DOCUMENTOS DE TRABAJO

Labor Earnings Dispersion in Chile: Decomposition, Dynamics and the Role of Firms

Rosario Aldunate
Gabriela Contreras
Matías Tapia

N° 892 Noviembre 2020
BANCO CENTRAL DE CHILE

BANCO CENTRAL DE CHILE

CENTRAL BANK OF CHILE

La serie Documentos de Trabajo es una publicación del Banco Central de Chile que divulga los trabajos de investigación económica realizados por profesionales de esta institución o encargados por ella a terceros. El objetivo de la serie es aportar al debate temas relevantes y presentar nuevos enfoques en el análisis de los mismos. La difusión de los Documentos de Trabajo sólo intenta facilitar el intercambio de ideas y dar a conocer investigaciones, con carácter preliminar, para su discusión y comentarios.

La publicación de los Documentos de Trabajo no está sujeta a la aprobación previa de los miembros del Consejo del Banco Central de Chile. Tanto el contenido de los Documentos de Trabajo como también los análisis y conclusiones que de ellos se deriven, son de exclusiva responsabilidad de su o sus autores y no reflejan necesariamente la opinión del Banco Central de Chile o de sus Consejeros.

The Working Papers series of the Central Bank of Chile disseminates economic research conducted by Central Bank staff or third parties under the sponsorship of the Bank. The purpose of the series is to contribute to the discussion of relevant issues and develop new analytical or empirical approaches in their analyses. The only aim of the Working Papers is to disseminate preliminary research for its discussion and comments.

Publication of Working Papers is not subject to previous approval by the members of the Board of the Central Bank. The views and conclusions presented in the papers are exclusively those of the author(s) and do not necessarily reflect the position of the Central Bank of Chile or of the Board members.

Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180, Santiago, Chile
Teléfono: (56-2) 3882475; Fax: (56-2) 3882231

Documento de Trabajo
N° 892

Working Paper
N° 892

Labor Earnings Dispersion in Chile: Decomposition, Dynamics and the Role of Firms*

Rosario Aldunate
Central Bank of Chile

Gabriela Contreras
Central Bank of Chile

Matías Tapia
Central Bank of Chile

Abstract

We use a matched employer-employee census of formal employment in Chile to characterize the distribution of formal labor earnings between 2005 and 2016. We decompose the overall dispersion in earnings across workers into between and within-firm components, and then use a two-way fixed effect model with no complementarities for individual earnings (the AKM model) to compare the contributions of firm and worker heterogeneity toward changes in dispersion. First, we find a decline in the dispersion of labor earnings throughout the decade, which is driven almost completely by a reduction in the variance of average earnings between firms. The dispersion of earnings within firms, which explain more than half of the overall dispersion and correlates strongly with productivity at the firm level, did not change. Second, AKM estimates show that systematic differences across workers explain the bulk of earning differences, and that the reduction in worker heterogeneity was the main driver towards a more compact earnings distribution, an effect that was complemented by weaker sorting patterns. Finally, although our results suggest that the AKM model provides a good first-order approximation to the labor earnings determination process, we use an alternative specification that allows for worker-firm complementarities. This estimation suggests a stronger role for sorting and an even weaker role for firms in explaining labor earnings differences in Chile.

Resumen

En base a un censo de empleo formal en Chile con datos a nivel de firma-trabajador caracterizamos la distribución de ingresos laborales formales entre 2005 y 2016. Descomponemos la dispersión agregada de ingresos entre trabajadores en diferencias en ingresos promedio entre firmas y diferencias en ingreso al interior de las firmas, y después utilizamos un modelo de ingresos individuales con efectos fijos y sin complementariedades (el modelo AKM) para comparar el rol de que juegan la heterogeneidad de firmas y trabajadores para explicar la dispersión y su evolución en el tiempo. Nuestros resultados muestran una caída en la dispersión de ingresos durante la década, la cual es asocia casi completamente a una caída en las diferencias en salarios promedio entre empresas. La dispersión de ingresos al interior de las empresas, que explica más de la mitad de la dispersión total y que correlaciona fuertemente con la productividad de las firmas, no cambió durante el período. En segundo lugar, las estimaciones en base al modelo AKM muestran que la dispersión de ingresos se explica fundamentalmente por diferencias sistemáticas entre trabajadores, mientras que las diferencias sistemáticas entre firmas juegan un rol menor. La reducción en la dispersión se asocia fundamentalmente a una caída en la heterogeneidad entre trabajadores, un efecto que se vio reforzado por un debilitamiento en los patrones de sorting trabajador-firma. Finalmente, aunque el

*We thank Stephane Bonhomme, Fatih Guvenen, Wilson Guzman, Steven Durlauf, and seminar participants at SECHI Meetings 2019, the Central Bank of Chile Workshop on Labor Markets and Productivity, and the RIDGE Workshop on Poverty and Inequality 2019 for valuable comments and suggestions. The views expressed in this paper are exclusively those of the authors and do not necessarily reflect the position of the Central Bank of Chile or its Board members. Any errors or omissions are responsibility of the authors. Emails: raldunate@bcentral.cl, gcontreras@bcentral.cl and mtapia@bcentral.cl.

modelo AKM es una buena aproximación al proceso de generación de ingresos, también utilizamos una especificación alternativa que permita que existan complementariedades entre trabajadores y empresas. Esta estimación sugiere una mayor importancia del sorting y una importancia aún menor de las firmas para explicar los patrones de ingreso en Chile.

I Introduction

The analysis of the determinants and dynamics of differences in earnings across workers is a central issue in economics, as it relates directly to several core questions that are relevant both for academic and policy purposes. For example, a precise assessment of earnings differences provides a measure of income inequality that can become a valuable guide for policy analysis. Understanding its determinants can also provide insights that lead to policy actions, from the design of education policies to the structure of the tax system. The discussion on earning differences is also, fundamentally, a question about equilibrium wage functions, endowments, and the operation of the labor market. Therefore, it relates to aspects such as the distribution of production functions at the firm level, the distribution of worker skills and human capital, and the way in which the labor market allocates different types of workers across different types of firms. As a consequence, earnings differences are also associated with the efficiency of the equilibrium allocation, and how they relate to the specific characteristics of the production function.

Conceptually, equilibrium earnings differences across workers can emerge from at least three sources. On the one hand, they can reflect fundamental differences in the market valuation of workers endowments, with different types of workers systematically getting different wages. On the other hand, they can emerge from fundamental differences across firms, with certain types of employers systematically paying more than others. Finally, earnings differences can be exacerbated or dampened by the way in which workers sort across firms in equilibrium, a process that is likely to be significantly affected by the potential presence of production complementarities between firms and workers.

Over the last decade, the increasing availability of large matched employer-employee datasets has rekindled interest in this topic, with a large recent literature (Song et al. (2019), Card et al. (2018), Alvarez et al. (2018), Bonhomme et al. (2019), among many others) using different methodologies to analyze the role of firm and worker components in explaining the dispersion of earnings, and its evolution over time. Most of these papers have relied on some variation of the model proposed by Abowd, Kramarz, & Margolis (1999) (henceforth AKM), which expresses log-earnings as a linear function of worker effects, firm effects, covariates, and idiosyncratic error terms, ruling out the existence of systematic worker-firm complementarities.

This article contributes to this literature by providing new evidence on earnings dynamics and determinants for Chile, a small open economy and a middle-income country that has experienced significant economic growth over the last three decades, and where earnings inequality has been relatively large by international standards. Chile is also particularly interesting in the case of estimates of the AKM model, as some recent literature (Bonhomme et al., 2019, 2020) has argued that limited worker mobility can bias the estimation of the separate worker and firms effects of the AKM model. Chile has a very large degree of labor mobility, one of the highest among OECD countries (Albagli et al., 2017). We exploit this feature by providing a comparison between a stylized version of the AKM model and the model proposed by Bonhomme et al. (2019) (henceforth BLM), which allows for worker-firm complementarities. Interestingly, as discussed below, our results show that, even in the context of an economy with high mobility and a very large connected set, there are relevant differences in the estimated importance of firm effects and sorting patterns across both methodologies.

Our dataset is a census of tax statements of all firms that hired formal salaried workers between 2005 and 2016. Thus, it covers all formal employment in the country and has the advantage of reporting uncapped labor earnings. The dataset also contains information on the balance sheet of firms, which allows us to connect measures of inequality and firm characteristics, such as productivity.

This is also an interesting period for the Chilean economy, characterized by relatively fast wage growth, a significant increase in the share of college educated workers and female participation, and a commodity boom associated to the price of copper.

We highlight three main results.

First, the dispersion of labor earnings in Chile decreased over the sample period, a result that holds under different measures of earnings inequality. This result is specially strong for the upper end of the income distribution. Almost the entire reduction in overall labor earnings inequality over the decade is due to the decreasing dispersion of average earnings between firms, a pattern that is observed in almost every economic sector and particularly in bigger and more productive firms. Within-firm dispersion, which explains more than half of the overall dispersion in earnings, remained roughly constant throughout the period. We show there is a close relationship, at the firm level, between productivity and within-firm dispersion.

Second, estimates based on the AKM model show that systematic differences between workers explain roughly 60% of the overall dispersion in earnings. The reduction in the heterogeneity of worker effects is also by far the main driver in the reduction of inequality, with a reduction in the strength of sorting also playing a role. Firm effects, which seem to be unrelated to measures of firm productivity, play a relatively small role.

Third, a comparison of the results of a stylized version of the AKM model and the framework in BLM shows some interesting differences. While the contribution of worker heterogeneity is roughly similar across both models, sorting patterns are more important in BLM, explaining a fifth of overall dispersion. Conversely, AKM attributes a larger role to differences in firm effects. Thus, while the estimation of the (large) importance of worker effects is robust across methodologies, there are relevant differences in the importance given to sorting patterns and firm effects. As mentioned before, the fact that these differences between both methodologies persist in an economy with high worker mobility is interesting in terms of the methodological discussion between different estimation procedures for matching functions.

The structure of the paper is as follows. Section II describes the data and the procedure to construct the baseline sample, and characterizes the labor earnings distribution. Section III analyzes the relative importance of between and within-firm variance to explain aggregate dispersion. Section IV decomposes the change in labor earnings inequality into changes in firm average earnings, worker sorting, and worker segregation, estimating the AKM and BLM models and contrasting their results. Section V concludes.

II Labor Earnings of Formal Employment in Chile

II.1 Data

Our main source of data is a matched employer-employee dataset provided by the Chilean Internal Revenue Service (SII, by its Spanish acronym) between the years 2005 and 2016[1]. It covers all firms that operate in the formal sector and all formal wage employment in Chile, which adds up to nearly 60% of the total employment in the country[2]. All firms must report annually the specific months in which each worker was employed as well as the annual sum of labor earnings paid to each individual[3]. To

[1]Affidavit N. 1887 of *Servicio de Impuestos Internos*.

[2]In 2017, data from the New National Employment Survey (NENE by its Spanish acronym) conducted by the National Statistics Institute shows that the share of formal wage jobs to total employment was 57%. The remaining 43% is divided into informal wage jobs (12%), self-employed workers (22%), employers (4%), domestic services (4%), and contributing family workers (1%).

[3]Labor earnings (net of social security payments) include the base salary, incentives and rewards, payments for agreements, sales commissions, and overtime payments, so they provide a comprehensive measure of labor compensation. Thus, every month, we can identify the employment status of any given worker, and a measure of her average monthly labor earnings in that year, as well as a monthly measure of total employment and the distribution of average monthly earnings within the firm. To work with real earnings, we adjust all peso-values in the dataset to be equivalent to 2015 pesos using

preserve anonymity, each firm and worker in this administrative dataset are assigned unnamed identifiers by SII, allowing us to track workers in the formal sector across job transitions. Between 2005 and 2016, we have information for 650 thousand firms and 9.7 million workers through 30 million labor relations (90 million observations).

We match this dataset with two other sources, protecting the confidentiality of all tax information. First, using information provided by the Chilean Register Office (*Servicio de Registro Civil e Identificación* in Spanish), we can match the basic demographic characteristics (gender and date of birth) of each worker. Second, we use the definition created by National Accounts of the Central Bank of Chile to assign each firm to an economic sector, and we aggregate this information to classify firms across nine broad sectors[4]: Agriculture (Agr.); Mining (Min.); Manufacturing (Manufac.); Utilities (Ut.); Construction (Constr.); Retail, hotels, and restaurants (RHR); Transportation and communication (TC); Financial services (FS); and Other services (OS).

A relevant advantage of using a census of administrative data, vis a vis income surveys, is having a more precise depiction of the complete earnings distribution. Most previous studies that characterize inequality using Chilean data have relied on household income surveys [5], and have the drawback of potentially underestimating the upper-tail of the income distribution, as high-income households are often underrepresented in random samples and frequently under-report their income (Atkinson et al., 2011). Thus, it seems likely that evidence from surveys, like The Socioeconomic Characterization Survey (CASEN by its Spanish acronym)[6] can underestimate income dispersion. (Burkhauser et al., 2018). As a way of comparison, a simple calculation of the Gini coefficient using the SII data for 2015 shows a value 0.53, 0.1 larger than the equivalent calculation using CASEN and the New Income Supplementary (NESI by its Spanish acronym) surveys (Figure A.1)[7]. More tellingly, according to the SII data, the top 5% of workers holds 30% of total labor earnings in the formal sector, while according to the

CPI.

[4]Firms that are not classified in a specific sector, as well as the observations of workers while they remain employed in the firm are excluded from our analysis. We also exclude firms in the public sector (public administration, public health, and public education), as well as workers' observations while they are employed there, as it is likely that labor compensation decisions in those firms are not directly related to profit maximization and might not reflect market forces.

[5]See, for example, Solimano & Torche (2008), Eberhard & Engel (2009), Sapelli (2013), Parro & Reyes (2017), among others.

[6]Conducted by the Ministry of Social Development since 1985, CASEN is a cross-sectional household survey that includes information about demographics, education, employment, health, housing, and different sources of incomes. It is the most popular survey for economic studies in the country.

[7]To ensure a proper comparison, we replicate the definition of labor earnings reported in the tax statements into the different surveys, considering only formal wage employment and using monthly earnings for the case of the tax data.

CASEN and the NESI surveys the same group of individuals would only account for 15% and 10% of earnings, respectively. These results support the notion that studying inequality with administrative data can provide a more precise picture compared to household surveys.

Moreover, this dataset is also better suited to provide a quantitative assessment of inequality when compared to other administrative sources available for Chile. In particular, the Unemployment Insurance dataset (*Seguro de Cesantía* in Spanish) includes, similarly to the tax records, information on the workers' monthly earnings, the firm where they worked, and the economic sector of the firm. However, reported earnings in this dataset are truncated at the top, as the unemployment insurance legislation defines an earnings cap for mandatory contributions. Thus, all workers with actual earnings above the cap are top-coded. This restriction affects slightly more than 2% of workers in the data each month. As a result, data from the Unemployment Insurance by construction cannot correctly represent the top 2 percent. Our database, on the other hand, is uncapped at the top, allowing us to study the very top of the labor earnings distribution, which according to the literature can be an important driver of inequality (Piketty & Saez, 2003). In fact, if we apply the Unemployment Insurance earnings cap in the SII data, the calculated Gini coefficient is 0.47, 0.06 below the one using the uncapped SII data.

II.2 Baseline Sample

To construct our baseline sample we follow the basic methodology proposed by Song et al. (2019) and adapt it to our data. First, we focus on inequality among workers who are active participants of the formal labor market. Therefore, we want to exclude individuals who work sporadically or whose main source of earnings comes from self-employment or the informal sector. As a result, our baseline sample is comprised of workers aged 20 to 60 years, who - summing across all their jobs in a given year - earn at least the equivalent of the minimum wage for one quarter full-time (13 weeks or 3.25 months).

Second, as we are interested in the dispersion of earnings within-firms, we only include firms and workers employed in firms with at least 20 employees, as measures of within-firm variance for smaller firms might be too noisy.

Third, although with our dataset we can identify the specific months of employment for each worker, we choose to work with annual earnings. The main reasoning behind this methodological choice is that monthly earnings can underestimate inequality, as differences in earnings across workers can come from an intensive margin (some workers have higher monthly wages than others) or from an extensive one (some workers have longer employment spells and receive more wages) (Hoffmann & Malacrino, 2019). As an illustration, a calculation of the Gini coefficient that uses annual wages yields a value of 0.59, compared to 0.53 using monthly labor earnings[8].

Finally, the decomposition procedure requires that all workers are linked to a unique firm in any given year. Workers who hold more than one job in a year are associated to their primary job, defined as the relationship where they received their largest earning in that year. As in Song et al. (2019), earnings of primary jobs are adjusted to include the earnings accrued in other jobs, as if the individual had only worked at the main firm during that year. Thus, earnings at the firm level are the set of adjusted earnings for all workers for whom the firm was their primary job[9].

Following these criteria, we build an unbalanced panel between 2005 and 2016 of 50,589 firms and 6,192,080 employees. As discussed, the sample only includes workers with a strong attachment to the formal labor market and private sector firms with at least 20 workers. On average, we work with nearly 3 million labor relations each year. Our results are robust to different definitions of the degree of attachment to the formal labor market (for example, if we only consider workers employed at least 6 months, instead of earning at least 3.25 times the monthly minimum wage). They also look very similar if we include all industries, or if we change the thresholds for age and firm size. This suggests that the restrictions on the sample, while useful for methodological reasons and to have a direct comparison with recent literature, do not have a first-order effect on our estimates.

[8]Using data for 2015 and all workers with non-zero earnings.

[9]For more details about the construction of our sample, see Appendix C.

Table 1

Summary Statistics

Year	Number of workers	Number of firms	Log Earnings Mean	Log Earnings SD	Age Mean	Proportion of men	Firm size Mean
2005	2,276,504	18,341	15.06	0.96	36.26	0.69	1,065
2006	2,406,445	18,893	15.08	0.96	36.39	0.68	1,141
2007	2,649,363	20,174	15.09	0.96	36.47	0.67	1,194
2008	2,787,933	20,600	15.12	0.96	36.59	0.66	1,256
2009	2,663,678	19,787	15.16	0.97	36.98	0.66	1,244
2010	2,858,165	20,716	15.20	0.96	36.99	0.66	1,442
2011	3,015,632	21,498	15.25	0.96	37.05	0.65	1,459
2012	3,195,109	22,077	15.31	0.96	37.18	0.65	1,804
2013	3,310,591	22,612	15.37	0.94	37.40	0.64	1,846
2014	3,280,532	22,424	15.40	0.95	37.64	0.64	1,786
2015	3,345,747	23,264	15.42	0.94	37.88	0.63	1,850
2016	3,421,542	24,273	15.43	0.93	38.14	0.63	1,781

Note: Statistics are computed for the baseline sample. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Table 1 provides key summary statistics for formal sector workers and their firms in our baseline sample. As described in the introduction, our sample covers a period of relatively rapid wage growth. Between 2005 and 2016, mean log earnings increased 37 log points, which represents an average annual increase of 3.4%, compared to a 2.7% average growth in real GDP per capita. Also, the simple standard deviation of labor earnings peaks in 2009 and falls afterwards, ending up in 2016 4 log points below the starting point in 2005. Another notable feature is the ageing of the working population, as in a decade the average age of workers increases by almost two years. This is consistent with the joint effect of the country's demographic trends, delayed entry to the labor market due to an increase in tertiary education enrollment, and with a longer working-life for older workers[10]. The data also shows the decreasing share of male workers in total employment, in line with a higher female participation rate[11].

II.3 Main Features of the Labor Earnings Distribution

To provide an initial overview of labor market dynamics in Chile, we start by characterizing the main changes in the distribution of labor earnings between 2005 and

[10]According to the National Employment Survey data (Institute of National Statistics), the participation rates of population older than 60 years old rose from 28 to 35% between 2005 and 2016, while that of individuals younger than 25 fell from 39 to 35%.

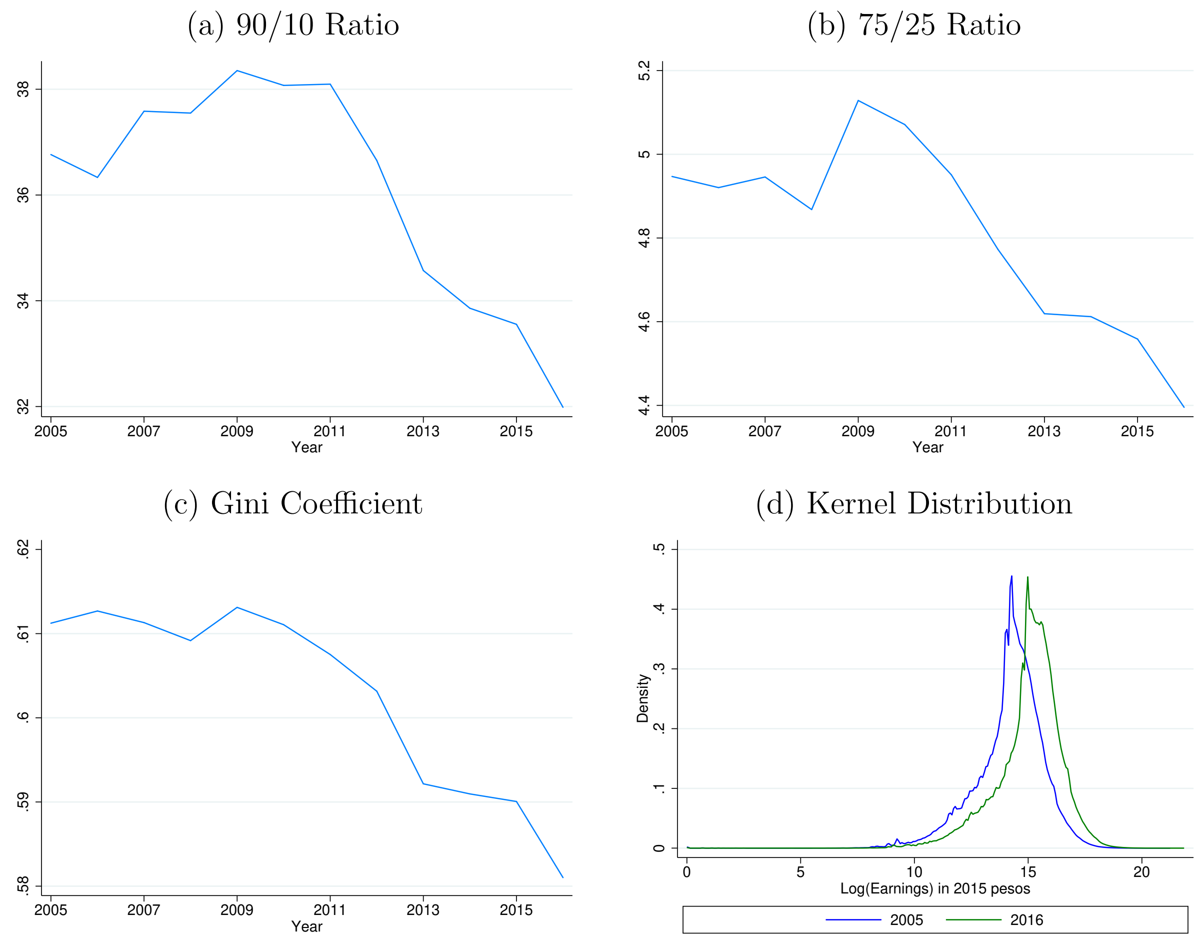
[11]Data from the National Employment Survey shows that the female participation rate increased from 43% in 2005 to 48% in 2017.

2016. To this regard, we use the complete universe of workers, before focusing in the baseline sample that will be used throughout most of our empirical analysis. We look at the evolution of four indicators: the 90-10 and the 75-25 percentile ratios, the Gini coefficient, and the Kernel distribution of labor earnings (Figure 1). The main takeaway from all these measures is a reduction in earnings dispersion over the period. Growth at the top percentiles of the earnings distribution was slower than growth at the bottom, leading to a reduction in both the 90-10 ratio (from 36 to 31) and the 75-25 ratio (from 4.9 to 4.3). Measures of the complete distribution show a mild reduction in the Gini coefficient in the second half of the sample (from 0.61 to 0.58) and a shift towards the right in the Kernel distribution[12].

[12] It seems likely that these patterns can be partially explained by the increase of 34% in the real minimum wage in these eleven years. Engbom & Moser (2017) find that the rise of 119% in the real minimum wage in Brazil between 1996 and 2012 induced a 11 log points decline in the variance of wages, with spillovers for workers up to the 80th percentile. Although the increase of the minimum wage in Chile was smaller than in Brazil, it is still potentially a relevant factor behind the reduction in inequality.

Figure 1

Labor Earnings Inequality Measures of Formal Wage Employment in Chile, All Workers

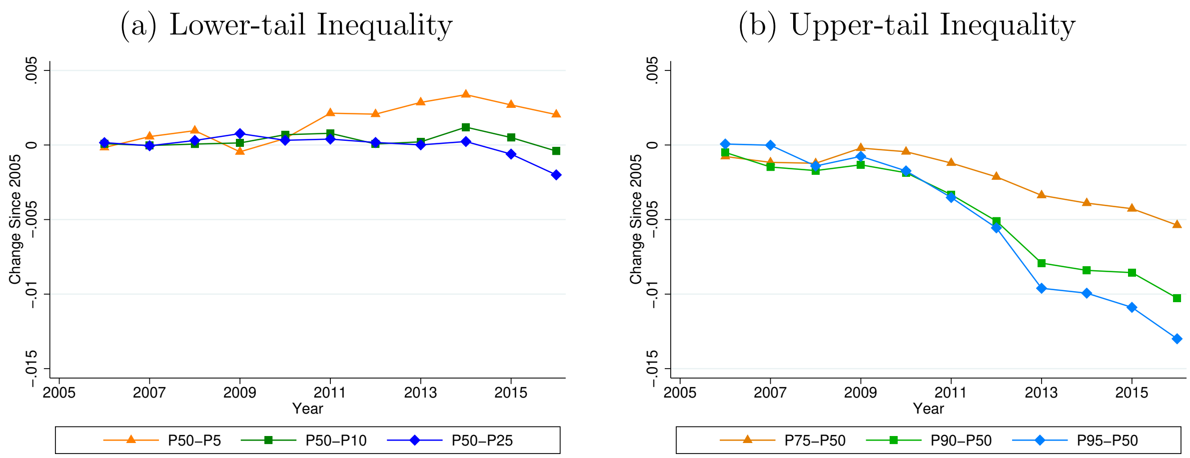


Note: This sample includes all firms and all workers with non-zero earnings.
Source: Authors' calculations based on SII data.

To provide a better characterization of earnings dynamics, and the underlying forces behind the reduction in earnings dispersion, we now, up to the end of the paper, focus on our baseline sample, which provides a cleaner comparison between workers with similar attachment to the labor market. We begin by following Alvarez et al. (2018) and plot the variation in different log percentile earning ratios with respect to 2005. Lower-tail inequality is measured by the ratio of wages at the middle of the earnings distribution relative to those that are near the bottom (i.e., the 50th percentile relative to the 5th, 10th, and 25th percentiles), while upper-tail inequality is characterized by the ratio of wages near the top of the distribution to those at the middle of the earnings distribution (i.e., the 75th, 90th, and 95th percentiles relative to the 50th percentile).

Figure 2

Log Percentile Ratios of the Earnings Distribution in Chile

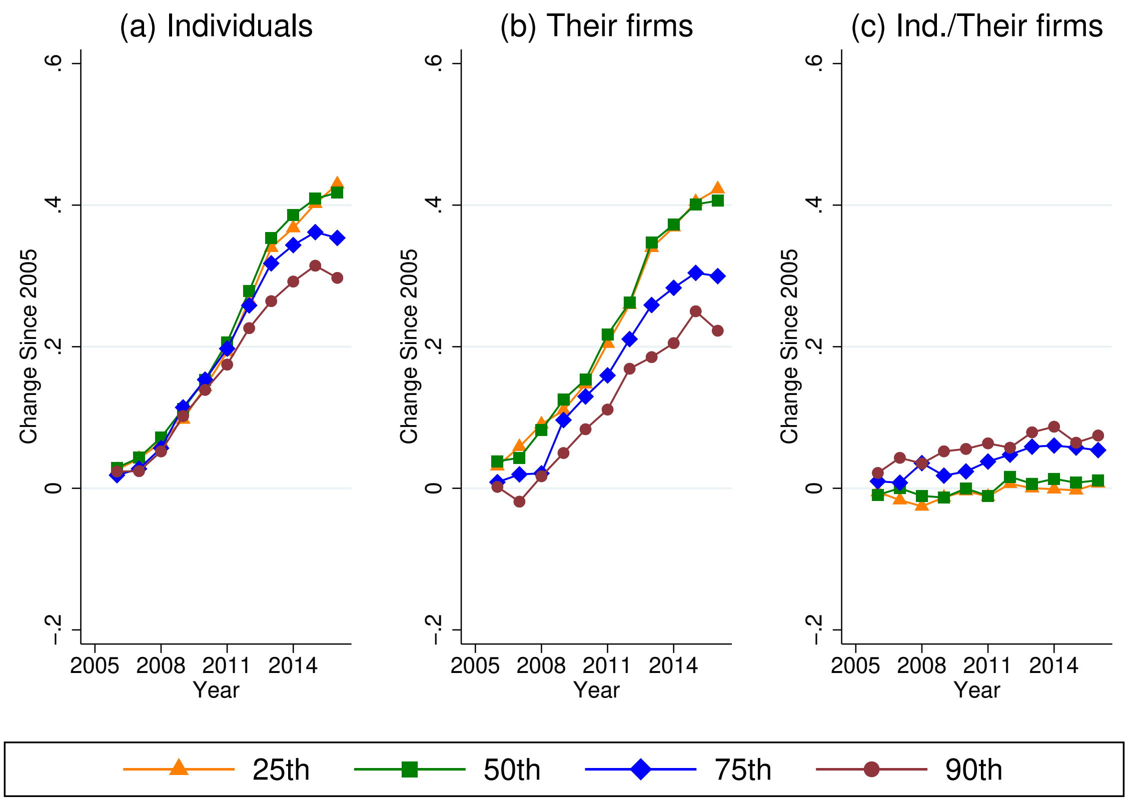


Note: Statistics are computed for the baseline sample. The ratio is normalized to zero in 2005. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Figure 2 reveals that the reduction in upper-tail inequality was strongest for the top of the distribution. While the 95-50 ratio fell 1.3 log points between 2005 and 2016, the 90-50 ratio and the 75-50 experienced a reduction of 1.0 and 0.5 log points, respectively. In contrast, lower-tail inequality actually increased slightly, especially in the bottom part of the distribution (the 50-5 ratio increased by 0.2 log points). This differs from the results in Alvarez et al. (2018), which show that lower-tail inequality fell sharply in Brazil between 1996 and 2012, while upper-tail dispersion remained constant.

Figure 3

Change in Selected Percentiles of Annual Labor Earnings Relative to 2005



Note: Statistics are computed for the baseline sample. See text for details.
Source: Authors' calculations based on the SII and Register Office data.

The inequality reduction in the upper-tail of the earnings distribution can decomposed as the result of two mechanisms. First, it could reflect that average wage growth in firms that had a larger share of high-earnings workers was slower, a force that would reduce the dispersion of wages between-firms. Conversely, it could reflect a reduction in dispersion within-firms: the earnings of workers in the upper percentiles grew less than the firm average, therefore compressing the wage distribution inside the firm. These two forces can be easily illustrated by two extreme alternatives. On one end, if firms are heterogeneous, and each type of firm completely specialized in hiring workers from a specific earnings percentile, there would be no within-firm dispersion, and all differences in earnings would be differences in average earnings across firms. On the other end, if all firms were completely identical, they would all have the same composition of workers, which would be representative of the whole economy. In that world, average earnings across all firms would be the same, and dispersion in the economy would only come from within-firm differences.

To account for these two mechanisms, Panel A of Figure 3 shows the growth of individual labor earnings for four selected earnings percentiles - 90th, 75th, 50th, and 25th- normalized by their level in 2005. Panel B plots the component associated to changes between-firms, measured as the increase - relative to their 2005 values- of average labor earnings (across workers) in firms that hire employees of that specific percentile. Panel C shows the within-firm component, which by construction can be calculated as the difference between individual earnings growth (Panel A) and average firm growth (Panel B) at each percentile. Therefore, it captures the relative performance of workers in each percentile with respect to their co-workers.

Panel A confirms that during this period the increase in annual earnings for the lower-tail exceeded that of the top percentiles, particularly with respect to the 90th percentile. However, Panel C shows that this was not driven by an overall reduction in within-firm dispersion: while earnings growth for workers in the lower percentiles was similar to the average growth of their firms, the worker-coworker gap actually increased for individuals in the upper-tail[13]. Therefore, the reduction in inequality was in fact driven by the evolution of the firms' average earnings. As Panel B shows, average earnings in firms that hired workers in the lower-tail increased more rapidly.

Figure A.2 in the appendix decomposes Panel A of the last exercise into four groups by the workers' age: 20 to 29, 30 to 39, 40 to 49, and 50 to 60 years old. For the three younger groups, the growth rate of earnings has been similar across earning groups. Dispersion, however, has fallen within the older group. This somehow surprising result suggests that the reduction in upper-tail inequality is driven mostly by workers in the latter part of their work life.

Following Song et al. (2019), this decomposition can be extended to the complete earnings distribution (Figure 4). Analogously to Panel A in Figure 3, the line in blue plots, for each percentile, the variation in log earnings between 2005 and 2016. By construction, that variation can be decomposed between the red line, the increase in average labor earnings in firms hiring each percentile, and the green line, the relative individual growth with respect to the corresponding firm average (thus, the average across all percentiles is zero).

Dispersion across a given set of percentiles increases when the blue line slopes upward (workers at the top of the distribution experience higher earnings growth than

[13]From 2005 to 2016, the worker-coworker gap for individuals in the 25th percentile marginally increased by 0.5 log points. In contrast, the worker-coworker gap for workers in the upper percentiles increased by 7 log points by the end of the period.

those at the bottom) and decreases when it has a negative slope. A flat region - as seen between percentiles 21st and 46th - means that relative earnings for those percentiles remained constant during the sample period. Therefore, dispersion in the lower-tail of earnings (below the 20th percentile) increased, while it fell in the upper-half (above the 47th percentile).

The decomposition shows that the reduction in dispersion between percentiles 47 and 97 was driven by the fact that workers in the upper percentiles were employed at firms where average wages grew less, even though their relative wages within the firm increased. This differs from the case of workers at the very top of the distribution: above the 98th percentile, the wages of these workers grew less than the average wages at their firms (6 log points more in average).

Below the 20th percentile, the rise in lower-tail inequality is explained by the within-firm component: the wages of workers at the lower end lagged behind the average earnings paid by the firms in which they were employed. Hence, the increase in lower-tail dispersion is explained by changes within-firms.

Figure 4

Change in Inequality of Annual Earnings
across Percentiles from 2005 to 2016

Note: Statistics are computed for the baseline sample. See text for details.
Source: Authors' calculations based on SII and Register Office data.

These results show that the evolution of between-firm dispersion plays a big role in explaining the dynamics of wage inequality in Chile between 2005 and 2016. The results found for Brazil, between 1996 and 2012, are in the same line: the significant reduction in inequality is almost entirely explained by the growth of the average wage paid by firms (Alvarez et al., 2018). In the United States, Barth et al. (2016) find that the increase in inequality is due to the growth of the average salary paid by firms, since the red and blue lines are practically the same.

In the appendix, we repeat this last exercise for different age and gender subsamples. The patterns across groups are similar to the aggregate: an increase in lower-tail inequality that comes from those workers lagging behind their coworkers, an intermediate region with no changes in dispersion, a reduction in upper-tail inequality driven by the evolution of average firm earnings, and, last, a drop in earnings dispersion due to within-firm changes for the very top of the earnings distribution.

Despite these general trends, analyzing these subgroups separately uncovers differences between workers. We find that the decline in upper-tail inequality due to between-firm changes is more frequent for women than for men (Figure A.3), and for older workers than for younger individuals (Figure A.4). The opposite is true for the rise in lower-tail variance attributable to within-firm variations.

The results in this section highlight how the evolution of earnings across income percentiles can be understood as the result of average changes occurring at the firm level and relative earnings movements inside the firms. This same logic is applied in the next section to decompose the overall dispersion of earnings in the economy into between-firms and within-firms components. Even though these accounting exercises cannot identify the fundamental economic forces that underlie these trends, they provide valuable insights into the equilibrium outcomes resulting from those forces.

III Labor Earnings Inequality Within and Between-Firms

Extending the argument presented in the previous section, wage dispersion in the economy can be understood as the combination of differences in average wages across firms and the dispersion of wages inside firms. To compare the relative importance of these forces, we follow the methodology proposed by Fortin et al. (2011); Alvarez et al. (2018); Song et al. (2019), among others. The log earnings of worker i employed by firm j in period t ($y_t^{i,j}$) can be decomposed into the average earnings for firm j ($\bar{y}_t^j$) and the difference between the individual's earnings and the firm average:

$$y_t^{i,j} = \bar{y}_t^j + [y_t^{i,j} - \bar{y}_t^j] \tag{1}$$

As detailed in Appendix D, the decomposition at the individual level can be extended to break down overall inequality into the dispersion of average earnings across firms and a within-firm component, calculated as a weighted mean of the earnings dispersion within each firm, weighted by the firmÂ´s share of total employment:

$$var(y_t^{i,j}) = \underbrace{var_j(\bar{y}_t^j)}_{\text{Between-firm dispersion}} + \underbrace{\sum_{j=1}^{J} \omega_j^t \times var_i(y_t^{i,j} | i \in j)}_{\text{Within-firm-}j\text{ dispersion}} \tag{2}$$

With a similar intuition to the one for earnings growth, Alvarez et al. (2018) pose two extremes to explain this decomposition: one case where all firms pay the same wage on average and, therefore, all income dispersion comes from differences across workers within firms; and another where every firm pays the same wage to all their workers, so the dispersion of income is only due to differences in wages across firms.

Figure 5 provides an overview of the time series of the overall variance from 2005 to 2016 and its two components in equation (2) over our baseline sample. The first standout result from the decomposition is that the within-firm component is the main determinant of aggregate dispersion - explaining on average 54% of the variance of log earnings-, although the between-firm component is also very important[14]. Thus, the overall dispersion in earnings in Chile's formal labor market comes first from the distribution of earnings within firms - likely related to the structure of production tasks and the composition of the firm's workforce - and secondly from average differences between firms, emerging both from firm characteristics and the way different types of workers sort more intensively into certain types of firms. The fact that the relative weight of both components in explaining overall dispersion is relatively similar differs from the evidence from the United States (Song et al., 2019), where the dispersion within firms explains a much larger share of the overall variance than the dispersion between firms.

The second notable result is that the 9.7 log points reduction in the cross-section variance of earnings between 2005 and 2016 comes entirely from a reduction in the

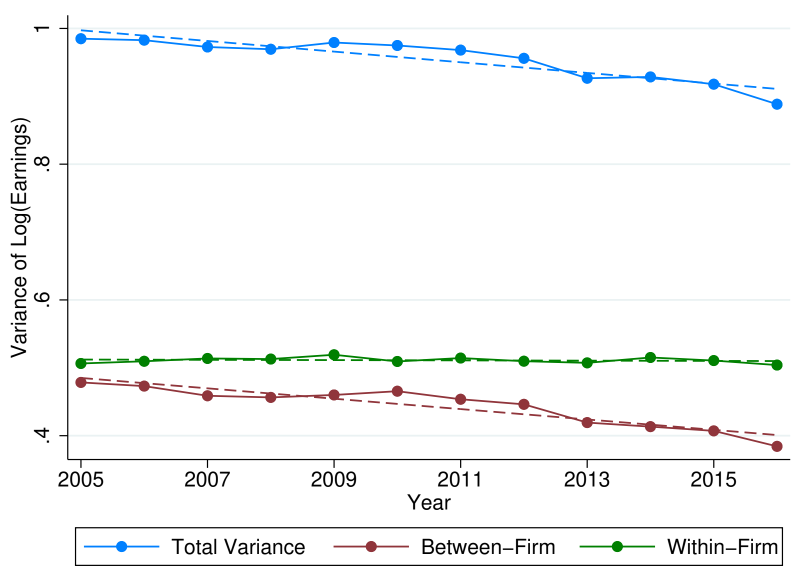
[14]The main qualitative results in Figure 5 are robust to different definitions of labor market participation. Figure A.5 shows the decomposition exercise for individuals who work for at least 3 or 6 months in a given year (instead of earning at least the equivalent of 3.25 times the monthly minimum wage cut-off of our baseline sample): in both cases, the within-firm component explains more than half of the overall variance and almost the entire reduction in total inequality over 2005 to 2016 is due to the decreasing dispersion of average earnings between firms.

between-firms component, as the within-firm dispersion remained roughly constant, similarly to the evidence for the United States and Brazil (Alvarez et al., 2018). Thus, while the composition of wages inside firms did not change, differences in average wages across firms were significantly dampened during the decade.

Naturally, these patterns in inequality and its components can be driven by multiple margins, from changes in the characteristics of the active sets of firms and workers, to changes in the allocation of workers across firms. This is, the extensive margins of firm turnover - defined as exit/entry dynamics that change the composition of firms- and worker turnover - defined as workers that are changing jobs- could explain a relevant share of the previous results. Table B.1 addresses this concern and compares the results of the baseline sample with two alternative exercises based on Barth et al. (2016): one that only considers workers employed in the same firm for two consecutive years and another which only allows for firms that are active during two consecutive years. The variance decomposition as well as the evolution of the components of dispersion is very similar across samples, suggesting that these adjustment margins do not play a large role in explaining the general results[15].

Figure 5

Variance of Log Earnings Between and Within Firms



Note: Statistics are computed for the baseline sample. See text for details.
Source: Authors' calculations based on SII and Register Office data.

[15]As could be expected, and consistent with the results in Barth et al. (2016), the dispersion of earnings among job keepers is smaller than across the whole sample of workers (0.90 compared to 0.95).

Likewise, the aggregate results might be driven by the behavior of a particular set of firms or hide a significant degree of heterogeneity. For instance, the reduction in between-firms dispersion might be associated to sectoral patterns related, for example, to the large changes in commodity prices observed throughout this period. Similarly, firms with specific characteristics, as size or productivity, might have been affected more intensely by factors that can lead to changes in employment and compensation decisions, such as the availability of new production technologies or exposure to international trade. To address this issue, we compare the relative importance of between and within-firm dispersion for different subsamples based on the size, productivity, and industry of the firms.

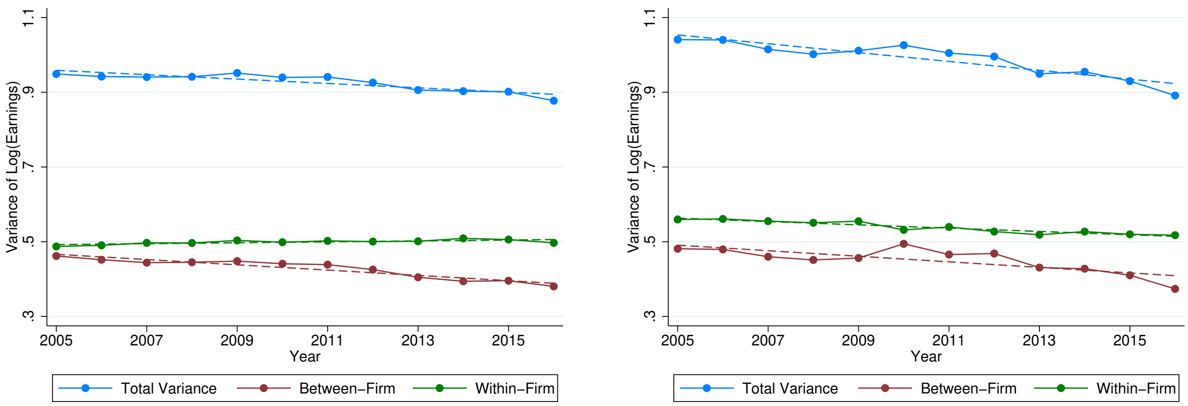
First, we separate the sample between small to medium-sized firms (20 to 999 workers) and big firms (those with more than 1,000 employees)[16]. Figure 6 indicates that there is a significant difference in the dynamics of dispersion between both groups, as the reduction in earnings variance between 2005 and 2016 for workers in big firms doubles the reduction for smaller firms (15 versus 7 log points, respectively). This is the joint result of a larger reduction in between-firm variance across big firms (10 versus 8 log points, respectively) and by the fact that within-firm inequality actually increased by 1 log point for small and medium-sized firms, whereas it fell by 4 log points for larger firms. In contrast, in Brazil the reduction of earnings inequality in small firms between 1996 and 2012 was mainly due to a fall in the between-firm component (Alvarez et al., 2018).

[16]Big firms account for slightly less than 2% of firms but 32% of employment, on average.

Figure 6

Variance of Log Earnings Between and Within Firms by Firm Size

(a) Firms with 20 to 999 employees (b) Firms with 1,000+ employees



Note: Statistics are computed for the baseline sample. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Second, we distinguish between low and high-productivity firms[17], as we can match our employer-employee dataset with information on the firms' balance sheets reported by the Annual Affidavit on Income[18]. Figure 7 shows that the variance of earnings is lower and more stable within the group of low-productivity firms. The decomposition patterns observed in the overall sample no longer hold inside each productivity group, as the relative importance of the between component becomes much smaller. This reflects that a significant share of the between-firms dispersion on the aggregate sample comes from differences in average wages between low and high productivity firms. The reduction in earnings dispersion for workers in high productivity firms is solely driven by a reduction in the differences in average earnings between firms.

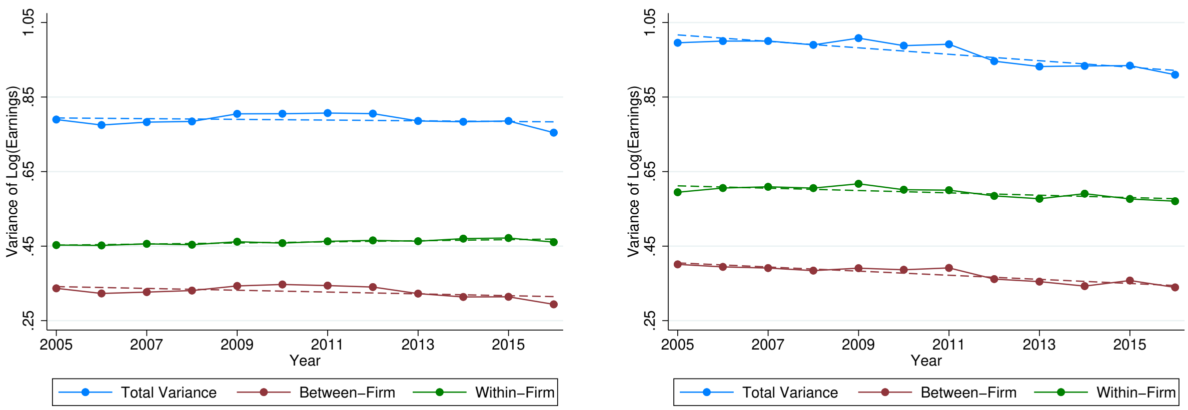
[17]Low-(high-) productivity firms have a demeaned revenue per full-time equivalent worker below (above) its sector' median demeaned productivity.

[18]Firms that have to pay annual revenue taxes are mandated to file this tax form (Form N.22). After merging the two datasets, we lose 21% of firms and 24% of firm/year observations, as some firms filing Affidavit N.1887 are not required to report the Annual Affidavit on Income, and others do not provide complete information on all variables needed to calculate productivity. Nevertheless, the decomposition into between and within-firm components using this new aggregate sample is very similar to the one with the whole baseline sample.

Figure 7

Variance of Log Earnings Between and Within Firms by Labor Productivity

(a) Low-Productivity Firms (b) High-Productivity Firms



Note: Statistics are computed for firms that are included in the baseline sample and report the Annual Affidavit on Income. Labor productivity is defined as revenue per full-time equivalent worker. Low (high) productivity firms have a demeaned labor productivity below (above) the sector median demeaned productivity. See text for details.
Source: Authors' calculations based on SII and Register Office data.

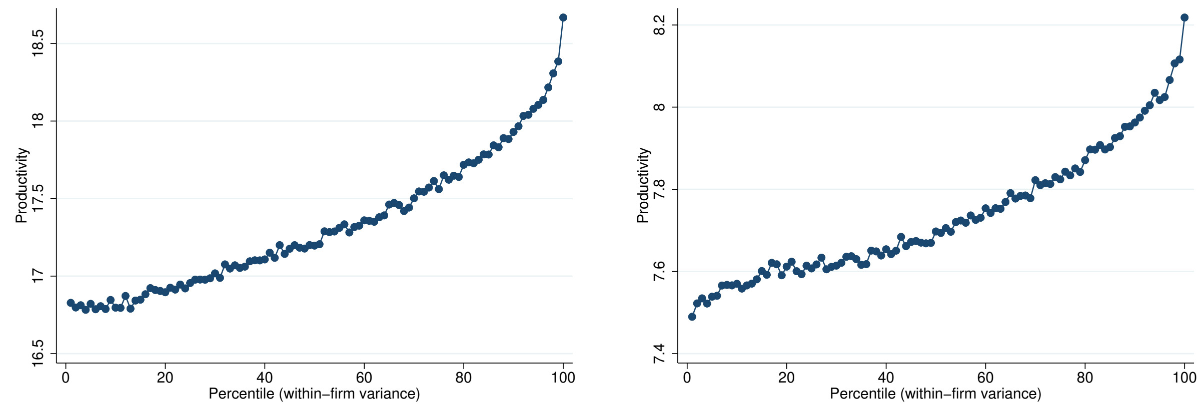
An interesting related exercise is to analyze how within-firm dispersion, which explains more than half of overall inequality, relates to firm productivity. Figure 8 plots the average productivity across firms for each percentile of within-firm variance. We use both labor productivity, defined as revenue per full-time equivalent worker, and the productivity measure used in Ackerberg, Caves, & Frazer (2015) (henceforth ACF), calculated as a function of intermediate costs, capital, and the number of full-time equivalent workers.

There is a very strong positive correlation between within-firm inequality and productivity, which is robust to the two alternative specifications for productivity. This finding seems consistent with the notion that more productive firms might also be more complex organizations, with a larger set of tasks associated with their production function (Becker & Murphy, 1994; Chaney & Ossa, 2013). Under this interpretation, wage dispersion can be the equilibrium result of hiring a broader set of worker types, which can be relate to the concept of occupation layers as in Caliendo et al. (2015).

Figure 8

Average Productivity and Within-Firm Variance

(a) Labor Productivity (b) ACF



Note: Statistics are computed for firms that are included in the baseline sample and report the Annual Affidavit on Income. Labor productivity is defined as revenue per full-time equivalent worker. The second measure of productivity is calculated according to ACF and defines labor as number of full-time equivalent workers. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Third, we characterize the dispersion of earning across different economic sectors. Figure 9 plots the time evolution of the variance decomposition for workers in each sector between 2005 and 2016. Sectors with high average wages (such as Utilities, Financial Services, and Mining) also have a large earnings dispersion [19] The figure also shows that sectors with high earnings variance generally have a larger within-firm component, expect for Utilities. These results are likely to be related to differences in the complexity of the production structure within each sector, as well as the span of occupations and tasks associated to them.

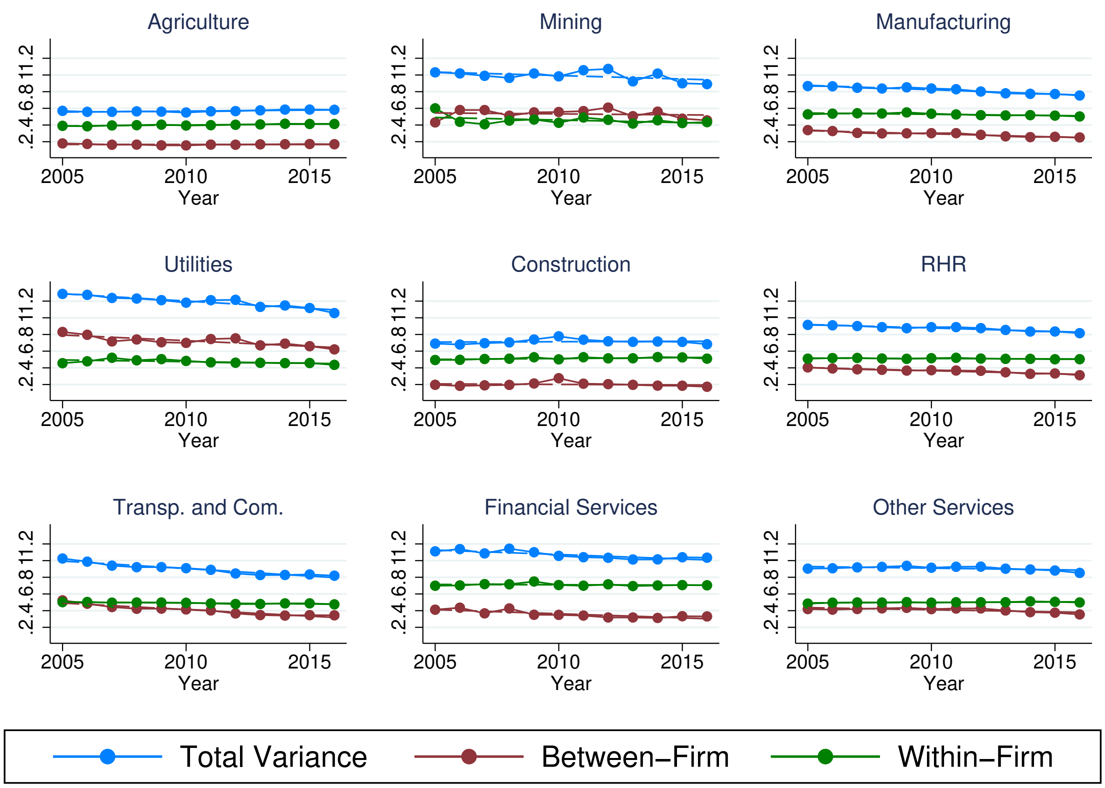
In terms of dynamics, labor earnings variance fell across all sectors except for Agriculture, with Utilities, Transportation and Mining experiencing the largest reductions. On average, the drop in earnings was larger in those sectors with large initial variance. The reduction in the between-firm component observed in the aggregate data also holds for almost all sectors, with the exception of Mining, a sector that represents a relatively small share of total employment. Thus, the aggregate reduction in inequality driven by

[19] Figure A.6 shows the average distribution of employment across sectors for workers of different earnings percentiles. Agriculture and Construction are on average low-paying sectors, as their share of employment decreases significantly for workers in the higher earnings percentiles. On the other side, the share of employment in Mining, Financial Services and Utilities is larger for higher income percentiles, though very small in terms of total employment. The predominance of the financial sector in the upper-tail of the income distribution has also been documented for developed countries (Kaplan & Rauh, 2009; Denk, 2015),

the between-firm component is not associated to the behavior of a particular sector, or-at least not solely- by the reduction in differences between firms operating in different sectors, but represents a general trend. On the other hand, the aggregate stability in the within-firm component hides significant sectoral heterogeneity. Within-firm inequality increased during the period in four sectors that represent 52% of total employment (Agriculture; Construction; Other Services; and Financial Services), while it decreased in the five remaining sectors. For the case of Brazil, Alvarez et al. (2018) show that all sectors, including Agriculture, experienced a drop in inequality due to a reduction in the dispersion of average earnings paid by firms. In addition, they also find a greater proportion of within-firm variance in Agriculture and Construction, as we do.

Figure 9

Variance of Log Earnings Between and Within Firms by Economic Sector



Note: The following is the list of acronyms and abbreviations used: Retail, hotels, and restaurants (RHR) and Transportation and communication (Transp. and Com.). Statistics are computed for the baseline sample. See text for further details.
Source: Authors' calculations based on SII and Register Office data.

IV Labor Earnings Inequality: Econometric Approach

The previous sections highlighted the role of earnings differences between and within firms in explaining the dispersion of earnings across the Chilean formal labor market, and its evolution across time. While very informative, this approach cannot account for the separate role of both firms and worker characteristics in the distribution of equilibrium earnings, or for how those earnings relate to the specific ways in which workers are sorted across firms. As discussed in the introduction, that requires estimating an earnings function for firm-worker matches, and making assumptions on the characteristics of the production process associated to the match, in order to disentangle the contributions of both parties. We present two approaches to the estimation of the matching process. First, the widely used approach by Card, Heining, & Kline (2013) (henceforth CHK) to implement the AKM model developed by Abowd, Kramarz, & Margolis (1999) and solved by Abowd et al. (2002). Secondly, we present an alternative approach developed by Bonhomme et al. (2019) to account for potential complementarities between workers and firms.

IV.1 Main features of the AKM model

By its elegance and relative simplicity, the AKM model has become a workhorse for addressing multiple topics related to labor market outcomes, including the most recent literature on inequality. The model assumes that the production technology for a given worker-firm match, in terms of equilibrium log earnings, is additive in worker and firm characteristics, ruling out potential firm-worker complementarieties. Following CHK, the log of real earnings for a given worker-firm match can be expressed as:

$$\ln(w_{it}) = \alpha_i + \psi_{\mathbf{J}(i,t)} + x_{it}'\beta + \varepsilon_{it} \tag{3}$$

where α_i corresponds to a worker-fixed effect, $\psi_{\mathbf{J}(i,t)}$ to a firm-fixed effect, $x_{it}'\beta$ is a measure of time-varing observable characteristics, and ε_{it} is an error component that will include a match-specific component. In our data, since we do not have information on the worker's formal education, our set of time-varing observable characteristics only includes a polynomial of age[20] and year-fixed effects. To study how worker and firm

[20]Given the collinearity between year and age, we normalize age as $\frac{age-40}{40}$, as in Card et al. (2018) and Song et al. (2019). This allows us to identify worker effects and time-varing observable characteristics separately.

effects have changed over the decade, we follow CHK and split our sample into three 4-years intervals: 2005-2008, 2009-2012, and 2013-2016[21].

Conceptually, the "worker effect" α_i captures time-invariant unobserved characteristics of the individual. These include the returns to human capital associated to formal education or acquired through previous labor market experience and learning by doing. In practice, it will also reflect the average additional gains obtained throughout the estimation period, to the extent that they are not fully captured by the time-varying components. It also accounts for any innate talents and skills that are relevant to the earnings-generation process. These are all market returns, in the sense that they capture the individual's ability to systematically generate earnings when employed, regardless of the specific match. The "firm effect" $\psi_{\mathbf{J}(i,t)}$ measures time-invariant earnings differences associated with working in a particular firm. This effect, as explained by CHK, captures a firm-specific pay premium (or discount) that can be related to systematic firm characteristics such as productivity, but also to the permanent component of specific features as the rent-sharing process the firm uses or its reliance on efficiency wages. Additionally, the specification assumes that the error component has mean zero and can be split in three terms: a match-specific effect, which by assumption is independent of the firm-worker effects, so systematic complementarities are ruled out; a unit root component, and a transitory error.

Separate identification of the worker and firm effects requires job transitions of workers across firms: in a world with zero mobility, worker and firm effects are indistinguishable. As explained by CHK, this implies that these effects can only be identified within a "connected set" of firms, which are linked by the job transitions of workers moving among them. To maximize the number of worker/year observations, we consider the largest connected set in the data as our baseline sample. As discussed in more detail below, the large job turnover on the Chilean labor market works in our favor, as almost all workers in the initial dataset are linked, and therefore part of the largest connected set.

As in CHK, the estimates from equation (3) can be used to decompose the variance of log earnings as:

$$\begin{aligned} Var(\ln(w_{it})) = {} & Var(\alpha_i) + Var(\psi_{\mathbf{J}(i,t)}) + Var(x'_{it}\beta) \\ & + 2Cov(\alpha_i, \psi_{\mathbf{J}(i,t)}) + 2Cov(\alpha_i, x'_{it}\beta) \\ & + 2Cov(\psi_{\mathbf{J}(i,t)}, x'_{it}\beta) + Var(\varepsilon_{it}) \end{aligned} \tag{4}$$

[21]CHK consider overlapping intervals. We choose to follow Song et al. (2019) and work with adjacent year intervals.

IV.2 Results of the AKM Model

We estimate equation (3) for the three 4-years adjacent intervals: 2005-2008, 2009-2012, and 2013-2016. As discussed, identifying worker and firm effects separately requires a "connected set" of firms, which are related through job transitions. In order to define our largest "connected set", we follow the algorithm proposed by Abowd et al. (2002). To allow a better comparison with the previous literature, we put an additional restriction on our sample and analyze men between 20 and 60 years old.

Table 2

Summary Statistics for Overall Sample
and Individuals in Largest Connected Set

	2005-08		**2009-12**		**2013-16**	
	All	Set	All	Set	All	Set
N. of worker/year obs.	6,829,409	6,822,363	7,685,455	7,679,304	8,461,632	8,456,136
N. of workers	2,429,097	2,425,626	2,704,550	2,701,564	2,914,281	2,911,360
Log annual earnings						
Mean	15.175	15.175	15.337	15.337	15.516	15.516
Median	15.079	15.079	15.258	15.258	15.467	15.467
Std. Dev.	0.968	0.969	0.972	0.972	0.948	0.948

Note: Statistics are computed for men in the baseline sample. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Table 2 shows, for each interval, the number of worker-year observations, the number of individuals, and the mean, median, and standard deviation of log labor earnings. The combination of all restrictions leads to a sample of 6.8-8.5 million individual-year observations on labor earnings for about 2.4 - 2.9 million individual workers for each interval. Similarly to Table 1 with the broader sample, the standard deviation of labor earnings rises slightly between the first two periods and then falls between the second and the third. Overall, we observe a fall of 2 log points in the standard deviation of labor earnings between the first and last 4-years intervals. Mean log earnings rise rapidly throughout the period, at an average annual rate of 3.4%.

Table 2 also contrasts the summary statistics obtained from the set of all observations in the sample and those that are part of largest connected set. The descriptive statistics are identical in both samples, as the connected set includes 99.9% of workers in all intervals. As mentioned earlier, this reflects the very large turnover observed in Chile, one of the largest in the OECD (Albagli et al., 2017), which implies that almost all workers in the labor market are linked. This is an advantage of our dataset, which contrasts favourably with previous papers in which the connected set is relatively smaller, and therefore a larger share of workers is excluded from the analysis.

Table 3

Estimation Results for the AKM model, Fit by 4-years Period

	2005-08	**2009-12**	**2013-16**
Sample summary statistics			
Number of worker effects	1,912,043	2,152,920	2,353,606
Number of firm effects	28,106	29,288	32,059
Sample size	6,308,491	7,130,353	7,898,046
SD $\ln(w_{it})$	0.954	0.957	0.929
Summary of AKM parameter estimates			
SD (α_i)	0.740	0.735	0.707
SD $(\psi_{\mathbf{J}(i,t)})$	0.274	0.279	0.286
SD $(x'_{it}\beta)$	0.222	0.247	0.198
Corr$(\alpha_i, \psi_{\mathbf{J}(i,t)})$	0.300	0.296	0.269
Corr$(\alpha_i, x'_{it}\beta)$	0.046	0.019	0.087
Corr$(\psi_{\mathbf{J}(i,t)}, x'_{it}\beta)$	0.086	0.093	0.106
RMSE (residual)	0.363	0.371	0.374
Adj R^2	0.855	0.850	0.838
Comparison match model			
RMSE (match residual)	0.316	0.319	0.326
Adj R^2	0.884	0.882	0.868
SD (match effect)	0.079	0.076	0.074

Note: Statistics are computed for men in the largest connected set. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Table 3 summarizes the estimation results of the AKM model for Chile in each of the three 4-years intervals. A first standout result is that the standard deviation of the firm fixed effects is less than half of the standard deviation of the worker fixed effects: thus, the main driver of earnings dispersion comes from differences between workers, rather than from differences across firms. This result is similar to the one obtained by CHK, and highlights the importance of worker level heterogeneity (and its relationship with human capital) in explaining the distribution of wages in an economy. Also, in line with the previous literature (Alvarez et al., 2018; Card et al., 2018), there is a positive correlation between firm and worker effects: on average, workers with stronger market skills are more likely to work in high-paying firms. This result suggests that there is positive assortative matching in the equilibrium market allocation, which could signal complementarities in production, especially if high-paying firms are also those that are more productive.

Second, the dispersion of worker and firm effects move in opposite directions over time: workers became less heterogeneous throughout the decade, while differences across firms increased, although these changes are modest given the relatively short period under analysis. Sorting patterns also became weaker, as the correlation between firm and worker effects decreased. The residual standard deviation of labor earnings (measured by the RMSE) is small and rises slightly over the years.

Following CHK, Table 3 also contrasts the results of the AKM model - which by construction rules out systematic complementarities between worker and firm attributes- and an estimation with unrestricted match effects, which includes separate dummies for each worker-firm combination. The match model has a better fit relative to the AKM models (3% higher adjusted R^2), thus revealing that a systematic match component has a role in explaining labor earnings, although the fit of the AKM model is already quite large. However, both the standard deviation of the match effects and the adjusted R^2 of the AKM model fall - slightly - over time. This does not support the notion that assuming additivity in the earnings function has lead to significant specification errors, as otherwise the decline in the fit of the AKM model would have gone hand in hand with an increase in the standard deviation of the match effects.

In the appendix, we implement two additional robustness exercises suggested by CHK as checks for the presence of job-match effects in labor earnings. The first exercise searches for large mean residuals, whose existence would violate the assumption of no-complementarities between worker and firm attributes, as they would point out to systematic earning effects for particular match types. To do this, CHK plot the average estimated mean residuals, grouping the results by deciles of worker and firm effects (resulting in 100 worker -firm decile combination cells). Figure A.7 presents this exercise for 2009-12. Mean residuals are small, lower than 1% for most cells, suggesting that there are no systematic components in earnings specific to those particular combinations. The sole relevant exception are matches between the first decile of individuals working and the first decile of firms, which have a large mean residual that suggest a systematic match component not captured by the structure of the model. However, as this cell has only a very small share of total observations, it seems likely that its effect on the overall estimation is small.

The second exercise is an event-study analysis for the earnings effects of job transitions, classifying the origin and destination positions by quartiles of estimated firm effects, implicitly defining a job ladder for worker transitions. Figure A.8 shows that earnings gains and losses are roughly symmetric for workers who move up or down the job ladder: labor earnings gains for those switching from a first quartile firm to a fourth quartile firm are very similar to the losses for those who move in the opposite direction. Workers who switch to a firm of a different quartile experience large labor earnings changes, whereas people staying in the same quartile face small changes. This result is consistent with the idea that, on average, transitions are not driven by idiosyncratic job-match effects. However, it does not necessarily discard the existence of systematic worker-firm complementarities, as the composition of worker types involved in those transitions can change along the job ladder. There is also little evidence of transitory earnings shocks in the year before or after a job transition. These results suggest that the AKM model provides a good first-order approximation to the labor earnings determination process.

Next, we use equation (4) to decompose the variance of log labor earnings using the parameters estimated with the AKM model in Table 3. This provides a more detailed characterization of the determinants of earnings inequality that complements the analysis in the previous section. Table 4 shows the resulting decomposition using the same three 4-years intervals. First, we find that the dispersion of workers characteristics explains almost 60% of the dispersion of earnings: this is, more than half of earnings inequality in the formal market does not depend on firms or the specific allocations of workers, but directly from the differences in the market valuation of workers attributes. This result is slightly larger than the evidence for the United States (Song et al., 2019), Germany (Card et al., 2013), and Brazil (Alvarez et al., 2018), where worker heterogeneity accounts for 50 to 56% of the total variance. Similarly to the United States, firm heterogeneity plays a small direct role, explaining only 10% of total inequality. This figure is significantly smaller than the results for Germany and Brazil, where 20% of the total variance comes from firm heterogeneity.

In any case, one must keep in mind that the distribution of the market valuation of workers attributes is in itself an equilibrium outcome, which depends on the specific characteristics of the Chilean economy. Thus, the distribution of worker effects is the result both from distribution of the skills and talents across workers and the specific way in which a given labor market values those endowments. This distinction is especially relevant for cross-country comparisons, as the distribution of the market valuation of a given set of workers would probably vary across different countries.

Table 4

Decomposition of the Decline in Labor Earnings Inequality, for Men

	2005-08	**Share**	**2009-12**	**Share**	**2013-16**	**Share**	**Change**	**Share**
Var$(\ln w_{it})$	0.910		0.915		0.864		-0.046	
Components of variance								
Var(WFE)	0.547	60	0.540	59	0.499	58	-0.048	-103
Var(FFE)	0.075	8	0.078	8	0.082	9	0.007	15
Var$(Covariates)$	0.049	5	0.061	7	0.039	5	-0.010	-21
Var$(Residuals)$	0.091	10	0.095	10	0.098	11	0.006	14
2Cov(WFE, FFE)	0.121	13	0.121	13	0.109	13	-0.013	-27
2Cov$(WFE, Covar.)$	0.015	2	0.007	1	0.024	3	0.009	20
2Cov$(FFE, Covar.)$	0.010	1	0.013	1	0.012	1	0.002	3
Sum of firm components								
Cov$(\ln w_{it}, FFE)$	0.141	16	0.145	16	0.142	16	0.001	3

Note: Statistics are computed for men in the largest connected set. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Consistent with our previous results, we find a decrease of labor earnings dispersion of 4.6 log points between the first and the last time interval. This is accompanied by a similar reduction in the dispersion of the worker effects, while the variance of the firm attributes increases. Hence, we can infer that the reduction of the differences in the market valuation of workers was the main driver of the decline of inequality over the sample period. This is consistent both with a change in the distribution of the intrinsic attributes of workers and with a change in the relative return of those attributes[22]. The data for Chile suggests that both effects are playing in the role in the same direction: according to the CASEN Survey, the share of workers with complete tertiary education increased from 18 to 27% between 2003 and 2017, while the college wage premium with respect to workers with no education fell from 6.2 to 4.0 during the same period. The dynamics underlying the evolution of labor earnings dispersion in Chile appear to be very different from those in Brazil, where the reduction in differences across firms accounts for 40% of the decrease in inequality, while workers only account for 30% (Alvarez et al., 2018).

As discussed earlier, there is evidence of positive sorting, which contributes to 13% of the total variance, as on average high-earnings workers are more likely to be employed on high-paying firms, exacerbating earnings differences. However, as mentioned earlier, the correlation between worker and firm effects weakens over time, explaining nearly a quarter of the decline in labor earnings inequality. This finding is very similar to

[22]The previous findings are robust to the sample restriction of including only firms and individuals hired in firms with at least 20 full time equivalent employees. Table B.2 in the appendix provides similar results using an alternative cutoff of firms with only 3 FTE.

the evidence documented for Brazil, where this term explains 23% of the reduction of inequality between 1996 and 2012 (Alvarez et al., 2018). This result suggests a change in sorting patterns that could come from several sources, from changes in production technologies to the changes in the distribution of worker abilities. To provide a clearer description of the variation in worker allocations, we follow CHK and plot the joint distributions of the estimated worker and firm effects in the first and last periods, using the same decile combination cells as in Figure A.7. For both periods, sorting appears to be particularly strong at the upper end of the distribution, with higher labor earnings workers being allocated to firms that pay a larger premium (Figure A.9). However, this group of workers experiences a slight fall in its concentration over the decade, revealing a reduction in sorting for the higher-decile individuals.

A relevant question for the interpretation of sorting is what characteristics are captured by worker and firm effects. In the case of firms, high-paying firms might also be more productive, in which case the earnings function might be more closely related to a measure of the actual productivity of the match, and the discussion of earnings dispersion might be related more directly to a discussion on allocative efficiency. We follow this idea using the simple measure of firm labor productivity (revenue per full-time equivalent worker demeaned by sector), and include it both as an additional control in the standard AKM specification as well as a covariate in a regression that excludes firm fixed effects. Results are reported in Table 5, which compares the previous decomposition with the two decomposition exercises that account for firm productivity. Firm productivity appears to be largely irrelevant, both in explaining the dispersion of earnings or in correlating with worker characteristics in the equilibrium allocation, and the estimated firm effects are largely unaffected. This implies that firm effects are not driven by observed labor productivity, and are therefore related to other unobservable characteristics.

Table 5

Decomposition of the Decline in Labor Earnings Inequality

(a) With Firm Fixed Effect and Without Labor Productivity

	2005-08	**2009-12**	**2013-16**	**Change**
Var$(\ln w_{it})$	0.869	0.882	0.840	-0.029
Share of each component				
Var(Worker FE)	62	60	58	-155
Var(Firm FE)	8	8	9	43
Var(Covariates)	6	7	5	-35
Var(Covariates Worker)	5	6	5	-24
Var(Covariates Firm)				
Var(Time Effect)	0	1	0	-9
Var(Residuals)	11	11	12	22
2 Cov(WFE,FFE)	12	12	12	-14
2 Cov(WFE,Covariates)	2	1	3	33
2 Cov(FFE,Covariates)	1	1	1	7

(b) With Firm Fixed Effect and Firm Labor Productivity

	2005-08	**2009-12**	**2013-16**	**Change**
Var$(\ln w_{it})$	0.869	0.882	0.840	-0.029
Share of each component				
Var(Worker FE)	61	60	58	-153
Var(Firm FE)	7	8	9	59
Var(Covariates)	6	7	5	-43
Var(Covariates Worker)	5	6	5	-25
Var(Covariates Firm)	0	0	0	-4
Var(Time Effect)	0	1	0	-9
Var(Residuals)	10	11	12	22
2 Cov(WFE,FFE)	10	12	11	23
2 Cov(WFE,Covariates)	4	1	3	-5
2 Cov(FFE,Covariates)	2	1	2	-2

(c) Without Firm Fixed Effect and With Labor Productivity

	2005-08	**2009-12**	**2013-16**	**Change**
Var$(\ln w_{it})$	0.869	0.882	0.840	-0.029
Share of each component				
Var(Worker FE)	74	74	73	-95
Var(Firm FE)				
Var(Covariates)	8	9	6	-58
Var(Covariates Worker)	6	7	5	-34
Var(Covariates Firm)	1	1	1	-7
Var(Time Effect)	0	1	0	-10
Var(Residuals)	11	12	13	25
2 Cov(WFE,FFE)				
2 Cov(WFE,Covariates)	7	5	8	28
2 Cov(FFE,Covariates)				

Note: Statistics are computed for men in the largest connected set working in firms that report the Annual Affidavit on Income. Labor productivity is defined as revenue per full-time equivalent worker, demeaned by sector. See text for details.

Source: Authors' calculations based on SII and Register Office data.

Finally, we follow Song et al. (2019) to connect the results of Table 4 with the earlier decomposition between differences across firms in average wages and wage differentials inside the firm. Specifically, we use the estimated parameters from equation (4) to decompose the within and between-firm components of inequality, following equation (1). Table 6 shows the result of this additional decomposition, using the same notation as CHK. As in the earlier exercise in Figure 5, within-firm earnings dispersion explains the largest part (55%) of the overall variance in earnings. However, we can now study in more detail the underlying forces behind within-firm dispersion. Almost three quarters of within-firm dispersion comes from differences in worker firm effects. Therefore, a very large part of earnings differentials inside the firm are a reflection of systematic earnings gaps between workers of different types: firms hire heterogeneous workers, and pay them pay heterogeneous wages. For between-firm variance, the main driver also comes from systematic differences across workers, and the fact that average worker types vary across firms (19% of total variance). The firm specific premium plays a relatively small direct role (9%), although it is also important for the sorting pattern of higher-paid workers into higher-paying firms (13%). These results highlight, once again, the importance of worker characteristics in explaining overall earnings differentials, both within the firm - with a large degree of inside dispersion that comes directly from having workers with different market skills - and across firms - as the average composition of firms varies, with different firms employing different types of workers. This is exacerbated by the fact that high-type workers are more likely to work at high-type firms.

In terms of dynamics, the within-firm component remained relatively stable through the decade. Thus, the whole reduction in overall labor earnings inequality over the period was due to the smaller dispersion of average earnings between firms, which is reflected in the fact that the average composition of workers across firms became more similar. There are two main forces behind this result. First, sorting patterns became weaker, so high-earning individuals became less likely to work in high-paying firms. Second, segregation also fell, meaning that high-earning workers became less likely to sort together in a given firm. These patterns could be associated with several drivers, such as changes in technology and the sectoral structure of the economy. Interestingly, firm-level pay premium became more diverse, although it cannot offset the composition effects.

These results mirror the evidence for the United States, where most of the rise in inequality is due to changes in the composition of workers (Song et al., 2019) and the change in firm-pay premium does not play a role in explaining overall dispersion.

Table 6

Decomposition of the Decline in Labor Earnings Inequality
into Between and Within-Firm Variance

	2005-08	Share	2009-12	Share	2013-16	Share	Change	Share
Var($\ln w_{it}$)	0.910		0.915		0.864		-0.046	
Between-firm: Var($\bar{w}_t^j$)	0.420	46	0.421	46	0.376	43	-0.045	-97
Var($\overline{WFE}$)	0.181	20	0.174	19	0.146	17	-0.035	-76
Var($\overline{FFE}$)	0.075	8	0.078	9	0.082	9	0.007	15
Var($\overline{Covariates}$)	0.008	1	0.013	1	0.004	1	-0.004	-8
2Cov($\overline{WFE}, FFE$)	0.121	13	0.121	13	0.109	13	-0.013	-27
2Cov($\overline{WFE}, \overline{Covar.}$)	0.018	2	0.017	2	0.018	2	0.000	-1
2Cov($FFE, \overline{Covar.}$)	0.010	1	0.013	1	0.012	1	0.002	3
Within-firm: Var($w_{it} - \bar{w}_t^j$)	0.489	54	0.494	54	0.488	57	-0.001	-3
Var(WFE-$\overline{WFE}$)	0.366	40	0.366	40	0.353	41	-0.013	-28
Var($Covar.$-$\overline{Covar.}$)	0.041	4	0.048	5	0.035	4	-0.006	-13
Var($Residuals$)	0.091	10	0.095	10	0.098	11	0.006	14
2Cov(WFE-$\overline{WFE}$,$Covar.$-$\overline{Covar.}$)	-0.003	0	-0.010	-1	0.006	1	0.010	21
2Cov(WFE-$\overline{WFE}$,$Residuals$)	0.000	0	0.000	0	0.000	0	0.000	1
2Cov($Covar.$-$\overline{Covar.}$, $Residuals$)	0.000	0	0.000	0	0.000	0	0.000	0
Segregation index	0.332		0.323		0.293			

Note: Segregation index is defined as: $\frac{Var(\overline{WFE})}{Var(WFE)}$. Statistics are computed for men in the largest connected set. See text for details.

Source: Authors' calculations based on SII and Register Office data.

IV.3 A Further Look at Complementarities between Firms and Workers

As discussed in the previous section, the AKM model assumes a linear structure for the effects of worker and firm heterogeneity in earnings. The presence of systematic complementarities between workers and firms would constitute a violation to the assumptions of the model. Table 7 calculates the serial correlation of worker and firm fixed effects over the same three 4-years intervals as the previous exercises, and shows that the correlation over time is not equal to one. This is particularly the case for the estimates of firm effects, and is consistent with the presence of complementarities, as the fixed effects might not be properly identified.

Table 7

Serial Correlation of Worker and Firm Fixed Effect

	Worker Fixed Effect			Firm Fixed Effect		
	2005-08	**2009-12**	**2013-16**	**2005-08**	**2009-12**	**2013-16**
2005-08	1.000			1.000		
2009-12	0.832	1.000		0.574	1.000	
2013-16	0.759	0.848	1.000	0.543	0.612	1.000

Note: Statistics are computed for men in the largest connected set. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Thus, it seems natural to compare the estimates of the AKM model with the results of a more general model that does not assume linearity and can account for complementarity patterns in the earnings functions of firm-worker matches. In particular, we use the BLM empirical framework developed by Bonhomme et al. (2019), which explicitly allows for nonlinear interactions between workers and firms. Given the difficulty of solving a model where millions of workers and dozens of thousands firms could be connected, the authors choose to reduce the dimension of the problem by modelling the unobserved firm heterogeneity by classes of firms.

As mentioned in the introduction, comparing the results of the AKM model and the more general framework is particularly interesting in the case of Chile, as Bonhomme et al. (2020) show that limited mobility can be a significant source of bias in the estimates of worker and firm effects in linear models. Given the large degree of mobility in the Chilean formal labor market, reflected in the size of the connected set, it is natural wonder whether a similar pattern emerges in this context.

Table 8 compares the average estimates of versions of AKM and the BLM models[23]. Both models ares estimated for 3-years rolling windows between 2005 and 2016 (detailed estimation for the ten intervals are presented in Figure A.10). Although, the estimated contribution of worker heterogeneity is almost the same for both models, there are some discrepancies over the roles of firm dispersion and sorting, which appear to be in line with the discussion in Bonhomme et al. (2020), even in a context in which limited mobility does not appear to be an issue. Sorting is significantly stronger in the BLM estimates, explaining 20% of overall earnings variance while in the AKM estimation it accounts for 14% . Conversely, time-invariant firm characteristics play a smaller role in explaining variance in BLM, with a meagre 4% compared to 14% in this version of the

[23]Since the BLM model does not include workers characteristics as covariates, we exclude them from both estimations.

AKM estimation. Thus, estimates from the BLM model reduce the importance of firm heterogeneity in explaining earnings difference - reinforcing the relative importance of worker heterogeneity -, but enhance the importance of sorting patterns between high-type workers and high-paying firm classes in amplifying earnings differences.

Table 8

Decomposition of Labor Earnings Inequality
(%, average 2005-2016)

	Without Complementarities between WFE and FFE (AKM Model)	**With Complementarities between WFE and FFE (BLM Model)**
Var(Worker FE)	72	76
Var(Firm FE)	14	4
2 Cov(WFE,FFE)	14	20

Note: Statistics are computed for men in the largest connected set. See text for details.
Source: Authors' calculations based on SII and Register Office data.

V Conclusions

We use a matched employer-employee census of formal employment of Chile between 2005 and 2016 to analyze labor earnings inequality in the context of a small open economy with a high level of labor mobility. We document a decline in labor earnings inequality in this decade, mostly explained by a compression in the upper-tail of the distribution. This result is mainly driven by a reduction in the dispersion of average wages between firms while the within-firm component has remained roughly constant, even if both components explain a similar proportion of the overall dispersion. The pattern we obtain for the aggregate holds in almost every economic sector and particularly in bigger and more productive firms.

In addition, we estimate a two-way fixed effect model (AKM model) for individual earnings to compare the contributions of firm and worker heterogeneity toward changes in dispersion. We find that time-invariant unobserved characteristics of workers, associated to the market returns on their skills, account for nearly 60% of overall dispersion. The reduction in the heterogeneity of worker effects is by far the main driver of the decline in dispersion, especially associated to a reduction of the average worker effect between firms.

Finally, we compare our results with a model that allows for worker-firm complementarities (BLM model). While the large role of worker effects is robust across both methodologies, there are important differences in the importance of firm effects and sorting patterns.

Our results are consistent with evidence for other countries, which have also found that the evolution of inequality is almost completely explained by the movements in the between-firm component (Song et al., 2019; Alvarez et al., 2018). Therefore, an interesting question for future research can be to have a better understanding of the stability of the earnings distribution within firms, and of its determinants. Our results suggest that firm productivity correlates strongly with within-firm dispersion and that, consequently, movements in within-firm dispersion at the aggregate level can be associated with the evolution of the productivity distribution.

References

Abowd, J., Kramarz, F., & Creecy, R. (2002). Computing Person and Firm Effects Using Linked Longitudinal Employer–Employee Data. *Center for Economic Studies, US Census Bureau*.

Abowd, J., Kramarz, F., & Margolis, D. (1999). High Wage Workers and High Wage Firms. *Econometrica*, *67*(2), 251–333.

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification Properties of Recent Production Function Estimators. *Econometrica*, *83*(6), 2411–2451.

Albagli, E., Chovar, A., Luttini, E., Madeira, C., Naudon, A., & Tapia, M. (2017). Labor Market Flows: Evidence from Chile Using Micro Data from Administrative Tax Records. *Working Paper*, *812*, Central Bank of Chile.

Alvarez, J., Benguria, F., Engbom, N., & Moser, C. (2018). Firms and the Decline in Earnings Inequality in Brazil. *American Economic Journal: Macroeconomics*, *10*(1), 149–189.

Amar, M. (2010). Les Tres Hauts Salaires du Secteur Privé. *INSEE Première*, *1844*, 4.

Atkinson, A., Casarico, A., & Voitchovsky, S. (2018). Top Incomes and the Gender Divide. *Journal of Economic Inequality*, *16*(2), 225–256.

Atkinson, A., Piketty, T., & Saez, E. (2011). Top Incomes in the Long Run of History. *Journal of Economic Literature*, *49*(1), 3–71.

Barth, E., Bryson, A., Davis, J., & Freeman, R. (2016). It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States. *Journal of Labor Economics*, *34*(S2), S67–S97.

Becker, G., & Murphy, K. (1994). The Division of Labor, Coordination Costs, and Knowledge. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*, 299–322.

Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., & Setzler, B. (2020). How Much Should We Trust Estimates of Firm Effects and Worker Sorting? *Working Paper*, *27368*.

Bonhomme, S., Lamadon, T., & Manresa, E. (2019). A Distributional Framework for Matched Employer Employee Data. *Econometrica*, *87*(3), 699–739.

Brown, C., Hamilton, J., & Medoff, J. (1990). *Employers Large and Small.* Harvard University Press.

Burkhauser, R., Herault, N., Jenkins, S., & Wilkins, R. (2018). Survey Under–Coverage of Top Incomes and Estimation of Inequality: What is the Role of the UK's SPI Adjustment? *Fiscal Studies*, *39*(2), 213–240.

Caliendo, L., Monte, F., & Rossi-Hansberg, E. (2015). The Anatomy of French Production Hierarchies. *Journal of Political Economy*, *123*(4), 809–852.

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and Labor Market Inequality: Evidence and some theory. *Journal of Labor Economics*, *36*(S1), S13–S70.

Card, D., Heining, J., & Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *Quarterly Journal of Economics*, *128*(3), 967–1015.

Chaney, T., & Ossa, R. (2013). Market size, Division of Labor, and Firm Productivity. *Journal of International Economics*, *90*(1), 177–180.

Denk, O. (2015). Financial Sector Pay and Labour Income Inequality: Evidence from Europe. *OECD Economics Department Working Papers*(1225).

Eberhard, J., & Engel, E. (2009). The Educational Transition and Decreasing Wage Inequality in Chile. *Working Document Department of Economics Yale University*.

Engbom, N., & Moser, C. (2017). Earnings Inequality and the Minimum Wage: Evidence from Brazil. *CESifo Working Paper Series*(6393).

Fortin, N., Bell, B., & Böhm, M. (2017). Top Earnings Inequality and the Gender Pay Gap: Canada, Sweden, and the United Kingdom. *Labour Economics*, *47*, 107–123.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition Methods in Economics. In *Handbook of Economics* (Vol. 4, pp. 1–102). Elservier.

Hoffmann, E., & Malacrino, D. (2019). Employment Time and the Cyclicality of Earnings Growth. *Journal of Public Economics*, *169*, 160–171.

Kaplan, S., & Rauh, J. (2009). Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes? *Review of Financial Studies*, *23*(3), 1004–1050.

Lam, D. (1994). The Variance of Population Characteristics in Stable Populations, with Applications to the Distribution of Income. *Population Studies*, *38*(1), 117–127.

Muñoz, E., Rucci, G., F., S., & Urzúa, S. (2018). Wage Inequality in Latin America: Learning from Matched Employer-Employee Data. *IDB Working Paper Series*, *1004*.

Parro, F., & Reyes, L. (2017). The Rise and Fall of Income Inequality in Chile. *Latin American Economic Review*, *26*(3), 1–31.

Piketty, T., & Saez, E. (2003). Income Inequality in the United States, 1913-1998. *Quarterly Journal of Economics*, *118*(1), 1–41.

Sapelli, C. (2013). Movilidad Intrageneracional del Ingreso en Chile. *Estudios Públicos*, *131*, 1–35.

Solimano, A., & Torche, A. (2008). La Distribución del Ingreso en Chile 1987-2006: Análisis y Consideraciones de Política. *Working Paper*, *480*, Central Bank of Chile.

Song, J., Price, D. J., Guvenen, F., Bloom, N., & Von Wachter, T. (2019). Firming Up Inequality. *Quarterly Journal of Economics*, *134*(1), 1–50.

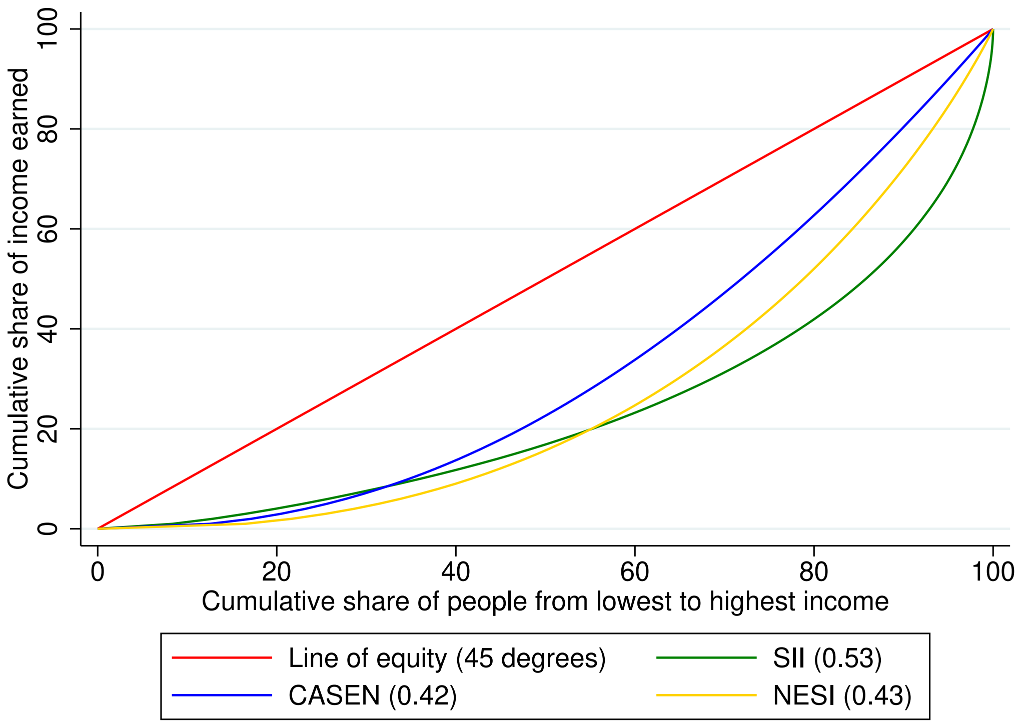
Appendix

.

Appendix A Additional Figures

Figure A.1

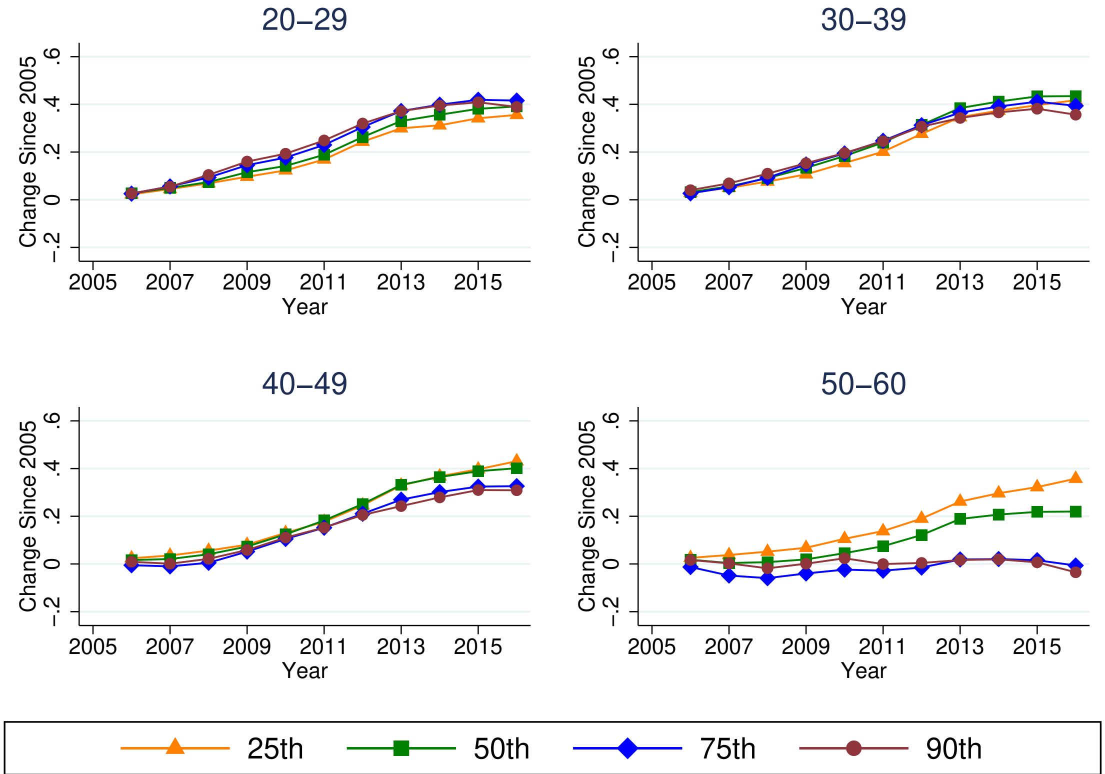
Lorenz Curve using alternative Data Sources
(2015)



Note: In parentheses, we present the Gini Coefficient for each series. We include all firms and all workers with non-zero labor earnings from the IRS monthly data. For the different surveys, we consider only formal wage employment to ensure a proper comparison with the SII data.
Source: Authors' calculations based on the data from SII, Ministry of Social Development (CASEN Survey), and Institute of National Statistics (NESI Survey).

Figure A.2

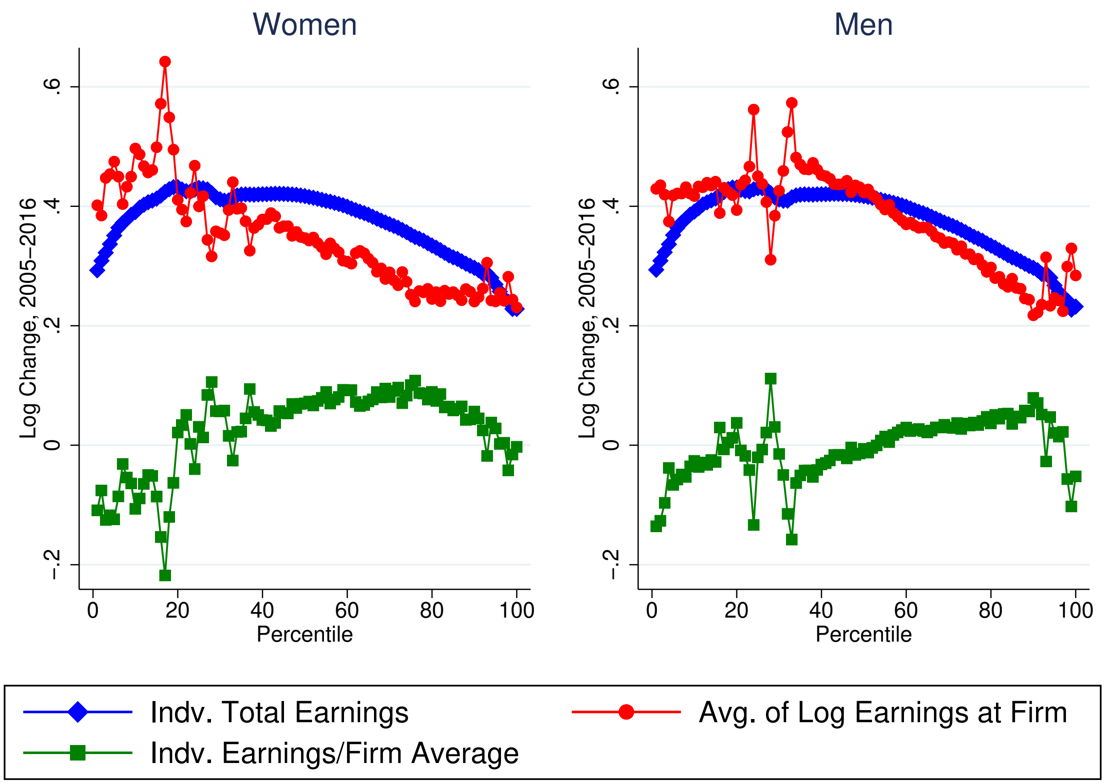
Change in Selected Percentiles of Annual Earnings Relative to 2005



Note: Statistics are computed for the baseline sample.
Source: Authors' calculations based on the SII and Register Office data.

Figure A.3

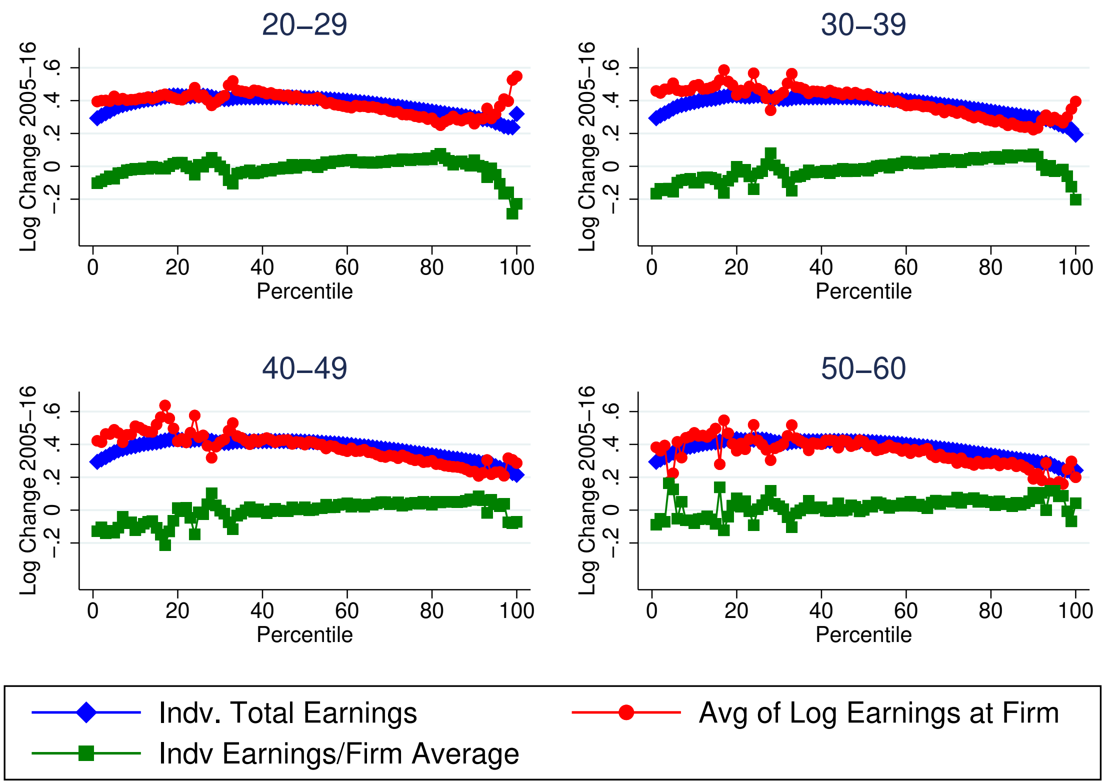
Change in Inequality across Percentiles from 2005 to 2016 By Gender



Note: Statistics are computed for the baseline sample. Percentiles are based on all individuals, regardless of gender. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Figure A.4

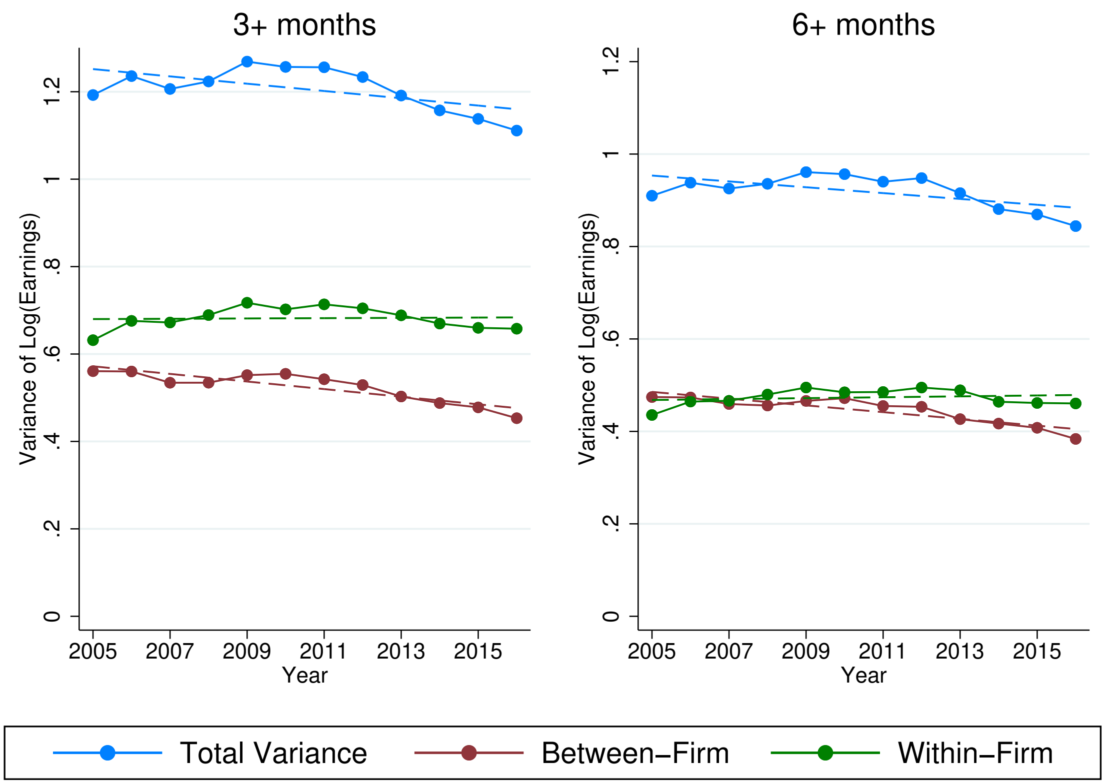
Change in Inequality across Percentiles from 2005 to 2016 By Age Group



Note: Statistics are computed for the baseline sample. Percentiles are based on all individuals, regardless of age. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Figure A.5

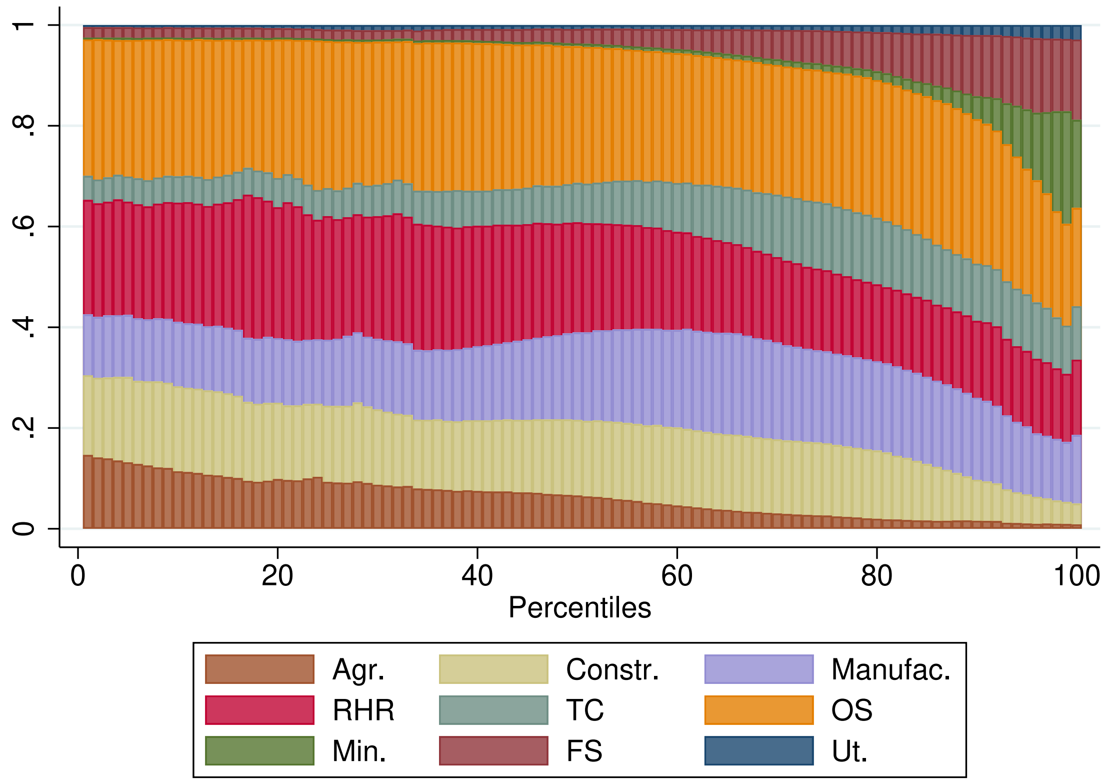
Variance of Log Earnings Between and Within Firms



Note: The definition of strong attachment to the formal labor market consider working at least 3 or 6 months in a given year, no matter if the annual earning is less than 3.25 times the monthly minimum wage.
Source: Authors' calculations based on SII and Register Office data.

Figure A.6

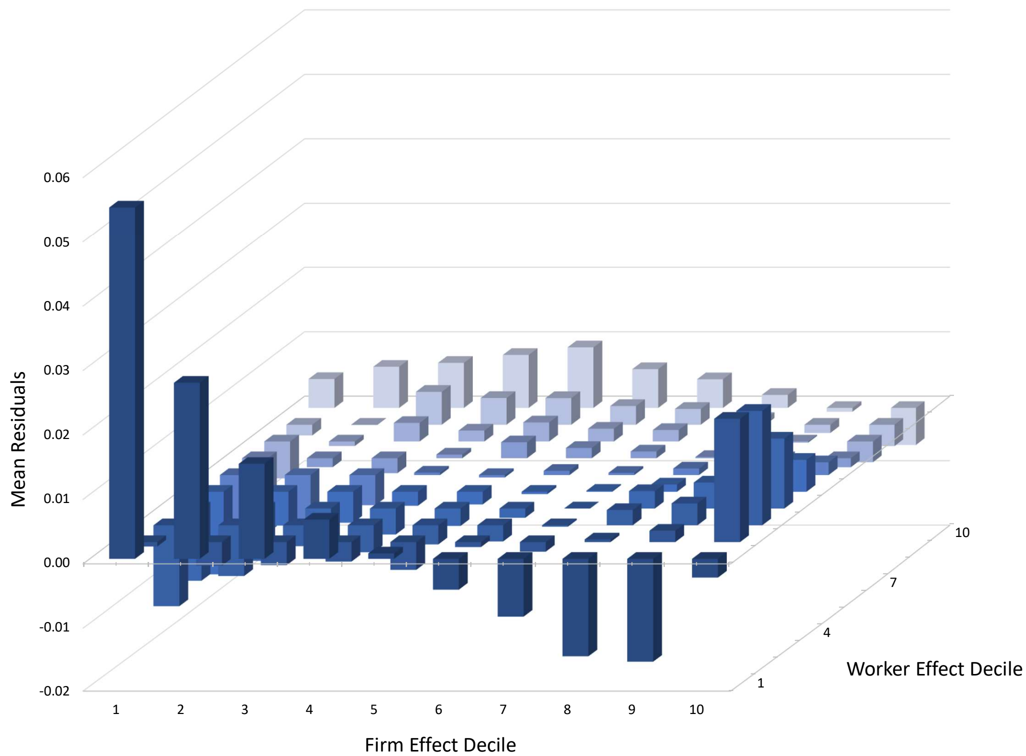
Share of Employment by Labor Earnings Percentile (average 2005-2016)



Note: The following is the list of acronyms and abbreviations used: Agriculture (Agr.); Mining (Min.); Manufacturing (Manufac.); Utilities (Ut.); Construction (Constr.); Retail, hotels, and restaurants (RHR); Transportation and communication (TC); Financial services (FS); and Other services (OS). Statistics are computed for the baseline sample. See text for further details.
Source: Authors' calculations based on SII and Register Office data.

Figure A.7

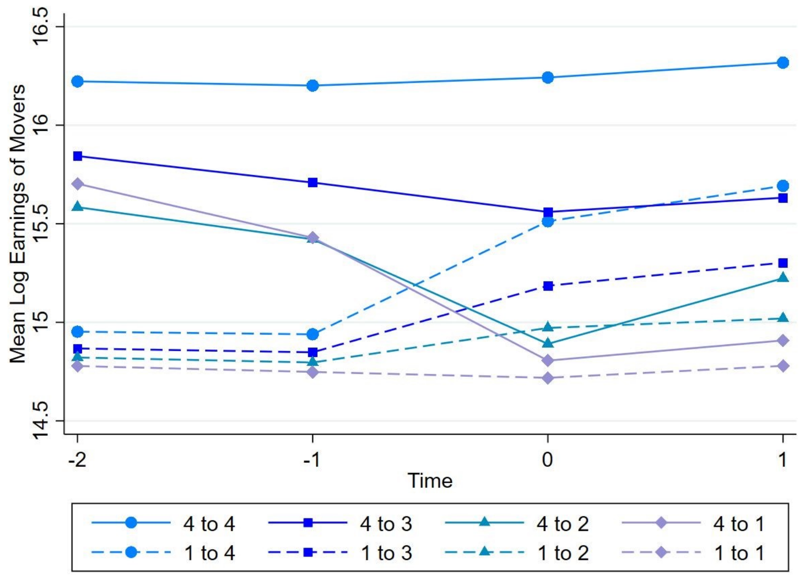
Mean Residuals by Worker and Firm Effects Deciles, 2009-2012



Note: Statistics are computed for men in the largest connected set.
Source: Authors' calculations based on SII and Register Office data.

Figure A.8

Mean Labor Earnings of Movers by Firm Effects Quartiles for Origin and Destination Firms, 2009-2012

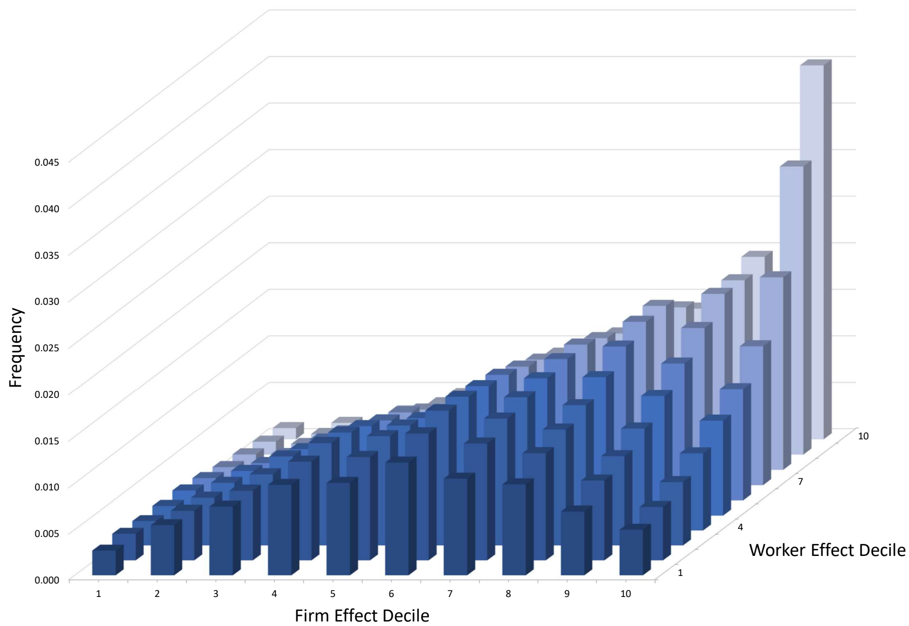


Note: Statistics are computed for men in the largest connected set.
Source: Authors' calculations based on SII and Register Office data.

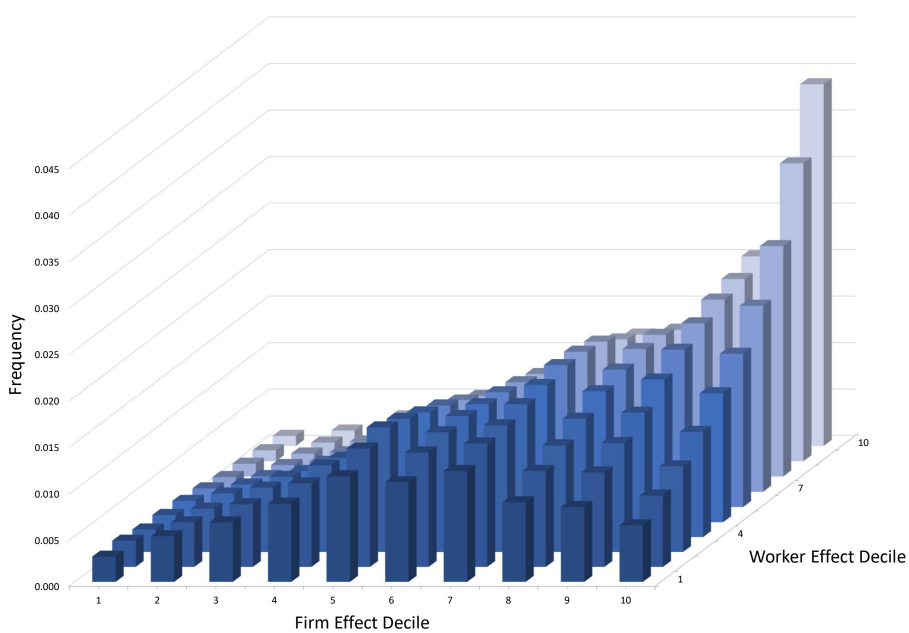
Figure A.9

Joint Distribution of Worker and Firm Effects

(a) 2005-2008



(b) 2013-2016



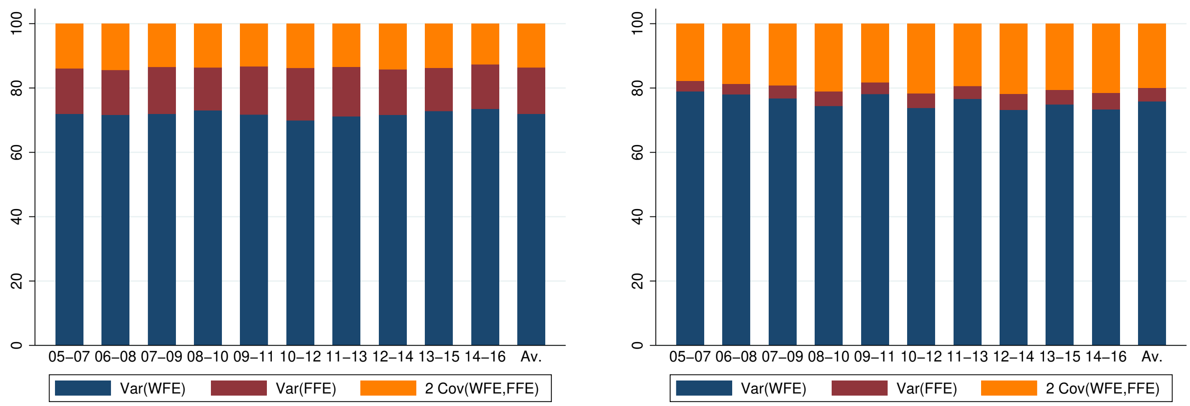
Note: Statistics are computed for men in the largest connected set.
Source: Authors' calculations based on SII and Register Office data.

Figure A.10

Role of Complementarities between Firms and Workers
in Explaining Labor Earnings Inequality

(a) Without Complementarities
(AKM Model)

(b) With Complementarities
(BLM Model)



Note: Statistics are computed for men in the largest connected set.
Source: Authors' calculations based on SII and Register Office data.

Appendix B Additional Tables

Table B.1

Variance of Log Earnings Between and Within Firms,
for stayers (firms and employees) and overall

2005-16	**Average**	**Change**
(a) Job keepers		
Var (log earnings)	0.90	-0.10
Between-firm	0.43	-0.08
Within-firm	0.47	-0.01
(b) Continuing firms		
Var (log earnings)	0.95	-0.10
Between-firm	0.44	-0.09
Within-firm	0.51	-0.01
(c) All employees & all firms		
Var (log earnings)	0.95	-0.10
Between-firm	0.44	-0.09
Within-firm	0.51	0.00

Note: Statistics are computed for the baseline sample. See text for details.
Source: Authors' calculations based on SII and Register Office data.

Table B.2: Decomposition of the Decline in Labor Earnings Inequality

(a) Using firms with 20+ FTE for AKM and 20+ FTE for decomposition

	2005-08	**2009-12**	**2013-16**	**Change**
Var($\ln w_{it}$)	0.910	0.915	0.864	-0.046
Share of each component				
Var(Worker FE)	60	59	58	-103
Var(Firm FE)	8	8	9	15
Var(Covariates)	5	7	5	-21
Var(Residuals)	10	10	11	14
2 Cov(WFE,FFE)	13	13	13	-27
2 Cov(WFE,Covariates)	2	1	3	20
2 Cov(FFE,Covariates)	1	1	1	3

(b) Using firms with 3+ FTE for AKM and 3+ FTE for decomposition

	2005-08	**2009-12**	**2013-16**	**Change**
Var($\ln w_{it}$)	0.897	0.910	0.871	-0.026
Share of each component				
Var(Worker FE)	60	58	57	-143
Var(Firm FE)	11	11	13	37
Var(Covariates)	5	6	4	-38
Var(Residuals)	10	11	11	27
2 Cov(WFE,FFE)	12	12	11	-32
2 Cov(WFE,Covariates)	1	0	2	44
2 Cov(FFE,Covariates)	1	1	1	6

(c) Using firms with 3+ FTE for AKM and 20+ FTE for decomposition

	2005-08	**2009-12**	**2013-16**	**Change**
Var($\ln w_{it}$)	0.913	0.919	0.870	-0.043
Share of each component				
Var(Worker FE)	59	58	57	-103
Var(Firm FE)	8	9	10	17
Var(Covariates)	5	6	4	-23
Var(Residuals)	10	11	12	15
2 Cov(WFE,FFE)	14	14	13	-32
2 Cov(WFE,Covariates)	2	1	3	23
2 Cov(FFE,Covariates)	1	1	1	3

Note: Statistics are computed for men in the largest connected set.
Source: Authors' calculations based on SII and Register Office data.

Appendix C Data Procedures

Our raw dataset comprises of 647,598 firms and 9,713,145 workers in an unbalanced panel between 2005 and 2016, summing up to 89,758,464 observations (see Column 1 in Table C.1). First, we drop 329,522 workers from the sample (1,905,652 observations), because we cannot match their non nominated identification number with the information provided by the Register Office (Column 2).

We then impose two restrictions to include only workers with a strong attachment to the formal labor market. We consider only workers aged 20 to 60 years, dropping 7,092,066 observations (Column 3). Also, we eliminate 9,606,967 observations of workers earning less than 3.25 times the monthly minimum wage in a given year (Column 4).

Next, we drop 13,144 observations of workers for whom we are not able to choose their primary job in a given year and 18,616,192 observations of labor relations that are not the primary job (Column 5). We also drop 14,945 employers and the employees who worked primarily at those firms (summing 7,518,636 observations), because they belong to the public sector or we cannot assign them to an economic sector (Column 6).

Finally, we only consider employers and workers hired in firms with at least 20 employees, dropping 9,794,556 observations as a result (Column 7). We ended up working with an unbalanced panel of 50,589 firms and 6,192,080 employees (35,211,241 observations).

Table C.1

Number of observations per year after each step of the cleaning procedure

Year	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2005	5,813,522	5,576,596	5,189,149	4,547,067	3,484,390	2,982,182	2,276,504
2006	6,118,651	5,894,236	5,459,181	4,770,395	3,622,872	3,111,627	2,406,445
2007	6,874,756	6,648,510	6,107,453	5,375,502	3,946,414	3,410,558	2,649,363
2008	7,114,617	6,906,554	6,333,572	5,581,909	4,117,480	3,546,542	2,787,933
2009	6,671,855	6,504,250	6,016,894	5,213,900	4,002,547	3,404,899	2,663,678
2010	7,213,709	7,052,919	6,506,968	5,686,587	4,236,449	3,632,221	2,858,165
2011	7,765,780	7,614,095	6,998,421	6,151,902	4,464,810	3,830,579	3,015,632
2012	8,190,746	8,049,048	7,367,105	6,524,083	4,703,070	4,048,803	3,195,109
2013	8,431,562	8,307,393	7,603,492	6,759,540	4,898,040	4,193,121	3,310,591
2014	8,342,755	8,243,133	7,553,263	6,695,119	4,894,275	4,170,533	3,280,532
2015	8,390,979	8,305,587	7,614,103	6,730,696	4,953,645	4,276,488	3,345,747
2016	8,829,532	8,750,491	8,011,145	7,117,079	5,200,451	4,398,254	3,421,542
Total	89,758,464	87,852,812	80,760,746	71,153,779	52,524,443	45,005,807	35,211,241

Note: See text for details.
Source: Authors' calculations based on SII and Register Office data.

Appendix D Decomposition of Variance

We start by using the following decomposition of an individual wage:

$$y_t^{i,j} = \underbrace{\bar{y}_t^A}_{\text{economy average}} + \underbrace{(\bar{y}_t^j - \bar{y}_t^A)}_{\text{employer deviation}} + \underbrace{(y_t^{i,j} - \bar{y}_t^j)}_{\text{worker deviation}}$$

We rearrange terms:

$$y_t^{i,j} - \bar{y}_t^A = (\bar{y}_t^j - \bar{y}_t^A) + (y_t^{i,j} - \bar{y}_t^j)$$

We take variance on both sides, and get:

$$\sum_{j=1}^{J}\sum_{i=1}^{N_j}(y_t^{i,j} - \bar{y}_t^A)^2 = \sum_{j=1}^{J}\sum_{i=1}^{N_j}(\bar{y}_t^j - \bar{y}_t^A)^2 + \sum_{j=1}^{J}\sum_{i=1}^{N_j}(y_t^{i,j} - \bar{y}_t^j)^2 + 2\sum_{j=1}^{J}\sum_{i=1}^{N_j}(\bar{y}_t^j - \bar{y}_t^A)(y_t^{i,j} - \bar{y}_t^j)$$

We only consider the last term:

$$\begin{aligned}
\sum_{j=1}^{J}\sum_{i=1}^{N_j}(\bar{y}_t^j - \bar{y}_t^A)(y_t^{i,j} - \bar{y}_t^j) &= \sum_{j=1}^{J}\sum_{i=1}^{N_j}\bar{y}_t^j \times y_t^{i,j} - \bar{y}_t^j \times \bar{y}_t^j - \bar{y}_t^A \times y_t^{i,j} + \bar{y}_t^A \times \bar{y}_t^j \\
&= \sum_{j=1}^{J}\bar{y}_t^j \times y_t^j - \sum_{j=1}^{J}(\bar{y}_t^j)^2 \times N_j - \bar{y}_t^A\sum_{j=1}^{J} y_t^j + \bar{y}_t^A\sum_{j=1}^{J}\bar{y}_t^j \times N_j \\
&= \sum_{j=1}^{J}\bar{y}_t^j \times y_t^j - \sum_{j=1}^{J}\bar{y}_t^j \times y_t^j - \bar{y}_t^A\sum_{j=1}^{J} y_t^j + \bar{y}_t^A\sum_{j=1}^{J} y_t^j = 0
\end{aligned}$$

Then,

$$\sum_{j=1}^{J}\sum_{i=1}^{N_j}(y_t^{i,j} - \bar{y}_t^A)^2 = \sum_{j=1}^{J}\sum_{i=1}^{N_j}(\bar{y}_t^j - \bar{y}_t^A)^2 + \sum_{j=1}^{J}\sum_{i=1}^{N_j}(y_t^{i,j} - \bar{y}_t^j)^2$$

$$\sum_{j=1}^{J}N_j\sum_{i=1}^{N_j}\frac{1}{N_j}(y_t^{i,j} - \bar{y}_t^A)^2 = \sum_{j=1}^{J}\sum_{i=1}^{N_j}(\bar{y}_t^j - \bar{y}_t^A)^2 + \sum_{j=1}^{J}N_j\sum_{i=1}^{N_j}\frac{1}{N_j}(y_t^{i,j} - \bar{y}_t^j)^2$$

$$N \times var(y_t^{i,j}) = \sum_{j=1}^{J}N_j \times (\bar{y}_t^j - \bar{y}_t^A)^2 + \sum_{j=1}^{J}N_j \times var_i(y_t^{i,j}|i \in j)$$

We divide both sides by N, and get:

$$var(y_t^{i,j}) = \underbrace{var_j(\bar{y}_t^j)}_{\text{Between-firm dispersion}} + \sum_{j=1}^{J}\underbrace{\omega_j \times var_i(y_t^{i,j}|i \in j)}_{\text{Within-firm-}j\text{ dispersion}}$$

Documentos de Trabajo **Banco Central de Chile**	**Working Papers** **Central Bank of Chile**
NÚMEROS ANTERIORES	PAST ISSUES
La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: www.bcentral.cl/esp/estpub/estudios/dtbc. Existe la posibilidad de solicitar una copia impresa con un costo de Ch$500 si es dentro de Chile y US$12 si es fuera de Chile. Las solicitudes se pueden hacer por fax: +56 2 26702231 o a través del correo electrónico: bcch@bcentral.cl.	Working Papers in PDF format can be downloaded free of charge from: www.bcentral.cl/eng/stdpub/studies/workingpaper. Printed versions can be ordered individually for US$12 per copy (for order inside Chile the charge is Ch$500.) Orders can be placed by fax: +56 2 26702231 or by email: bcch@bcentral.cl.

DTBC – 891
Determinantes del rating soberano: el caso de Chile
Natalia Gallardo, Mauricio Hitschfeld

DTBC – 890
The impact of the Covid public policies on the Chilean households
Carlos Madeira

DTBC – 889
Surveying the survey: What can we learn about the effects of monetary policy on inflation expectations?
Michael Pedersen

DTBC – 888
Does the Exposure to the Business Cycle Improve Consumer Perceptions for Forecasting? Microdata Evidence from Chile
Fernando Faure, Carlos A. Medel

DTBC – 887
Railroads, specialization, and population growth in small open economies: Evidence from the First Globalization
Andrés Forero, Francisco A. Gallego, Felipe González, Matías Tapia

DTBC – 886
High Dimensional Quantile Factor Analysis
Andrés Sagner

DTBC – 885
Heterogeneous Paths of Industrialization
Federico Huneeus, Richard Rogerson

DTBC – 884
Does the Commodity Super Cycle Matter?
Andrés Fernández, Stephanie Schmitt-Grohé, Martín Uribe

DTBC – 883
Twitter-Based Economic Policy Uncertainty Index for Chile
Andrés Sagner, Juan Sebastián Becerra

DTBC – 882
Corporate-Sector Functional Currency: An International Comparison
Jorge Fernández, Fernando Pino, Francisco Vásquez

DTBC – 881
Back testing fan charts of activity and inflation: the Chilean case
Jorge Fornero, Andrés Gatty

DTBC – 880
Financing Firms in Hibernation during the COVID-19 Pandemic
Tatiana Didier, Federico Huneeus, Mauricio Larrain, Sergio L. Schmukler

DTBC – 879
Choice Aversion in Directed Networks
Jorge Lorca, Emerson Melo

DTBC – 878
Big G
Lydia Cox, Gernot Muller, Ernesto Pasten, Raphael Schoenle, Michael Weber

DTBC – 877
Sticky Capital Controls
Miguel Acosta-Henao, Laura Alfaro, Andrés Fernández

BANCO CENTRAL
DE CHILE