition, productivity, and labour regulations in a compensating differentials theory of informality.” *Review of Economic Studies*, 88, 2970–3010. [1408]

Hoffmann, Eran B. and Davide Malacrino (2019), “Employment time and the cyclicity of earnings growth.” *Journal of Public Economics*, 169, 160–171. [1409]

Martínez, Tomás and Ursula Mello (2021), “Trade shocks and higher-order earnings risk in local labor markets.” Working Paper. [1423]

McKinney, Kevin L., John M. Abowd, and Hubert P. Janicki (2022), “U.S. long-term earnings outcomes by sex, race, ethnicity, and place of birth.” *Quantitative Economics*, 13, 1879–1945. [1407]

Meghir, Costas, Renata Narita, and Jean-Marc Robin (2015), “Wages and informality in developing countries.” *American Economic Review*, 105, 1509–1546. [1408, 1427, 1428]

Moffitt, Robert A. and Peter Gottschalk (1995), “Trends in the variances of permanent and transitory earnings in the U.S. and their relation to earnings mobility.” Boston College Working Papers in Economics. [1409]

Morchio, Iacopo and Christian Moser (2020), “The gender gap: Micro sources and macro consequences.” Working Paper. [1409]

Ponczek, Vladimir and Gabriel Ulyssea (2021), “Enforcement of labour regulation and the labour market effects of trade: Evidence from Brazil.” *The Economic Journal*, 132, 361–390. [1408]

Puggioni, Daniela, Mariana Calderón, Alfonso Cebreros Zurita, Leó Fernández Bujanda, José Antonio Inguanzo González, and David Jaume (2022), “Inequality, income dynamics, and worker transitions: The case of Mexico.” *Quantitative Economics*, 13, 1669–1705. [1407]

Ribas, Rafael Perez, and Sergei Suarez Dillon Soares (2008), “Sobre o painel da pesquisa mensal de emprego (pme) do ibge.” IPEA Discussion Paper 1348. [1416]

Sabelhaus, John and Jae Song (2010), “The great moderation in micro labor earnings.” *Journal of Monetary Economics*, 57, 391–403. [1409]

Seminario-Amez, Cristhian (2021), “Informal rungs on the job ladder: Theory and evidence from Brazil.” Working Paper. [1408]

Shorrocks, Anthony (1978), "Income inequality and income mobility." *Journal of Economic Theory*, 19, 376–393. [1407]

Ulyssea, Gabriel (2010), "Regulation of entry, labor market institutions and the informal sector." *Journal of Development Economics*, 91, 87–99. [1427]

Ulyssea, Gabriel (2018), "Firms, informality, and development: Theory and evidence from Brazil." *American Economic Review*, 108, 2015–2047. [1408]

Ulyssea, Gabriel (2020), "Informality: Causes and consequences for development." *Annual Review of Economics*, 12, 525–546. [1408, 1427, 1428, 1430]

Co-editor Giovanni L. Violante handled this manuscript.

Manuscript received 24 February, 2021; final version accepted 18 February, 2022; available online 23 March, 2022.