I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size

Jaime Arellano-Bover

DISCUSSION PAPER SERIES

Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size

Jaime Arellano-Bover
Yale University and IZA

ABSTRACT

Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size[*]

I study the long-term effects of landing a first job at a large firm versus a small one using Spanish social security data. Size could be a relevant employer attribute for inexperienced workers since large firms are associated with greater training, higher wages, and enhanced productivity. The key empirical challenge is selection into first jobs – for instance, more able people may land jobs at large firms. I address this challenge developing an instrumental-variables approach that, while keeping business-cycle conditions fixed, leverages variation in the composition of labor demand that labor-market entrants face. I find that initially matching with a larger firm substantially improves long-term outcomes such as lifetime income, and that these benefits persist through subsequent jobs. Additional results point to mechanisms related to search frictions and better skill-development at large firms. Together, these findings shed light on how heterogeneous firms persistently impact young workers' trajectories.

JEL Classification:	E24, J23, J24, J31, J62
Keywords:	first job, employer size, firm heterogeneity, young workers, lifetime income, on-the-job skills

Corresponding author:
Jaime Arellano-Bover
Yale University
37 Hillhouse Avenue
New Haven CT, 06511
USA
E-mail: jaime.arellano-bover@yale.edu

1 Introduction

Firms are heterogeneous along many dimensions including pay, productivity, training, management quality, or technology adoption.[1] The experiences of similar workers in different workplaces can be worlds apart. Consider a young person entering the labor market. Suppose that her first job is at a productive firm that trains its workers, is technologically advanced, has knowledgeable managers, and employs many coworkers with whom to interact. Alternatively, imagine she starts at an unproductive firm with no training schemes, outdated technologies, unsophisticated management, and few coworkers. From a long-term view, will it matter if she starts in the first or second firm? Why?

On the one hand, young workers are mobile (Topel and Ward, 1992), so initial matches might not be relevant in the long run; there will be time to find a good job later on. On the other hand, first employers could affect career paths: search for ensuing jobs could vary based on first-employer quality, and opportunities to learn useful skills might differ across firms. For a young adult in her formative years, these distinctions could persistently impact her working life. Research on firm-driven wage inequality has focused on contemporaneous worker-firm relationships (e.g. Card et al., 2013, 2018; Song et al., 2019). However, we know little of how workers are impacted by past employment at heterogeneous firms.

In this paper I use social security data from Spain to study how first-employer heterogeneity impacts young workers' careers. I focus on firm size (number of employees) and document a causal relationship between holding the first job at a larger or smaller firm and long-term labor market outcomes. Size is an appealing firm attribute since it works as a sufficient statistic for various hard-to-observe characteristics (e.g. training, wages, productivity, management quality).[2] I develop an instrumental-variable (IV) approach to address non-random sorting of labor market entrants and firms. The empirical strategy—which keeps business cycle conditions at entry fixed—leverages the timing of large firms' idiosyncratic labor-demand shocks in relation to young people's labor market entry, thus providing plausibly exogenous variation in the chances of joining a larger or smaller first employer.[3] I find that initially matching with a larger firm persistently improves labor-market prospects. The estimated effect is substantial, with an elasticity between lifetime income and first-employer size equal to 0.12.

The IV strategy uses variation in regional labor demand *composition* across time. The

---

[1]See Card et al. (2018) for pay differentials, Syverson (2011) for productivity, Lynch and Black (1998) for training, Bloom and Van Reenen (2007) for management quality, and Fabiani et al. (2005) for technology adoption.

[2]A longstanding literature documents a positive correlation between employer size and wages (Moore, 1911; Brown and Medoff, 1989; Oi and Idson, 1999). Workers at large firms undergo higher levels of and more structured training (Lynch and Black, 1998). The conceptual link between managerial talent and size goes back to Lucas (1978). Bloom and Van Reenen (2006) show a positive correlation between management quality and size. The hierarchical production literature (e.g. Garicano, 2000; Fox, 2009; Garicano and Rossi-Hansberg, 2015) documents the relationship between organizational practices and size.

[3]The relevant thought experiment is random assignment of entrants to be hired by a larger or smaller firm. Firms that differ in size are typically different in other attributes; all of which form part of the thought experiment of being hired by potential first employers of different sizes. The thought experiment is *not* to exogenously increase the size of a given firm.

logic underlying the IV is that idiosyncratic shocks in the hiring decisions of a small number of large firms can meaningfully affect regional labor-demand composition. The IV aims to isolate changes in the composition of labor demand while controlling for its level and, thus, capture exogenous changes in the probability of being hired by a larger or smaller firm. This variation—occurring across years and regions—implies that time and place of labor market entry, together with who happens to be hiring, will lead young people to be exposed to different propensities to join larger or smaller firms. I operationalize the IV by building a Bartik-approach (shift-share) instrument—constructed using the small-large firm hiring patterns observed in the data—and assigning a predicted first-employer size to each worker based on her birth region, education, and typical graduation year given their age and education.

The Spanish context provides rich variation in large-firm hiring shocks. During 1985–2003, the years of labor market entry I study, Spain underwent an economic transformation following adhesion to the EU in 1986 (Chislett, 2002). This period was characterized by an opening to trade, growth in foreign firms' investments, market reforms, and expansion of the country's regional infrastructures. These factors led to great dynamism in large firms opening and expanding operations across different parts of the country. This variation allows me to keep constant the effects that cyclical conditions at entry might have on long-term outcomes, which has been the focus of previous work.[4] I keep cyclical conditions constant by controlling for regional unemployment rates, thus only using variation in large-firm hiring that is uncorrelated with business-cycle trends.

My results show that first-employer characteristics can shape workers' long-term career prospects. The raw data already display a positive correlation between lifetime income (a cumulative measure of many years of monthly labor income) and first-employer size. Figure 1 illustrates this unconditional correlation.[5] Adding controls and using the IV approach to account for workers' unobservable characteristics confirms the patterns in the raw data: I estimate a positive IV elasticity of lifetime income with respect to first-employer size, equal to 0.12. The magnitudes are meaningful: a one standard-deviation increase in log first-employer size is associated with a 27.7% increase in lifetime income. The first stage, which does a good job at predicting first-employer size, implies that, at least for some, luck plays a role in the key process of matching with heterogeneous first employers. The effect on lifetime income can be attributed both to an increase in average daily wages, explaining 74% of the effect on income, and an increase in total days worked, explaining the remaining 26%.

The IV estimate of the elasticity between lifetime income and first-employer size is four times larger than the OLS. In a context of heterogeneous effects, this is consistent with "compliers", those whose first-employer match is more susceptible to the labor-demand IV, benefiting the most from a first job at larger firms. Building on results from Angrist and Imbens (1995), I estimate "complier weights" that shed light on who are the people whose first jobs are most affected by the variation the IV captures. I find that compliers tend to be less edu-

---

[4]See Oyer (2006); Kahn (2010); Oreopoulos et al. (2012); Brunner and Kuhn (2014); Altonji et al. (2016); Fernández-Kranz and Rodríguez-Planas (2018); Schwandt and Von Wachter (2019).

[5]U.S. panel survey data show a similar correlation. See Appendix Figure A1.

Figure 1: Positive correlation between lifetime income and first-employer size
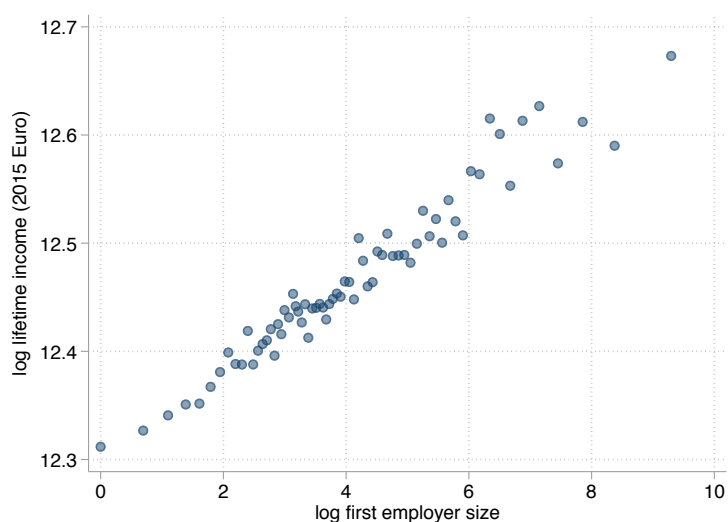
Notes: Conditional expectation of lifetime income as a function of first-employer size. Binned scatterplot. Log lifetime income (as defined in the text) on the vertical axis. Log size of worker's first employer on the horizontal axis. Sample of male workers of all education levels, born in Spain between 1968–1980.

cated and from less urban areas. This LATE result indicates that I capture the causal effect for younger entrants with lower earnings potential, who might be of particular interest.

Although my empirical analysis keeps constant business cycle conditions at entry, I document that the effect of starting one's career at a larger or smaller firm underpins part of the widely studied effects of entering the labor market during a recession. I quantify this relationship equipped with estimates of the first-employer size effect, estimates of the relationship between unemployment conditions and size of hiring firms, and existing estimates of the "graduating-in-a-recession" effect in the context of Spain from Fernández-Kranz and Rodríguez-Planas (2018). I find that 7%–15% of the losses from entering the labor market during a recession could be explained by the fact that during bad economic times young entrants are more likely to match with smaller first employers.

Aiming to understand the mechanisms behind the first-employer size effect, I first confirm that the lifetime effect is truly persistent, not solely stemming from time spent at the first job. Evidence for persistence includes the low fraction of lifetime income that is earned at the first job (due to job mobility and rising wages), and first-employer effects that are still present at age 35 (an age at which income trajectories have stabilized and 93.2% of workers have moved on from their first job). Based on the persistence of the results, I focus on the mechanisms that the literature identifies as main sources of life-cycle wage growth: job search and human capital accumulation (Rubinstein and Weiss, 2006).

Using the same IV approach, I find that first-employer size has a positive causal effect on the size of ensuing employers (second employer, and employer at age 35). That is, a larger first employer leads to larger subsequent ones. This finding is consistent with models of frictional search which show how on-the-job search can result in a job ladder where workers subsequently move "up" to more desirable firms (e.g. Burdett and Mortensen, 1998;

Christensen et al., 2005; Krolikowski, 2017). If large firms are generally more desirable, a larger first employer would result in a "higher" starting point in the ladder.[6] Workers would then only voluntarily switch jobs to move to an even more desirable firm. Persistent first-employer effects would arise from subsequent jobs that are of differential quality based on first-employer size.

I then document that the first-employer size effect is present even for the subset of workers who experience an unemployment spell between their first and second jobs. This result is suggestive of a human capital channel based on the insight, present in models of on-the-job search, that unemployment destroys search capital (disrupts a job ladder progression) but has lesser effects on human capital. Thus, long-term positive effects for those with a E-U-E first-to-second job transition are consistent with a human capital channel, but harder to explain with a pure search channel. Young workers in their formative period could acquire differentially valuable skills at large firms due to higher workforce training, learning from better peers and managers, or working in a more productive environment.[7]

Finally, as additional evidence consistent with a human capital channel, I show that the returns to *experience obtained at large firms* are higher than returns to experience obtained elsewhere. I reach this conclusion estimating monthly-panel wage equations that, while controlling for unobserved worker heterogeneity and *current* employer characteristics, allow differential returns to experience acquired at the largest firms. This approach compares wages of people who work for observably similar employers and have the same amount of experience, but acquired this experience in different—large or small—firms. The results show that one year of experience in the largest firm-size group is around 2.5 percentage points more valuable than one year of experience acquired elsewhere,[8] and that this differential is more important the younger a worker is.

The rest of the paper is organized as follows. Section 2 discusses related literature. Section 3 describes the data and measurement. Section 4 presents the analysis of the causal effects of first-employer size on long-term outcomes. Section 5 studies the persistence of these effects and mechanisms. In Section 6 I estimate the differential return to large-firm experience. Section 7 concludes. Several appendices contain supplementary results and robustness checks.

2   Relation and Contribution to the Literature

A growing literature studies firm-driven wage differentials, showing how similar workers can earn substantially different wages at different employers (e.g. Card et al., 2013; Goldschmidt and Schmieder, 2017; Sorkin, 2018; Abowd et al., 2018; Card et al., 2018; Song et al., 2019; Bonhomme et al., 2019). While this literature focuses on contemporaneous worker-

---

[6]U.S. evidence supports, among mature firms, a job ladder in terms of firm size (Haltiwanger et al., 2018). Sorkin (2018) documents a positive correlation between firm desirability and size.

[7]Kugler and Verhoogen (2012) document a positive correlation between plant size and the quality of inputs and outputs. Larger employers tend to be more productive (Moral-Benito, 2018), and faster to adopt new technologies (Fabiani et al., 2005). For learning from coworkers see Nix (2017) and Jarosch et al. (2018).

[8]As a benchmark, the average annual wage growth during the first eight years in the labor market is 10%.

firm relationships, little is known about persistent effects stemming from past employment at heterogeneous firms.[9] This paper is the first to establish a direct causal link between young workers' first-employer characteristics and long-term career outcomes, showing how early-career firm heterogeneity can have persistent implications for long-term inequality.[10] My IV setting, leveraging plausibly exogenous demand-side variation in the probability of joining large or small firms, contrasts the common approach of relying on realized workers' firm-switching to identify firm effects.

A long tradition, going back to Moore (1911), documents a robust positive correlation between earnings and employer size (Brown and Medoff, 1989; Oi and Idson, 1999).[11] The literature, however, has not agreed on why the premium exists, nor determined whether it has a causal component.[12] This paper documents, in a causal way, that matching with larger first employers leads to persistently better labor market outcomes. I also provide evidence consistent with large firms providing better skill-development opportunities.

Another literature documents sizable and persistent earnings losses associated with entering the labor market during bad economic times (Kahn, 2010; Oreopoulos et al., 2012; Brunner and Kuhn, 2014; Altonji et al., 2016; Fernández-Kranz and Rodríguez-Planas, 2018; Schwandt and Von Wachter, 2019). The "graduating-in-a-recession" effect is related to the first-employer size effect since inexperienced workers are more likely to be hired by large employers during booms.[13] Other work related to graduating in a recession includes Beaudry and DiNardo (1991), Devereux (2002), and Oyer (2006). In spite of this body of work, not much is known of the underlying mechanisms.[14] This paper improves our understanding of the mechanisms behind the importance of labor market entry, finding explicit support for both job search and human capital explanations. By studying first-employer heterogeneity, one of the suggested mechanisms in this literature, but doing so while keeping constant business-cycle fluctuations, I am able to quantify how much of the "graduating-in-a-recession" effect can be explained by the first-employer size effect.[15]

---

[9]Abowd et al. (2018) and Bonhomme et al. (2019) provide some evidence on dynamic implications of employment at heterogeneous firms. Abowd et al. (2018) argue that employment in year t in a top-paying firm leads to a higher probability of upward movements in the earnings distribution in year $t + 1$. Bonhomme et al. (2019) document how a worker's past firm-type may impact earnings after changing jobs.

[10]Other papers studying first jobs have focused on specialized workers not representative of the labor force such as Ph.D. economists, MBAs (Oyer, 2006, 2008) or CEOs (Schoar and Zuo, 2017).

[11]Lallemand et al. (2007) document the size premium for Spain and other European countries. Bloom et al. (2018) document that the firm-size premium, while considerable, has been declining in the U.S.. This declining trend is not present in other countries (Colonnelli et al., 2018).

[12]Some papers have tried to address endogenous sorting of workers across firm sizes (Idson and Feaster, 1990; Main and Reilly, 1993; Albæk et al., 1998). They rely, however, on exclusion assumptions of worker characteristics that could themselves depend on labor market outcomes (i.e. marital status or family composition).

[13]See Oreopoulos et al. (2012); Brunner and Kuhn (2014). See also Moscarini and Postel-Vinay (2012).

[14]Oreopoulos et al. (2012) do some work on mechanisms by documenting that graduating in a recession leads to higher job mobility and matches with lower-quality employers (measured by size and average wages). However, while the overall effect of graduating in a recession is causally identified, the subsequent sorting response of graduates across employer types is not. Heterogeneous responses attributed to employer quality could be driven by unobserved worker characteristics. Oreopoulos et al. (2012) describe this issue and discuss unreported estimates of the heterogeneous employer-driven response taking into account control functions with the fraction of workers starting to work at high-quality firms.

[15]Other work aiming to understand graduating-in-a-recession effects are Kwon et al. (2010); Liu et al. (2016); Wee (2016).

Previous work has assigned importance to young workers' initial job experiences. Some theoretical work focuses on skill-development reasons (e.g. Jovanovic and Nyarko, 1997; Gibbons and Waldman, 2006). Some empirical work has focused on the German apprenticeship system: von Wachter and Bender (2006) document long-lasting losses for those who involuntarily separate from their training firm;[16] Müller and Neubaeumer (2018) argue that training at a larger firm leads to lower unemployment later on. Kramarz and Skans (2014) study how parental ties induce first-employer matches but do not study long-term outcomes.

Finally, by documenting differential wage growth paths induced by heterogeneous first employers, this paper contributes to a vast literature on the sources of life-cycle wage growth (see Rubinstein and Weiss, 2006, for a review). Human capital is one of the main channels behind wage growth, and I provide evidence consistent with heterogeneous skill-development opportunities across employers (put forward theoretically by Rosen, 1972).[17]

3   Data, Sample Selection, and Measurement

3.1   Spanish Social Security Administrative Records

My principal data source is the Continuous Sample of Employment Histories (*Muestra Contínua de Vidas Laborales*, or MCVL), a 4% non-stratified random sample of Spanish Social Security administrative records, extracted yearly from 2004 to 2015. The sample is drawn from the population of those who in a given year have a relationship with Social Security (workers, unemployed receiving benefits, and pensioners). The data have a panel nature: those initially sampled are also selected each following year conditional on them still having a relationship with Social Security. The sample is refreshed each year to preserve representativeness.

The data include, at a monthly frequency, full labor market histories of sampled workers. Employment histories go as far back as 1967. Earnings start being recorded in 1980. Worker demographics include place of birth, date of birth, and sex. Educational attainment is also observed since this information is merged from municipal registries. While education is a key variable when studying youth labor market entry, many times it is not recorded in administrative datasets of employment and earnings, making MCVL well-suited for this topic. I group educational attainment levels into three categories: high school, vocational, and college. For each employment spell (employee-employer relationship) I observe its start and end date, an anonymized employer identifier, type of contract (permanent/temporary), professional category (*grupo de cotización*), and each month's payroll taxable base.

---

[16]von Wachter and Bender (2006) and Fackler et al. (2017) find larger losses for German workers who involuntarily separated from larger firms.

[17]I test for a human capital channel using mobility of young workers across different types of employers and unemployment. von Wachter and Bender (2008) study within-firm entry-cohort effects at German car manufacturing establishments and discern mechanisms with a similar logic. Their results suggest that, *within-firms*, differential skill investments do not explain persistent entry-cohort effects. In this paper, I argue that differential skill investments *across different firms* partly explain persistent first-employer effects.

The monthly taxable base is a censored measure of monthly earnings. It is bottom- and top-coded with limits that vary across years and professional category groups. I follow a procedure similar to Bonhomme and Hospido (2017) to impute monthly earnings for censored observations.[18] Censored observations are relatively few: 8.7% and 3% of observations are top- and bottom-coded, respectively, in the monthly panel. Since the taxable base of the self-employed is not a function of their monthly income, I do not observe earnings for self-employed workers.

The data include a flag for receipt of unemployment benefits. I use the type of benefits received (contributive or not), the benefits formula, and workers' employment and earnings histories to impute monthly unemployment benefits. I include unemployment income in lifetime income measures.

Social security records are matched with uncensored annual earnings tax data for the years 2006-2015. The downside from using tax records to study long-term effects is that the time series is significantly shorter and residents of two regions, the Basque Country and Navarre, are not in the data. I use tax data to show that the main lifetime results are robust when using measures of cumulative earnings derived from uncensored tax earnings.

Employers are represented in the data through their anonymized social security account numbers. For workers in the general regime of social security,[19] each firm has one account for each province in which it employs workers. There are 50 provinces in Spain which are further grouped into 17 autonomous regions. An employer identifier in the data thus represents a firm-province combination. This notion of employer is equivalent to a firm for single-establishment firms, and smaller than a firm—closer to an establishment—for firms operating in several provinces. Firm-province is the employer definition I use throughout the paper.[20] Since this paper focuses on size, and to the extent that large firms are large employers relative to other employers in the provinces in which they operate, using employer or firm size should not make much of a difference, other than compressing the size distribution. A drawback of this employer definition might arise from rare cases in which I assign a small first employer to workers who are in fact matched to a large firm in a province in which it has a small presence. Unfortunately, I do not observe firm size whenever it differs from firm-province.

For each employer I observe its location, sector, age, and number of workers. Number of workers is the measure of employer size I use. The data include a firm identifier which groups together employers belonging to the same firm. While this identifier allows me to identify two sampled employers that belong to the same firm I still use firm-province as the employer unit because employer size is observed at this level. Since I observe a sample of workers, I cannot "aggregate up" from employer size to firm size.

---

[18]This involves grouping worker-month observations into 5,480 cells c {professional category $\times$ age $\times$ quarter} and parametrically model earnings within-cell while imposing no restrictions across cells. I assume log-normality within each cell and estimate the parameters μ_c and σ_c^2 using maximum likelihood. I then use these parameters to simulate earnings observations for bottom- and top-coded observations.

[19]More than 95% of Spanish workers are in the general regime of Social Security (Bonhomme and Hospido, 2017). Certain civil servants and agricultural workers, for instance, are excluded from the general regime.

[20]This definition notwithstanding, I follow convention in related literature and I use the words *firm* and *employer* interchangeably.

In the original MCVL data employer size is only observed starting on 2004. However, I obtained a new data extract recording the evolution of size for the employers in my sample, going back to 1980. This extract allows measuring employer size at any point in time during the sample years of labor market entry, which in this study is key in order to avoid reclassification bias (assigning a large first-employer to a worker who had a small first-employer that grew).[21]

Throughout the paper I use supplementary data sources other than MCVL that I describe in Section B.1 of Appendix B.

3.2 Sample Selection

I use the employment history information to build a monthly panel of employment, earnings, worker characteristics, and employer characteristics. The panel covers 1984 to 2015. I do not use 1980–1983 earnings since they are missing in large proportions. If a worker has more than one employer in a given month, I add up earnings from the different employers while keeping the characteristics of the employer which provides higher earnings that month.

I limit the analysis to Spain-born male workers. The retrospective nature of the data suggest that the earlier years of the panel are not representative for women, who were more likely to enter and then leave the labor force (García Pérez, 2008; Bonhomme and Hospido, 2017). Focusing on those born in Spain makes it more likely that I observe the entire labor market history of workers in my sample. Furthermore, including foreign-born workers is at odds with my empirical strategy that relies on a person's region of birth. Since the lifetime analysis requires me to observe each worker a sufficient number of years, the data impose a tradeoff between how many cohorts I study and how many years I follow each worker. Balancing this tradeoff, I focus attention on the 1968–1980 birth cohorts while they are aged 16–35. I include those who, between labor market entry and age 35, predominantly work as wage earners.[22] These are the restrictions I impose for the panel analysis of Section 6, which result in around 125,000 workers and 16,000,000 worker-month observations.

The data requirements for the cross-sectional long-term analysis of Section 4 are more stringent since each observation aims to capture information about the full labor market history of a given worker. For each person, I require information on his first labor market experience, and enough lifetime information on employment and earnings. Thus, I impose additional restrictions for this part of the analysis that reduce the number of workers in the sample. I include those who, between 16 and 35 years, have sufficient attachment to the formal labor market: those who are employed for half or more of the months since labor market entry up until the year they turn 35. This type of sample selection criteria is present in other studies analyzing lifetime income (Guvenen et al., 2017). I further exclude workers who have their first job in the public sector, have their first job very late (later than age 22

---

[21]The special extract, prepared by MCVL staff, contains employer size back until 1980 for the employers who are the first or second employers of workers in my cross-sectional lifetime analysis sample. For the remaining employers, I observe size starting in 2004.

[22]I exclude those who are self-employed for 40% of the time or more during this period

for high school graduates, 25 for vocational, and 28 for college),[23] or in a Social Security regime different than the general regime. All these restrictions result in a sample of around 80,000 people, 50% of those in the initial raw data. Table 1 provides summary statistics for the resulting sample.

Table 1: Summary Statistics: Career Outcomes Sample

	N	Mean	Sd.	p25	p50	p75
education						
high school	79,941	.43	.50	0	0	1
vocational	79,941	.41	.49	0	0	1
college	79,941	.16	.39	0	0	1
between 16–35 years old						
number of employers	79,941	7.58	5.41	4	6	10
days worked	79,941	4,735	1,008	3,996	4,766	5,495
1st semester in labor market						
age	79,941	20.45	2.87	18	20	23
employers	79,941	1.23	.48	1	1	1
days worked	79,941	147.89	29.20	119	158	176
in region of birth	77,050	.88	.33	1	1	1
unemployment rate	79,941	13.58	5.86	8.89	12.93	16.85
lifetime income						
(cumulative income 16–35)						
0% discounting	79,941	280,745	118,698	198,773	254,142	333,516
3% discounting	79,941	193,194	78,752	138,359	177,426	230,360
lifetime income						
(excluding 1st semester in labor market)						
0% discounting	79,941	271,517	115,737	191,369	245,713	322,886
3% discounting	79,941	185,093	76,149	131,959	169,993	221,175
size of first employer						
first-employer size	79,941	299.94	1,389.22	5	19	94
log first-employer size	79,941	3.18	2.11	1.61	2.94	4.54
1–9 employees	79,941	.37	.48	0	0	1
10–19 employees	79,941	.14	.34	0	0	0
20–49 employees	79,941	.16	.37	0	0	0
50–249 employees	79,941	.19	.39	0	0	0
250+ employees	79,941	.15	.36	0	0	0

Notes: Summary statistics for cross-sectional lifetime analysis sample of Section 4. Includes Spain-born male workers born between 1968–1980 when they are between ages 16–35, who are predominantly wage earners in this period, who work for at least half the months since their first job until age 35, have their first job in the private sector, and do not enter their first job very late (i.e. over 22 for high school graduates, 25 for vocational, 28 for college). First labor market semester is defined as the first six continuous months after predicted graduation a person works for 100 days or more. Lifetime income is the sum of all monthly income (earnings and unemployment benefits) since the year a worker turns 16 until the year he turns 35. Lifetime income excluding 1st semester in the labor market only counts income starting after the first labor market semester. Income expressed in constant 2015 Euro.

3.3 Definitions and Measurement

First labor market experience. I define a worker's first labor market experience as the first six continuous months after predicted graduation that a person works for 100 days or more. This definition aims to capture the first relevant job after finishing formal education, while avoiding summer work or very temporary employment.[24]

---

[23]Those for whom I observe a late (relative to their education) first job in the data likely held their first job in informal employment or outside Spain.

[24]Panel (a) in Appendix Figure A7 plots the distribution of first labor market experience calendar years. Workers in my sample entered the workforce during the late 1980s, 1990s, and early 2000s. Panels (b) and (c) of

First-employer size. For each worker I assign as his first-employer size the size of his employer during his first labor market experience. In robustness checks I also use average size during the four years prior to the worker joining the firm. For the 20% of workers who have more than one employer during this semester, I assign the maximum size across these employers.

Lifetime income. I use measures of lifetime income as worker-level long-term outcomes in Section 4. These are meant to capture the whole stream of labor income a worker receives between labor market entry and some age T. I include as labor income both earnings and unemployment benefits. The lifetime income measure takes the following form:

$$Y_i = \sum_{t=16m1}^{Tm12} \frac{w_{it} + u_{it}}{(1+\delta)^{t-1}}. \tag{1}$$

Where w_{it} are monthly earnings, u_{it} are monthly unemployment benefits, and δ is a discount rate. I set $\delta = 0$ in the main analyses, but I show that the main results are robust to other commonly used annual discount rates.

The data impose a tradeoff between how many cohorts can be analyzed and how high is age T set. I set T=35 and analyze thirteen birth cohorts (1968–1980). While setting the top age at age 35 excludes several years of the working life, this is a meaningful measure since i) it captures a large amount of the working life (15 years on average in my sample), ii) it captures the fraction that is less time-discounted from the perspective of someone entering the labor market, and iii) reaches up until the mid 30s where incomes stabilize and trajectories are more easily predictable (past evidence indicates that the majority of lifetime wage growth occurs during the first 10 years of work (e.g. Topel and Ward, 1992; Rubinstein and Weiss, 2006); see Appendix Figure A4 for evidence for Spain on income profiles stabilizing during the mid 30s). Table 1 provides summary statistics for this measure. The median is 254,142 Euro (2015).[25]

Measures such as equation (1) are attractive for several reasons. First, they are conceptually relevant, reminiscent of utility expressions in life-cycle models. Second, they tone down business-cycle or transitory idiosyncratic shocks to income that might induce noise in workers' incomes at a single time period. Third, they naturally accommodate different income growth paths across education levels or occupations. And fourth, they account for non-employment spells and unemployment benefits in a natural way, bypassing traditional issues of self-selection into employment at a given time period. If the treatment of interest impacts employment outcomes at some point, not accounting for these periods could bias causal estimates. Accommodating these periods into the lifetime income measure (adding zeroes or unemployment benefits) deals with this issue.

Appendix Figure A7 plot the distributions of age and days worked during the first labor market experience.

[25]In order to study the long-term consequences of a worker's first job, the lifetime income variable in the analysis below nets out income earned before and during the first labor market semester (as defined above). Summary statistics for this variable are also included in Table 1. Its median is equal to 245,713 Euro (2015).

3.4 Large and small firms in Spain

Compared to other countries, Spain has few large firms. According to OECD data from 2013, 0.4% of Spanish enterprises have 250 employees or more. This percentage, while comparable to that from Portugal or Italy, is far below Germany (around 2%) or the U.S. (around 1.5%; see Appendix Figure A2). Some argue that size-dependent policies and regulations are partly responsible for this "distortion" in the firm-size distribution (IMF, 2015; Guner et al., 2007).

In relation to other contexts, thus, few young workers will be employed at large firms which, the literature suggests, tend to offer more desirable jobs (Sorkin, 2018). Firm attributes associated with a large size will likely be the same in Spain and other countries. However, compared with Germany or the U.S., the outside option of a young Spaniard who does not match with a large employer might disproportionately be a very small and possibly unproductive firm. In 2013, 16% of Spanish manufacturing workers were employed in a business with nine employees or less. This number was 5% for Germany and for the U.S. (see Appendix Figure A2). In my sample, 37% of workers hold their first job at an employer with less than 10 employees while 15% do so at a large employer with more than 250 employees.

4 Size of First Employer and Career Outcomes

This section lays out the relationship between the size of a worker's first employer and long-term career outcomes. After documenting some descriptive facts, I discuss the IV approach that accounts for endogenous sorting of workers and firms. The thought experiment I wish to capture is random assignment of young workers to be hired by firms of different sizes, with other firm attributes associated with size forming part of this thought experiment. I do not capture a hypothetical exogenous increase in the size of a given firm. Larger firms are characterized by attributes that could be relevant for young workers, likely driving any first-employer size effect.[26]

4.1 Descriptive Facts

There is an unconditional positive relationship between the size of a worker's first employer and his long-term career outcomes. Figure 1 above shows the unconditional relationship between the lifetime income measure and first-employer size. There is a strong positive relationship between the two variables which is linear in logs. The correlation coefficient is equal to 0.21.[27]

---

[26]Appendix E discusses what these attributes and mechanisms associated with firm size might be. Underlying my empirical approach is a presumption that any heterogeneity firms might display in how they impact their young workers' long-term outcomes can be ranked according to a scalar measure. As Appendix E lays out, there are reasons to believe size could be a good proxy for such a scalar measure (e.g. employer-provided training, higher productivity, use of new technologies).

[27]This relationship is not explained away by firms' sector of activity. Appendix Figure A5 shows that the pattern from Figure 1 holds within sectors (adjusting for 58 sector fixed effects).

I also provide evidence on the earnings and employment trajectories underlying the lifetime income measure. Figure 2 groups workers into five different groups based on the size of their first employer and plots the evolution of average quarterly earnings since labor market entry for each of these groups. First-employer size is a good unconditional predictor of subsequent earnings paths: the earnings profiles for these groups never cross. Similar patterns arise when looking at employment and daily wages (Appendix Figure A6).

Figure 2: Quarterly income trajectories by first-employer size
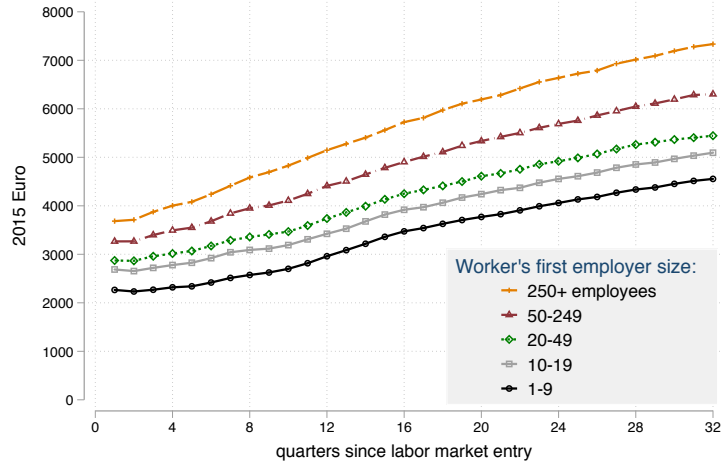
Notes: Evolution of average quarterly income since labor market entry, categorizing workers based on the size of their first employer. Sample of male workers of all education levels, born in Spain between 1968–1980.

4.2 Empirical Approach: Estimating Equation and IV

The goal is to estimate the elasticity of a worker's lifetime income with respect to the size of his first employer. This elasticity is given by β in the baseline estimating equation:

$$y_i = \beta s_{J(i)} + f(u_{r,t_0(e,c)}) + \delta_r + \delta_e + \delta_c + \varepsilon_i. \tag{2}$$

Where y_i is (log) lifetime income for worker i and $s_{J(i)}$ is the (log) number of employees of firm J where i held his first job.[28] c indexes birth cohorts, e refers to three educational attainment levels—high school, vocational, college—and r indexes regions of birth. $t_0(e,c)$ indexes a worker's *predicted* graduation year, which is a function of birth year c, and educational attainment e. Based on standard Spanish completion times, I assign year of predicted graduation as the year in which people with high school degrees turn 17, 20 for vocational education, and 23 for college education. The δs represent region of birth, education, and cohort fixed effects, while $f(u_{r,t_0(e,c)})$ is a flexible function of the unemployment rate in region r in year $t_0(e,c)$, capturing business cycle variation. All variation in equation (2) is cross-sectional since each worker only has one first job and one level of lifetime income.

---

[28]First-employer size is measured at the year the worker joined the firm. The main results are robust to alternative size measures.

OLS estimates of β are likely biased due to unobserved determinants of lifetime income that are plausibly correlated with first-employer size. For instance, large firms might be able to hire young workers who are more productive and would earn higher wages throughout their career no matter what. Similarly, young people who are able to match with a large firm might be more proactive in their job search strategies, a skill that can lend returns throughout the working life. These and related reasons are the motivation for my IV strategy.[29]

The IV approach uses variation in the *composition* of regional labor demand for inexperienced workers across time and space. It relies on the notion that the idiosyncratic hiring shocks of a *small* number of *large* employers can impact regional labor demand composition.[30] Expansions or openings of new operations will make large firms hire batches of inexperienced workers differentially across years. Depending on when and where a young worker enters the labor market, he will be exposed to different propensities to join larger or smaller firms.

A simplified example based on a true event illustrates the intuition behind the IV. Consider two high school graduates who were both born in the Spanish region of Asturias, one year apart from each other. The graduation year of the younger person is 1993 and coincides with the opening in the region of a large and modern plant of the U.S. multinational DuPont, which hires around 1,000 workers. The older worker's high school graduation was in 1992, one year earlier. This timeline suggests that the worker from '93 will be more likely to have his first job at DuPont than the worker from '92. Similarly, given low mobility across regions, a worker from '93 born in the neighboring region of Galicia will also be relatively less likely to start at DuPont than the '93 worker from Asturias.

The goal of the IV is to aggregate and summarize this type of large-firm hiring shocks across years and regions in my sample. Ideally, the DuPont example would be just one of many large-firm labor-demand shocks. Fortunately, the institutional and historical context provides a setting of rich variation. During the sample years of labor market entry (1985–2003), Spain was undergoing a period of economic transformation following adhesion to the European Union in 1986 (Chislett, 2002). This period was characterized by an internationalization of the economy: an increased openness to trade, and growth of foreign firms' investments in the country. It also featured reforms towards market liberalization, and large investments in regional infrastructures. This context led to great dynamism in large firms opening and expanding across the country, contributing to the variation that the IV approach leverages. Figure 3 illustrates this trend using register data on the population of establishments. For each region, the figure shows the number of establishments with 500+ employees in 1994 and in 2003.[31] In 15 out of 17 regions the number of large employers increased, and in most of them substantially so. This pattern holds even for regions that had initially fewer large firms.

I construct an index that captures the variation in labor-demand composition a given

---

[29]See Eeckhout and Kircher (2018) for a theoretical model on worker-firm sorting and firm size.

[30]Gabaix (2011) argues shocks to a small number of large firms can generate business cycles. My IV approach uses variation in large/small firm labor-demand composition while holding constant cyclical variation.

[31]The years of labor market entry in my sample are 1985–2003 but establishment census data starts being recorded in 1994. 500+ is the largest size category for which publicly available information is available.

Figure 3: A period of large-firm dynamism across Spanish regions 1994–2003

(a) 500+ establishments, 1994 and 2003

(b) 500+ establishments in 2003 (normalized 1994=1)
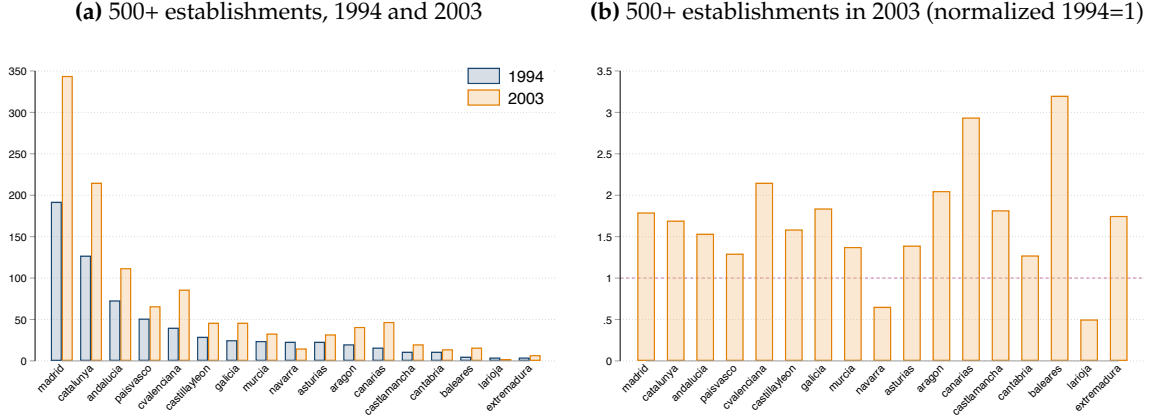
Notes: Source is the Central Business Register (*Directorio Central de Empresas*, or DIRCE). Evolution of the number of establishments with 500+ workers, across the 17 Spanish regions, during the second half of the sample period of labor market entry. Sample period of labor market entry is 1985–2003, DIRCE data only go as far back as 1994.

worker is exposed to when first entering the labor market. Critically, this should be done avoiding endogenous first-job search responses to large-firm hiring shocks. In particular, rather than assigning this index to workers based on the region and year in which they hold their first job, I assign this index to each worker based on her region of *birth* and *predicted* graduation year. This index will work as an IV, being used to predict the size of a worker's first employer.

In practice, I use the information on young workers' hires and their employers observed in the social security data to construct the IV. Let the IV for worker i be denoted by $\bar{s}_{-i}^{rec}$. In order to capture the labor demand composition worker i faces, $\bar{s}_{-i}^{rec}$ is equal to the (log) average first-employer size of i's "relevant peers": workers who have the same educational attainment as i, who are entering their first job in i's region of birth, and are doing so during i's predicted graduation year.[32] To be more precise, consider a worker i with education e_i, region of birth r_i, birth cohort c_i, predicted graduation year $t_0(e_i, c_i)$, and year of first job t_i. Also, let $\tilde{r}_i$ be the region where his first job is located. Subscript $l = 1, ..., N$ indexes workers in my sample and $\mathbb{1}\{\cdot\}$ is the indicator function. The IV approach predicts worker i's (log) first employer size, $s_{J(i)}$, with

$$\bar{s}_{-i}^{rec} = \ln\left(\frac{\sum_{l \neq i} exp(s_{J(l)}) \cdot \mathbb{1}\{\tilde{r}_l = r_i, e_l = e_i, t_l = t_0(e_i, c_i)\}}{\sum_{l \neq i} \mathbb{1}\{\tilde{r}_l = r_i, e_l = e_i, t_l = t_0(e_i, c_i)\}}\right). \tag{3}$$

4.3 IV Discussion

IV variation and the business cycle

While previous literature has been interested in the effects of cyclical conditions at the beginning of the working life (e.g. Kahn, 2010; Oreopoulos et al., 2012), I explicitly aim

---

[32]I follow a leave-one-out approach. That is, if individual i got his first job in his predicted graduation year and in his region of birth, I exclude his employer's firm size from the average $\bar{s}_{-i}^{rec}$.

to partial out this channel from the effect of starting out at a larger or smaller employer. That is, the business cycle is a potential confounder of the first-employer size effect since it could impact both the size of a worker's first employer (Moscarini and Postel-Vinay, 2012) and also lifetime income through other channels. However, the empirical approach summarized in equation (2) is aimed at shutting down any impacts that business cycle conditions at entry might have on long-run prospects: by including region and cohort fixed effects, and explicitly controlling for a flexible function of the unemployment rate a worker faces during labor market entry. By flexibly controlling for the unemployment rate at the time of labor market entry, I try to replicate the thought experiment of comparing workers who were randomly assigned to firms of different sizes but shared common business cycle conditions.

Figure 4 illustrates how cyclical conditions are held constant and the residual variation the IV approach uses. Panel (a) plots the correlation between the unemployment rate at entry and i) the IV $\bar{s}_{-i}^{rec}$ (*raw IV*), and ii) residuals from a regression of $\bar{s}_{-i}^{rec}$ on δ_r, δ_e, δ_c, and $f(u_{r,t_0(e,c)})$ (*residualized IV*). As expected, there is a negative correlation between the instrument and the unemployment rate during labor market entry (blue diamonds).[33] After controlling for fixed effects and a flexible function of the unemployment rate (orange circles), the remaining variation arises from the deviations from within-region, within-cohort, and within-education averages in workers' first-employer size that is orthogonal to unemployment rate fluctuations. This residual variation is—mechanically—unrelated to the unemployment rate, and meant to capture the changes in labor demand composition arising from large firms' idiosyncratic hiring shocks.[34] One could worry that using a single indicator such as the unemployment rate might not perfectly capture cyclical variation. Panel (b) on Figure 4 allays these concerns using data on regional GDP growth rates. This additional cyclical indicator, since it is excluded from the specification in equation (2), is not mechanically unrelated to the IV residual variation. Reassuringly, a similar pattern emerges. There is a positive correlation between the instrument and regional GDP growth (blue diamonds). This is, again, consistent with, unconditionally, large-firm hiring being more prevalent during good economic times. Controlling for fixed effects and the unemployment rate (orange circles) results in IV residual variation having a flat relationship with regional GDP growth.[35] Overall, the residual IV variation identifying β seems orthogonal to business-cycle conditions.[36]

---

[33]The fact that young workers who enter the labor market during a recession start a smaller firms has been documented for Canadian college-educated workers in Oreopoulos et al. (2012) and for Austrian non-college workers in Brunner and Kuhn (2014).

[34]Appendix Figure A8 makes a similar point focusing on the time series variation of a given region (one of the largest, Catalunya). It plots the unemployment rate in Catalunya, together with the time series variation in $\bar{s}_{-i}^{rec}$, and the residual variation after netting out fixed effects and unemployment. The blue dashed line represents the movements in labor demand composition that my empirical approach relies on.

[35]Appendix Table A1 shows the regression results underlying Figure 4.

[36]As a robustness check, I estimate versions of equation (2) that control for the regional unemployment rate *and* regional GDP growth.

Figure 4: IV residual variation uncorrelated with business-cycle variation

(a) Regional unemployment rate

(b) Regional GDP growth
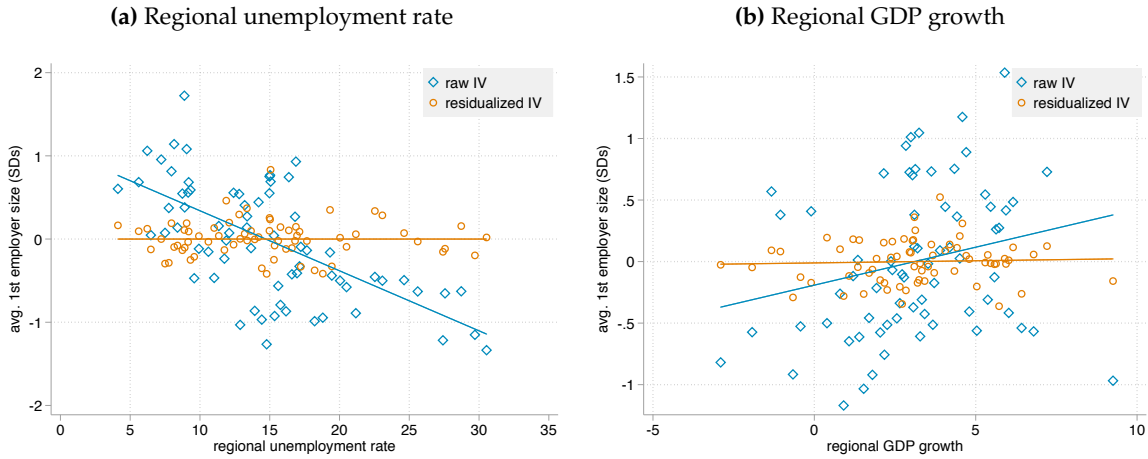
Notes: Binned scatterplots of the IV $\bar{s}^{rec}_{-i}$ described in the text (*raw IV*, blue diamonds) and residuals from a regression of $\bar{s}^{rec}_{-i}$ on region of birth, education, and cohort fixed effects and a flexible function of the regional unemployment rate at the worker's region of birth in his predicted graduation year (*residualized IV*, orange circles). Panel (a): Plotted against the regional unemployment rate at the worker's region of birth in his predicted graduation year. Panel (b): plotted against the regional GDP growth rate at the worker's region of birth in his predicted graduation year. Regression estimates in Appendix Table A1.

Instrument exclusion assumption

The instrument varies at the {region of birth $\times$ educational attainment $\times$ birth cohort}-level (except for the leave-one-out component), and follows the structure of the Bartik approach discussed by Goldsmith-Pinkham et al. (2018) and Borusyak et al. (2018). In my setting, identification is connected to workers' assignment to one of each {rec} cells (conditional on controls). The IV exclusion assumption relies on the absence of an unobservable {rec} component that impacts lifetime outcomes y_i and is correlated with the large-firm labor demand shocks the IV captures. Threats to identification fall under the umbrella of labor supply shocks at the cohort $\times$ region of birth $\times$ educational degree level.

What would constitute a violation of the IV exclusion assumption? Take the DuPont example from above and consider Asturian high school seniors in 1993 who would have gone to college in the absence of the DuPont shock. However, the arrival of the firm induces some to put an end to their formal education in order to try to get a DuPont job. Then, the 1993 Asturias high school cohort would be endogenously composed of more able young people (since in the absence of DuPont they would have attended college) *and* more likely to have a large first employer. This scenario would represent a violation of the exclusion assumption. Below I discuss in more detail the use of educational attainment in my empirical approach, and I find evidence that lessens this type of concern: the data do not support the hypothesis of young people changing their educational decisions in response to the variation captured in my IV.

IV and household characteristics at age 17

I use supplementary survey data to test for the plausibility of the exclusion assumption. In particular, I show that the IV is not correlated with $\{rec\}$-level observable characteristics at age 17. These characteristics include parents' employment and type of job, parents' education, or household income. The lack of correlation with these observable characteristics should diminish concerns about potential correlations with unobservable $\{rec\}$ characteristics. This test also allays concerns related to potentially endogenous large firms' decisions of when and where to expand based on unobserved cohort characteristics. I describe the test and show its results in Section B.2 of Appendix B.

Educational attainment and potential endogenous responses

I control for educational attainment and use it in the construction and assignment to workers of the instrument $\bar{s}_{-i}^{rec}$, making the labor demand predictor specific to each education group. A reasonable worry is that educational attainment could be endogenous in this setting, as opposed to predetermined like region and year of birth. This type of concern warrants consideration based on evidence on the countercyclicality of education enrollment decisions (e.g. Card and Lemieux, 2001; Petrongolo and San Segundo, 2002; Sievertsen, 2016).

Certain features of my empirical approach somewhat relax these worries. As described above, business cycle conditions are kept constant. Given this approach, the educational response that would be worrying would come from responses to the large-firm hiring shocks captured by the instrumental variable $\bar{s}_{-i}^{rec}$, while holding business cycle conditions constant.[37]

I test for endogenous education responses asking whether, after controlling for unemployment rates, regional labor demand composition influences education investment decisions. Section B.3 in Appendix B describes this test and its results. The key takeaway is that, reassuringly, there is no detectable correlation between the IV residual variation and education choices. Thus, I fail to reject the null hypothesis that, conditional on cyclical conditions, large-firm hiring shocks do not induce endogenous educational responses.

Autocorrelation of the instrument and persistent regional spillovers

One might worry that large-firm shocks might persistently change the economic landscape of a region through spillovers (Greenstone et al., 2010) and thus impact workers' lifetime outcomes through ways other than first-employer characteristics. In part, my empirical design allays these concerns thanks to (i) controlling for cyclical conditions, and (ii) other cohorts from a given region acting as controls. For instance, if DuPont changes general economic opportunities in Asturias after their arrival in 1993, the '92 and '94 cohorts would also enjoy these spillover effects and act as controls for the '93 cohort.

---

[37] Also note that an education enrollment response that is not followed by completion of the higher degree level would not be problematic for the exclusion restriction, it would simply reduce the relevance (predictive power) of the IV approach.

These types of worries would be more pressing if the large-firm hiring shocks that the IV leverages were very persistent. The nature of the IV approach is to capture idiosyncratic hiring of large employers that are not sustained over time (such as plant openings, expansions, or hiring in batches). In line with this, a low autocorrelation of the IV residual variation would be desirable. Collapsing the data at the $\{rec\}$-level, the residual variation of the IV features an estimated autocorrelation equal to 0.15.[38] This is a positive but low autocorrelation. The fact that it is small is reassuring. It being positive could be expected. For example, a new plant opening could see its hiring process expand over two calendar years.[39]

4.4 Lifetime Income: Results

Table 2 shows OLS, first stage, and IV-TSLS results of estimating β in equation (2) using the proposed instrumental-variables approach.[40] Throughout, I control for $u_{r,t_0(e,c)}$, unemployment rate in the region of birth at the year of predicted graduation, by fitting a separate quartic of $u_{r,t_0(e,c)}$ for each different educational attainment level. I cluster standard errors at the $\{$region of birth $\times$ educational attainment $\times$ birth cohort$\}$-level since this is the level through which the IV operates (Abadie et al., 2017). Column (5) shows first-stage results. The instrument does a good job at predicting first job size, with an excluded instrument F-statistic of 24.3.[41] Columns (1) and (6) show, respectively, the OLS and IV elasticities of lifetime income with respect to first-employer size. The OLS elasticity estimate is .028. The IV-TSLS estimate is significantly larger and equal to .117. This elasticity implies that, at least for the relevant compliers, matching with a 10% larger first-employer leads to 1.17% higher lifetime income.

A way of interpreting the estimated magnitude is using the standard deviation of log first employer size, which is equal to 2.1. We can expect that matching with a first-employer that is larger by one standard deviation in log size to increase lifetime income by 27.7%.[42]

4.5 Comparison of OLS and IV results

The IV estimate is about four times larger than the OLS. This is consistent with the first-employer size effect being heterogeneous across workers, and it being larger for those whose first-employer match is more susceptible to the labor demand IV. That is, suppose that some people benefit more than others from having their first job at a larger firm. Suppose as well that those who benefit the most tend to get a first job at a large firm if there is idiosyncratically high large-firm hiring in their birth region, but not otherwise. Then, a

---

[38] 0.15 is the coefficient of a regression of the cell-level residualized IV on its one-year lag and a constant (N=610, robust standard error=0.049).

[39] Davis and Samaniego de la Parra (2017) estimate that the mean duration from date of first posting of a single vacancy to start of employment is 57.3 days.

[40] Appendix Table A7 displays estimates of the reduced-form.

[41] Appendix Figure A9 displays a histogram with the IV residual variation. Appendix Figure A10 provides graphical evidence of the IV result, showing the variation that identifies β through the first and second stage of TSLS.

[42] From $100 \cdot [exp(2.1 \times .1166) - 1] = 27.74$

Table 2: Career outcomes and first-employer size: OLS and IV-TSLS estimates

	OLS				First Stage	IV-TSLS			
	lifetime income (1)	lifetime earnings (2)	average daily wage (3)	days worked (4)	first employer size (5)	lifetime income (6)	lifetime earnings (7)	average daily wage (8)	days worked (9)
first employer size	0.0276*** (0.0011)	0.0280*** (0.0011)	0.0311*** (0.0009)	-0.0031*** (0.0005)		0.1166** (0.0481)	0.1102** (0.0493)	0.0820** (0.0402)	0.0281 (0.0189)
labor demand instrument					0.0953*** (0.0193)				
F-stat excl. instr.					24.31				
SE Clusters	661	661	661	661	661	661	661	661	661
Observations	79941	79941	79941	79941	79941	79941	79941	79941	79941

Notes: OLS and IV-TSLS estimates of the elasticity of different long-term outcomes with respect to first-employer size. Columns (1)-(4) show OLS estimates. Column (5) shows the first stage. Columns (6)-(9) show IV-TSLS estimates, instrumenting for first-employer size using the labor-demand composition index defined in the text. Columns (1) and (6): Lifetime income defined as sum of total labor income (wages and unemployment benefits) after first job semester (defined in text) until age 35. Columns (2) and (7): Lifetime earnings defined as sum of total earnings after first job semester (defined in text) until age 35. Columns (3) and (8): Average daily wage defined as sum of total income over total days worked after first job semester (defined in text) until age 35. Columns (4) and (9): Total days worked after first job semester (defined in text) until age 35. All variables are in logs. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Standard errors clustered at the level of region of birth $\times$ education $\times$ birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

LATE interpretation (Imbens and Angrist, 1994) would explain the relatively-high IV magnitude. I now provide evidence consistent with this scenario.[43]

The first thing to ask is who, given the nature of the IV, are the likely compliers. First, note that the geographic dimension of the instrument works through region of *birth*. The minority of people who migrate across regions for their first job will be less likely to be compliers.[44] More generally, highly motivated individuals with clear career goals will be more likely to do their best to match with their preferred type of employer under all scenarios of labor demand composition. Compliers, those who only match with large firms in years of differentially high large-firm hiring, might thus be of initially lower ability. This could arise as a supply-side effect if lower ability young adults take a more passive approach towards job search. It could also be a demand-side effect if large firms are able to screen job applicants and hire in order of perceived ability. In both cases, the marginal large-firm hire will be the less able, and the average new hire in times of expansion (captured by the IV) will be less able than the average new hire in normal times.

The above discussion suggests that the less able or less knowledgeable young workers comprise the group of IV compliers, a possibility I explore and confirm more formally. Building on results by Angrist and Imbens (1995), I estimate a flexible first stage that permits characterizing which parts of the firm-size distribution and which type of workers are driving the IV two-stage least square (TSLS) estimate. The intuition behind the Angrist and Imbens (1995) result is that, with a multivalued treatment and a multivalued instrument, the TSLS estimate can be written as a weighted average of causal responses to a unit change in treatment along the treatment and instrument distributions for the relevant compliers. I develop an approach to estimate such weights, allow them to vary across different groups of people. I find these weights are consistently larger for those without a college degree and those born in less urban parts of Spain (see Figure 5; Appendix C derives the conceptual and estimation details).[45] Figure 5 indicates that the instrument has a greater impact in first employer size for those born in rural places and for those without a college education, especially so when shifting workers away from the bottom of the first employer size distribution.

Overall, the evidence suggests that the less able and less knowledgeable are more influenced by the labor demand variation my IV uses. Based on this result, a channel that would explain the larger magnitude of the IV estimates with respect to OLS is if younger and less knowledgeable workers have the highest long-term benefits of a larger first employer (due to, for example, worse outside options, having a harder time moving away from a bad first

---

[43]Unobserved ability bias, by which more productive workers match with larger firms, would by itself bring *down* IV estimates with respect to OLS. The current comparison does not mean that this form of positive sorting does not exist. Rather, it seems to suggest that heterogeneous effects and the LATE explanation I lay down trumps unobserved ability bias.

[44]Labor force survey data (EPA) from 1992–2015 show that 80% of employed persons lived in their region of birth. Table 1 shows that, in my sample, 87% of young workers held their first job in their region of birth.

[45]I classify workers as rural- or urban-born based on their province of birth and using data from Goerlich Gisbert and Cantarino Martí (2015). I classify as rural provinces those with over 15% of its population being rural. This number is around the population-weighted median across provinces in the original data, and close to the median in my sample.

Figure 5: Compliers are less urban and less educated: weight function from flexible 1st stage
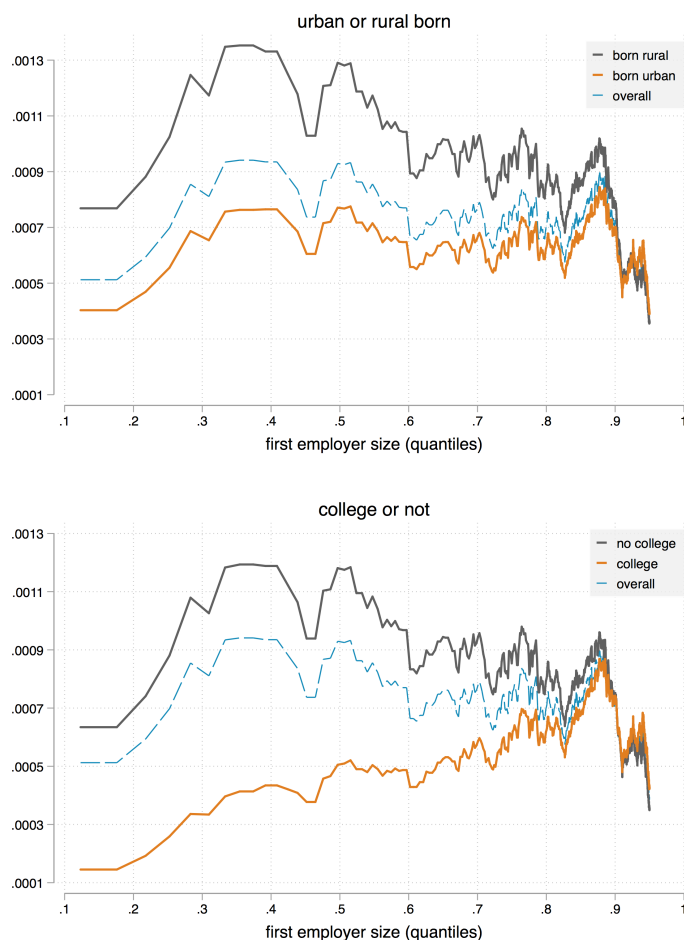
Notes: This figure plots the estimated differential impact of the instrumental variable on workers' first employer size, heterogeneously by workers' place of birth and education, and for different quantiles of the first employer size distribution. Estimates of the weight function from equation (C12) in Appendix C, for different subgroups and overall, as a function of first employer size, and holding the instrument constant in the 95th percentile.

job, or benefiting the most from large firms' on-the-job skill development opportunities).[46]

The LATE aspect of the results should be kept in mind when interpreting the IV estimates. However, even if the estimates of the first-employer size effect are not representative for all workers, I seem to be capturing the causal effect for younger workers with lower earnings potential (less educated and from less urban areas) who might be of special policy interest.

4.6 Lifetime Income: Robustness and Extensions

Robustness. The IV elasticity of lifetime income with respect to first-employer size is robust to several alternative specifications. Section B.4 of Appendix B shows that the results are stable when discounting the stream of income in lifetime income measures, when con-

---

[46]The evidence from Bonhomme et al. (2019), indicating that "lower-type" workers gain the most from employment at a "higher-type" firm, is consistent with this idea.

trolling for regional unemployment differently, when controlling for regional GDP growth, when controlling for unemployment rates in previous years or during the year of labor market entry in addition to unemployment during predicted graduation, when measuring employer size as an average over years prior to labor market entry, when controlling for first-employer sector, when using birth-province fixed effects, when controlling for population size in the first-employer's province, and when including region-specific time trends. Section B.5 shows that the first-employer size effect is robust to using uncensored measures of income constructed from 2005–15 tax records.

Varying elasticity across firm sizes. In Section B.6 of Appendix B I relax the constant elasticity assumption implicit in equation (2). The reason for doing so is the possibility that increments in first-employer size are differentially valuable at different parts of the employer size distribution. Using a control function approach I estimate an expanded version of equation (2) that allows a quartic polynomial of log first-employer size. The results display an intuitive and interesting non-linearity. The highest point of the first-employer size effect can be found around the 80th percentile of the empirical first-employer size distribution. By contrast, the elasticity is small and non-significant in both extremes of the distribution. This pattern is consistent with the returns to first-employer size mostly arising from the difference between joining or not one of the, relatively few, large employers. Conditional on starting out at a very small or a very large firm, differences in size do not seem to matter that much.

4.7 Wages, Employment, and Earnings

The lifetime income effect of matching with a larger first employer could combine effects on different margins: quantity of work, average wages, and unemployment benefits. Here, I decompose the lifetime income effect into its different components.

I estimate the elasticity β from equation (2) replacing lifetime income y_i with three different outcomes (in logs): average daily wages, total days worked, and lifetime earnings (which differ from lifetime income in that they do not count unemployment benefits). Table 2 shows OLS, first stage, and IV results from this exercise. Focusing on the IV estimates, the first thing to note in column (7) is that the elasticity of lifetime earnings is equal to .110, which is 94% of the elasticity of lifetime income equal to .117. Further, the elasticity of average daily wages is equal to .082, and the elasticity of total days worked is .028.

Taken together, these results imply that 94% of the lifetime income result come from increased earnings as opposed to unemployment benefits. Further, the increase in earnings can be attributable both to average daily wages (74%), and the amount of days worked (26%).[47]

---

[47]Section B.7 of Appendix B studies the effect of first-employer size on later job security.

4.8 How much of the *graduating-in-a-recession* effect can be explained by the first-employer size effect?

I quantify the relationship between the effect of starting one's career at a larger or smaller firm and the effect of entering the labor market during a recession, which has been the focus of previous work (e.g. Oyer, 2006; Kahn, 2010; Oreopoulos et al., 2012; Brunner and Kuhn, 2014; Altonji et al., 2016; Fernández-Kranz and Rodríguez-Planas, 2018; Schwandt and Von Wachter, 2019). The persistent positive effects of starting at a large firm could partly explain the findings of this literature (see Moscarini and Postel-Vinay, 2012; Oreopoulos et al., 2012; Brunner and Kuhn, 2014, for previous evidence on cyclical conditions and size of hiring firms).

I begin by estimating the following regression in my sample:

$$s_{J(i)} = \gamma u_{r,t_0(e,c)} + \delta_r + \delta_e + \delta_c + \varepsilon_i. \tag{4}$$

Where $s_{J(i)}$ is the (log) number of employees of employer J where worker i held her first job, and $u_{r,t_0(e,c)}$ is the unemployment rate in worker i's region of birth r during her predicted graduation year $t_0(e,c)$. The δs are region of birth, education, and birth cohort fixed effects. γ is the parameter of interest, representing the semi-elasticity between the size of a worker's first employer and the prevailing unemployment rate at the time of labor market entry.

Table 3: First-employer size and unemployment rate at entry

	first employer size		
	(1)	(2)	(3)
unemployment rate at entry	-0.0099***	-0.0117***	-0.0166***
	(0.0035)	(0.0044)	(0.0050)
SE Clusters	661	442	364
Sample	All	HS & Voc.	Less urban HS & Voc.
Observations	79941	66998	29724

Notes: OLS estimates of the semi-elasticity of first-employer size with respect to the unemployment rate during labor market entry. First-employer size in logs. Regressions at the worker level. All regressions control for region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Column (1) uses the whole sample. Column (2) uses the subsample of those without a college degree. Column (3) includes non-college workers who were born in the less urban provinces of Spain. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Table 3 shows OLS estimates of γ in equation (4) for the whole sample and different subgroups. The negative estimates are consistent with previous literature and the evidence found on Figure 4. The estimate for the full sample in column (1), equal to -.0099, is very similar to that found in Oreopoulos et al. (2012). In column (2) I estimate γ for the subsample of workers without a college degree. The estimated coefficient is equal to -.0117, which is somewhat larger than for the whole sample. This suggests that for this group of less educated workers the size of their first employer is more sensitive to the cyclical conditions at the time of entry. Column (3) focuses on the subgroup of non-college workers who were born in less urban provinces of Spain. As discussed in Section 4.5, this group of workers are likely to be compliers in my IV approach and thus mostly driving the first-employer size

causal effects. The estimate for this subgroup is even larger, equal to -.0166.

Next, I combine the estimates of γ with (i) the elasticity between lifetime income and first-employer size (equal to .117), and (ii) results from Fernández-Kranz and Rodríguez-Planas (2018), who estimate the *graduating-in-a-recession* effect in Spain. Using these results, Table 4 shows that between 7% and 15% of the losses from entering the labor market during a recession could be explained by the fact that during bad economic times young people are more likely to match with a smaller first employer. For non-college workers from less urban provinces, this fraction is between 12% and 15%.

Table 4: Benchmark: First-employer size effect explaining entering-in-a-recession effect

(1) Sample	(2) $\hat{\gamma}$	(3) $\hat{\gamma} \times 8$ (8 ppt = typical recession)	(4) % change in first-employer size	(5) % change in lifetime income ((4) × .117)	(6) % recession effect explained by size effect
All	-0.0099	-0.0792	-7.61 %	-0.89 %	7.12 - 13.91 %
HS & Voc.	-0.0117	-0.0936	-8.94 %	-1.05 %	8.40 - 10.94 %
Less urban HS & Voc.	-0.0166	-0.1328	-12.44 %	-1.46 %	11.68 - 15.21 %

Notes: Percentage of the effect of entering during a recession (Fernández-Kranz and Rodríguez-Planas, 2018) explained by the first-employer size effect for different subsamples. Column (2) reports the semielasticity between first-employer size and unemployment rate at entry (see equation (4) and Table 3). Column (3) shows the effect of a typical Spanish recession (increase in unemployment rate of 8%). Column (4) applies the formula $100 \cdot (\exp(x) - 1)$ to column (3) to display the percentage change in first-employer size associated with a typical recession. Column (5) maps the change in first-employer size into a change in lifetime income using the elasticity estimate of .117 from Table 2. Column (6) shows the losses in column (5) as a fraction of the losses from entering during a recession estimated in Fernández-Kranz and Rodríguez-Planas (2018), who report losses of 9.6%, 12.5%, and 6.4% for high school, vocational, and college workers respectively. For the whole sample (first row) I bound the fraction of the recession effect explained by the size effect using their vocational and college losses of 12.5% and 6.4%. For the high school and vocational workers (rows 2 and 3), I use as benchmark their vocational and high school losses of 12.5% and 9.6%.

5   Persistence and Mechanisms

In this section I discern potential mechanisms that could explain the first-employer size effect. I first show these effects are persistent—that is, not solely mediated by the time a worker spends at his first job. This persistence implies workers' trajectories in subsequent jobs are affected by the nature of their first employer. I then show evidence consistent with two channels that can generate persistent effects: a job ladder channel predicted by models of frictional search, and a human capital channel (heterogeneous skill development at heterogeneous employers).[48]

---

[48] An inherent caveat in the persistence analysis is lack of exogenous variation in the employment dynamics following a first-employer match. Ham and LaLonde (1996) discuss the issues arising when researchers have at their disposal exogenous variation in some initial treatment or intervention, but no exogenous variation driving the employment dynamics that arise afterward. For example, heterogeneous effects by tenure at the first job could arise *because of* different first-job tenures or, alternatively, due to a correlation between workers' propensity to stay on the job and heterogeneous treatment effects. Thus, analyses conditioning on employment patterns (e.g. time spent at first employer, number of jobs, unemployment spells) will be more descriptive in nature than the lifetime analyses from the previous section.

5.1 Persistence

Descriptive evidence

Descriptive patterns in the data are consistent with persistent effects. First, young workers are very mobile. Figure 6 shows that most workers do not stay at their first employer for a prolonged period of time. Around 50% of workers are at their first job for one year or less. Those who spend 1–2 or 2–5 years each amount to around 20%, and only around 10% of workers stay at their first job for 5 years or more. In spite of this high job mobility, first-employer size is a very good predictor of subsequent earnings and employment paths (see Figure 2 and Appendix Figure A6).

Figure 6: Time spent and income earned at the first job
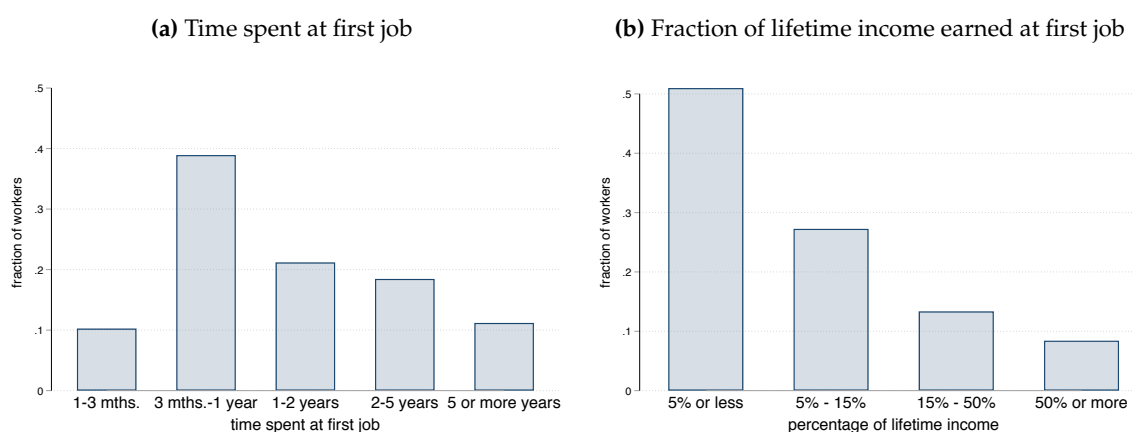
(a) Time spent at first job | **(b)** Fraction of lifetime income earned at first job

Notes: Panel (a): Distribution of time spent at first job. Panel (b): Distribution of the fraction of lifetime income earned at the first job. Lifetime income defined in text. First job is that held during the first continuous six months after predicted graduation in which a person works for 100 days or more. Sample of male workers of all education levels, born in Spain between 1968–1980.

High mobility and earnings growth together imply that only a small fraction of lifetime income is directly earned at the first job. Figure 6 shows that this share is rather low for most workers.[49] Income earned at the first job represents 5% or less of lifetime income for half of the workers in my sample. This number is 5–15% for 28% of workers, 15–50% for 12% of workers, and 50% or more for less than 10% of the workers. Appendix Figure A11 shows that the patterns of Figure 6 are very similar for the workers I previously identified as "likely compliers" (non-college, and born in less urban provinces).[50]

---

[49]To compute this fraction I use the lifetime income measure from equation (1), excluding income earned before or during the first labor market semester. For the numerator I do take into account all income earned at the 1st job (including the first labor market semester).

[50]Appendix Figure A12 provides additional evidence on mobility and time spent at the first employer, together with estimates of the first-employer size effect heterogeneously by time spent at the first job.

Income at age 35

I test for persistence more directly estimating a version of equation (2) in which the outcome variable is income earned during the calendar year a worker turns 35.[51] Table 5 shows the estimated elasticity, which is around 0.09. Appendix Table B3 shows the estimated elasticity is essentially the same when using uncensored earnings at age 35 from tax data.

Table 5: Income during age 35 and first-employer size: OLS and IV-TSLS estimates

	OLS	First Stage	IV-TSLS
	annual income age 35 (1)	first employer size (2)	annual income age 35 (3)
first employer size	0.0368*** (0.0015)		0.0894* (0.0538)
labor demand instrument		0.1010*** (0.0188)	
F-stat excl. instr.		28.89	
SE Clusters	661	661	661
Observations	70588	70588	70588

Notes: OLS and IV-TSLS estimates of the elasticity of annual income during age 35 with respect to first-employer size. Dependent variable is (log) total labor income (earnings and unemployment benefits) during the calendar year the worker turns 35. Includes workers who are employed for at least half of that year. Employer size in logs. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

This result is evidence of persistence at subsequent jobs since at this age only 6.8% of people in my sample are working at their first employer. Moreover, this is the last year of income that enters the lifetime income measure. Previous evidence indicates that the majority of earnings growth occurs in the first ten years of the working life (Topel and Ward, 1992; Rubinstein and Weiss, 2006).[52] Thus, first-employer effects at age 35 (when the average person in the sample has been in the labor market for 15 years) suggest permanent effects past the actual years I consider in my lifetime income measure.

Additional evidence on persistence

Section B.8 in Appendix B includes additional evidence of persistent effects. I estimate a time-varying version of the elasticity of lifetime income with respect to first-employer size and find an increasing first-employer size effect, meaning that a larger first employer results in higher earnings growth. I then show that workers with larger first employers experience greater wage growth when moving to their second job, holding constant first job tenure and second employer size.

---

[51]I do this using 88.3% of workers in my sample who work for at least half of the days in that year.I have estimated linear probability and probit models, neither of which indicate that first employer size impacts the probability of being in this group of 88.3% of workers.

[52]I provide related evidence for Spain in Appendix Figure A4.

5.2 Mechanisms: Search Frictions and Human Capital

So far we have seen that the lifetime first-employer size effect is sustained throughout the working life, time after workers move on to subsequent jobs. To understand the mechanisms behind the first-employer size effect I consider two main drivers of life-cycle wage growth: human capital accumulation and job search (Rubinstein and Weiss, 2006).[53] Section B.9 in Appendix B lays out a simple framework that complements the empirical evidence in this Section, illustrating the job ladder (search) channel first by itself, and then adding a human capital channel as well.

Search frictions and a job ladder

I find evidence consistent with a a job ladder channel. Using the IV framework from Section 4, I show that starting at a larger employer leads to employment at larger subsequent employers. The framework in Section B.9 of Appendix B illustrates this channel under the assumption that jobs at large firms are more desirable (Haltiwanger et al., 2018; Sorkin, 2018). I estimate equation (2) using (log) size of a worker's second employer and (log) size of his employer at age 35 as dependent variables. Results are in Table 6. The IV elasticities between first employer size and that of subsequent employers are between 0.36 and 0.37. Although this result is consistent with other channels apart from a job ladder (e.g. skills developed at large employers could be more valuable at other large employers), it does indicate a persistence in ensuing employers' characteristics that is characteristic of models with frictional search.

Table 6: Subsequent employers and first-employer size: OLS and IV-TSLS estimates

	OLS	First Stage	IV-TSLS	OLS	First Stage	IV-TSLS
	second employer size (1)	first employer size (2)	second employer size (3)	employer size age 35 (4)	first employer size (5)	employer size age 35 (6)
first employer size	0.3232*** (0.0048)		0.3610** (0.1513)	0.2582*** (0.0063)		0.3745** (0.1557)
labor demand instrument		0.0999*** (0.0198)			0.0954*** (0.0198)	
F-stat excl. instr.		25.46			23.3	
SE Clusters	661	661	661	661	661	661
Observations	72742	72742	72742	65217	65217	65217

Notes: OLS and IV-TSLS estimates of the elasticity of subsequent employers' size with respect to first-employer size. Columns (1)–(3) consider as outcome the size of a worker's second employer. Includes workers who change employers at least once before age 35. Columns(4)–(6) consider as outcome the size of a worker's employer at age 35. Includes workers for whom the size of their employer at age 35 is observed in the data. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. All size variables in logs. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

---

[53]Learning about a worker's unobserved abilities, or about job match quality is another potential driver of wage growth (see Jovanovic, 1979; Farber and Gibbons, 1996; Neal, 1999). Like recent work (e.g. Bagger et al., 2014; Jarosch, 2015), I mainly consider human capital and search. Larger firms could enhance learning about workers' abilities through job rotation across different tasks (Eriksson and Ortega, 2006).

Human capital

Better skill-development opportunities at larger firms could also lead to persistent effects and complement a job ladder channel.[54] How can we tell whether, in addition to search frictions, human capital acquisition drives the first-employer size effect? The key insight, present in models of on-the-job search, is that an involuntary unemployment spell cuts a job ladder progression. This is because an unemployed worker looking for a job does not have a current employer as an option to weigh against new offers. In this sense, this brings him to the "bottom" of the ladder.

I categorize workers based on whether they experience an unemployment spell between their first and second job or not.[55] Out of the 76,156 (95% of the sample) workers who had held at least two jobs by age 35, 34,507 (45%) experience unemployment between their first and second jobs. We would expect that a pure job ladder mechanism has little importance among this group of workers. Hence, evidence for persistent first-employer effects for this subsample would be consistent with a human capital channel.

I estimate the elasticity of different long-term outcomes with respect to first-employer size in the subsample of those experiencing unemployment between their first and second jobs. The IV results of equation (2) can be found in Table 7.[56] The key takeaway is that we still see similar long-term effects for this group of workers. For instance, the elasticity for lifetime income in column (2) is equal to .090, compared to the baseline estimate of .117.[57] Elasticities of comparable magnitudes to baseline also arise for average daily wages, lifetime earnings, subsequent employers' size, or income at age 35. The latter is noisily estimated but similar to baseline. These results are consistent with human capital playing a role. If a job ladder channel was the sole driver of the lifetime income effect we would not expect these effects for those whose progression after their first job is cut by unemployment.[58]

So far I have argued that search and human capital are candidate channels behind persistent first-employer effects. I have then provided evidence indicating that search alone is unlikely to explain the lifetime first-employer effect. Thus, the first piece of evidence in favor of a human capital explanation are persistent effects not simply accountable by search. Next, I show further evidence consistent with skills acquired at large employers being more valuable by studying the differential returns to experience from large vs. small employers.

---

[54]Rosen (1972) lays out how on-the-job learning can vary across jobs and its implications. Appendix E discusses firm size and skill development.

[55]I follow previous literature and categorize as having an unemployment spell workers who are not employed for at least two full months between the two jobs (Barlevy, 2008; Hagedorn and Manovskii, 2013).

[56]OLS results can be found in Appendix Table A5.

[57]That the elasticity for this group, while positive, is smaller than those with a job-to-job transition, is intuitively consistent with job ladder *and* human capital channels acting together. The standard errors of these estimates would, however, fail to reject a test of equality.

[58]Similarly, we might not expect persistent effects arising from human capital for those spending a very short amount of time at their first job. The evidence is consistent with this (see panel (d) in Appendix Figure A12).

Table 7: Career outcomes and first-employer size: 1st–2nd job unemployment gap sample

	First Stage	IV-TSLS						
	first employer size (1)	lifetime income (2)	average daily wage (3)	lifetime earnings (4)	days worked (5)	second employer size (6)	annual income age 35 (7)	employer size age 35 (8)
first employer size		0.0900** (0.0403)	0.0794** (0.0340)	0.0832** (0.0422)	0.0038 (0.0189)	0.5374*** (0.1928)	0.0882 (0.0599)	0.4298** (0.1761)
labor demand instrument	0.1132*** (0.0224)							
F-stat excl. instr.	25.45							
SE Clusters	654	654	654	654	654	654	654	653
Observations	34507	34507	34507	34507	34507	32965	30193	27881

Notes: IV-TSLS estimates of the elasticity of different long-term outcomes with respect to first-employer size, using the labor demand instrument detailed in the text. Estimated for the sample of workers who experience an unemployment gap between their first and second jobs (43% of original sample). I count as unemployment gaps in employment that are at least 2 months long. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. All measures are in logs. Column (1) shows the first stage for this sample. Columns (2)–(8) show the elasticity for different long-term outcomes measured between labor market entry and the year a worker turns 35: (2) lifetime income as defined in equation (1), (3) average daily wage, (4) lifetime earnings (lifetime income excluding unemployment benefits), (5) total days worked, (6) size of second employer, (7) annual income during the year worker turns 35, (8) size of worker's employer during year he turns 35. Column (8) excludes workers who worked for less than half the days of the year they turn 35. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

6 Differential Returns to Large-Employer Experience

This section provides additional evidence consistent with human capital mechanisms. I test for a differential return to large-employer experience using the data in its panel dimension, exploiting the richness of its monthly frequency. Observing employer transitions at the daily level and workers' histories since entry allows me to quantify actual experience at different employers measured in days. Experience at large firms could be correlated with unobserved worker characteristics and attributes of the current employer that affect wages.[59] To address this endogeneity problem, the empirical approach features worker fixed effects controlling for worker unobserved heterogeneity, and controls for observable characteristics of the current job (employer size, sector, location, type of contract). The remaining variation that I use compares workers who work for observably similar employers and have the same amount of experience, but acquired this experience in different—large or small—firms.

6.1 Empirical Approach

I estimate the following monthly wage equation:

$$\ln w_{it} = \alpha_i + \psi_{s(i,t)} + \gamma_1 bigExp_{it} + \gamma_2(bigExp_{it} \cdot Exp_{it}) + X_{it}'\delta + \varepsilon_{it}. \tag{5}$$

Where w_{it} is the monthly wage of worker i in month t, α_i are worker fixed effects, $\psi_{s(i,t)}$ are size-category fixed effects for worker i's employer at month t, $bigExp_{it}$ is the amount of actual experience (in days) that worker i has accumulated up until month t at employers with 250 or more employees, and Exp_{it} is the amount of total experience (in days, including both large and small employers).[60] X_{it} includes time-varying controls: a quadratic term for total experience, tenure at current employer (quadratic), age (quadratic), regional unemployment level (quadratic), size of municipality or urban area where employer is located (six categories), type of labor contract (permanent or fixed-term), sector fixed effects, and time (year-month) fixed effects.[61]

The parameters γ_1 and γ_2 capture the differential value of experience at large firms and how this differential varies over the working life. Let $Exp_{it} = bigExp_{it} + smallExp_{it}$ and

---

[59]Literature estimating the returns to general experience and tenure (seniority) includes Altonji and Shakotko (1987), Abraham and Farber (1987), Topel (1991), Altonji and Williams (2005), and Buchinsky et al. (2010). Fackler et al. (2015) document that stayers' wage growth is positively correlated with firm size.

[60]Employer size is not observed before 2004 except for the firms for which I obtained a special extract. To alleviate this missing data issue, in this section I use a measure of employer size that is fixed across time: median size across observed years. In spite of this, some employers' size information is missing (those who had disappeared by 2004). I treat "missing" as an additional size category in this analysis. Thus, $\psi_{s(i,t)}$ groups employers into six categories: missing size, 1–5 employees, 6–19, 20–49, 50–249, and 250+.

[61]This specification is similar to that in De La Roca and Puga (2017), who study worker learning in cities. To the extent that larger employers are located in larger cities, their results would capture part of the differential returns to experience from large firms. My specification controls for contemporaneous city size, including fixed effects for six city-size categories in X_{it} (using urban area size data from De La Roca and Puga, 2017). For additional evidence on returns to city and employer size in Spain see Porcher et al. (2019).

Z_{it} be equation (5) regressors, then

$$\frac{\partial \mathbf{E}(\ln w_{it}|Z_{it})}{\partial bigExp_{it}} - \frac{\partial \mathbf{E}(\ln w_{it}|Z_{it})}{\partial smallExp_{it}} = \gamma_1 + \gamma_2 Exp_{it}. \tag{6}$$

Worker fixed effects α_i prevent (time-invariant) unobserved worker heterogeneity (e.g. innate ability) to bias the differential return estimates. Controls for current-employer size, $\psi_{s(i,t)}$ (together with sector and location controls), imply that γ_1 and γ_2 are identified comparing workers who have different experience profiles but have the same amount of total experience and are currently working for similar employers. Comparing estimates of γ_1 and γ_2 with and without including $\psi_{s(i,t)}$ is an informative exercise. Intuitively, differential returns to large employer experience *not* controlling for current employer size will combine returns to skill *and* job search. Keeping constant current employer characteristics controls for returns to job search (at least among the observed employer characteristics). That is, if part of the benefits of past experience at a large firm comes from human capital *and* the possibility to be working at a large firm today, specifications including $\psi_{s(i,t)}$ will keep constant the latter channel, making the estimated returns more plausibly attributed to skill accumulation.[62]

6.2 Findings

Columns (1) and (2) of Appendix Table A6 show estimates of equation (5). Column (1) does not include current-employer size category fixed effects $\psi_{s(i,t)}$, while column (2) does. In both cases $\hat{\gamma}_1$ and $\hat{\gamma}_2$ indicate that large-employer experience has higher returns than other experience, and that the differential slowly decreases over time. The fact that $\hat{\gamma}_1$ from column (1) is significantly larger than that from column (2), indicates how a job ladder effect can be of importance. While $\hat{\gamma}_2$ indicates a decreasing differential, the rate of decline is small. Figure 7 helps understand the magnitude implied by the coefficients and its evolution over time. The figure plots the differential return to one year of large-employer experience (specification including $\psi_{s(i,t)}$).[63] One year of large-employer experience is associated with a return that is between 2–3 percentage points higher than a year of experience elsewhere. As benchmark, the average annual wage growth during the first eight years in the labor market is 10%. These results suggest that there is a differential return to large-employer experience, its magnitude is economically significant, and seems to be more relevant at the beginning of the working life.[64]

---

[62] Abraham and Farber (1987) make a similar point about the distinction of returns to experience *per se* and the returns to job search.

[63] Concretely, it plots $365 \cdot 100(\hat{\gamma}_1 + \hat{\gamma}_2 Exp)$ for different levels of Exp, together with 95% confidence intervals.

[64] Figure 7 displays marginal effects up until 12 years of (actual) experience since that is close to the average level of experience for workers in my sample at age 35, which is where the panel I use to estimate equation (5) ends.

Figure 7: Differential wage return to one year of large employer experience, by total experience
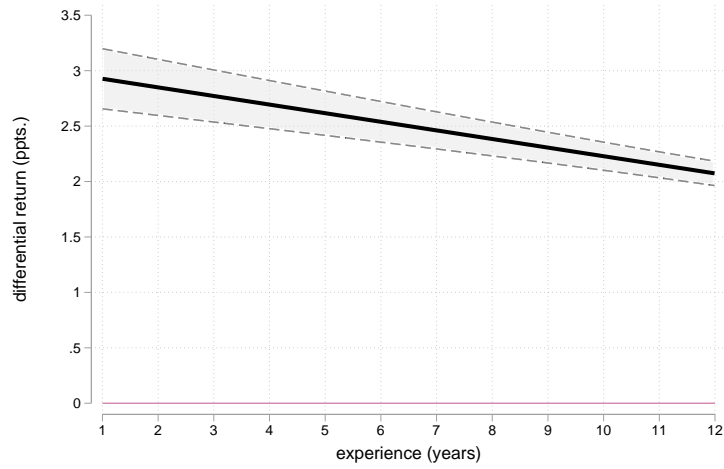
Notes: Monthly wage differential return to one year of experience at a large employer (250+ employees) with respect to a year of experience elsewhere (<250 employees) for different overall experience levels. Uses estimates of equation (5) (Table A6, column (2)) and plots $365 \cdot 100(\hat{\gamma}_1 + \hat{\gamma}_2 Exp)$ and a 95% level confidence interval computed using the delta method. Exp is measured in days, x-axis re-scaled for readability. Standard errors are clustered at the worker level.

Potential threats

In Appendix B, Section B.10 I address potential concerns that could bias the estimates of differential return to experience, or threaten their interpretation as return to skills. These include the possibility of large-firm experience working as a signal of (preexisting) high unobserved productivity, and possible bias arising from the additive separability assumption of worker and firm-size effects. The results from these robustness checks support the above interpretation of the findings.

Promotions

In Section B.11 of Appendix B I estimate the differential return of large employer experience on the probability of being promoted, with specifications similar to (5) in which the dependent variable is a promotion dummy. I show that, analogously to wages, time spent at a large employer is more valuable than that spent elsewhere in terms of future career progression.

7 Conclusion

This paper sheds new light on how firm heterogeneity affects workers' prosperity in the long term. My findings imply that working for different firm types as a young person can have effects that last throughout one's career, and that firm size is a relevant measure that captures meaningful employer characteristics. Compared to other firm attributes, size has the advantage that is easily observable to workers and policymakers and that, not being model-based, is transparently measured.

I am able to identify a causal link between first-employer size and long-term outcomes because, in spite of the importance of a first-job match, there is some randomness involved in the matching process. My IV approach leverages the part of this randomness that is driven by idiosyncratic hiring shocks of large firms. An inherent feature of the IV approach is that it estimates causal effects for workers whose first-employer match is most affected by idiosyncratic large-firm hiring shocks in their region of birth; that is, marginal large-firm hires. I find that these entrants are younger and with lower earnings potential. These "compliers" seem to disproportionately benefit from starting out at a large first employer. An interesting question outside the scope of this paper is to understand the characteristics of an equilibrium in which the marginal large-firm worker has the largest long-term benefits from such a match. It could be that any human-capital benefits a worker derives are proportional to the costs she generates for the firm (through training, or the time it takes to learn the ropes of the job). Firms might not want to, or not be able to, discriminate wages based on these costs and benefits. Even if firms did lower wages to equalize long-term benefits, there are reasons why young workers could turn down such a deal (e.g. unawareness of long-term benefits, or liquidity and credit constraints).

A human capital channel is consistent with the evidence. Firm heterogeneity in the provision of on-the-job skills has important implications. In the presence of imperfect wage adjustment and worker mobility, firms that increase young workers' productivity in persistent ways will not fully internalize this fact in their choices. Additionally, the efficiency losses some argue arise from size-dependent policies and regulations (IMF, 2015; Guner et al., 2008) could be underestimated if larger firms provide more valuable skills. It could be especially productive to acquire such skills early in the working life, when workers are in a formative period.[65]

A limitation of this paper is that I cannot test for certain channels that, in addition to human capital and search, could explain part of the first-employer size effect. In particular, social and professional networks that are developed in large firms could be larger or more valuable than those developed in small firms. These social ties could impact access to future jobs and be beneficial throughout the working life (Cingano and Rosolia, 2012). Since studying networks typically requires population data, this is an interesting potential channel outside the scope of this paper.

Finally, a better understanding of what it is that "good human-capital" firms do well could be informative for training and active labor market policies. Policy could also be used to encourage such firm practices. Overall, compared to what we know regarding the heterogeneous opportunities that open up from formal education of one type or another, we know little about the heterogeneous opportunities that might arise from spending key formative time as a young worker at one employer or another. This paper hopefully provides a first step towards increasing our understanding.

---

[65]The skill formation literature emphasizes the complementarity in skills —"skill begets skill"—and that skills developed at one stage of the life cycle impact the productivity of further skill development (Cunha et al., 2006). While this literature focuses on childhood, similar mechanisms could apply to skill development of young adults. Evidence indicating that cognitive abilities start declining when people reach their 20s (see Salthouse, 2009) is indicative of the importance of skill acquisition during the early career stages.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? Working Paper 24003, National Bureau of Economic Research.

Abowd, J. M., K. L. McKinney, and N. L. Zhao (2018). Earnings inequality and mobility trends in the united states: Nationally representative estimates from longitudinally linked employer-employee data. *Journal of Labor Economics 36*(S1), S183–S300.

Abraham, K. G. and H. S. Farber (1987). Job duration, seniority, and earnings. *The American Economic Review 77*(3), 278–297.

Acemoglu, D. and J.-S. Pischke (1999). The structure of wages and investment in general training. *Journal of Political Economy 107*(3), 539–572.

Albæk, K., M. Arai, R. Asplund, E. Barth, and E. S. Madsen (1998). Measuring wage effects of plant size. *Labour Economics 5*(4), 425–448.

Altonji, J. G., L. B. Kahn, and J. D. Speer (2016). Cashier or consultant? Entry labor market conditions, field of study, and career success. *Journal of Labor Economics 34*(1), 361–401.

Altonji, J. G. and C. R. Pierret (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics 116*(1), 313–350.

Altonji, J. G. and R. A. Shakotko (1987). Do wages rise with job seniority? *The Review of Economic Studies 54*(3), 437–459.

Altonji, J. G. and N. Williams (2005). Do wages rise with job seniority? A reassessment. *Industrial and Labor Relations Review 58*(3), 370–397.

Angrist, J. D. and G. W. Imbens (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association 90*(430), 431–442.

Antonacopoulou, E. P. and W. H. Güttel (2010). Staff induction practices and organizational socialization: A review and extension of the debate. *Society and Business Review 5*(1), 22–47.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review 104*(6), 1551–96.

Barlevy, G. (2008). Identification of search models using record statistics. *The Review of Economic Studies 75*(1), 29–64.

Beaudry, P. and J. DiNardo (1991). The effect of implicit contracts on the movement of wages over the business cycle: Evidence from micro data. *Journal of Political Economy 99*(4), 665–688.

Bernard, A. B. and J. B. Jensen (1999). Exceptional exporter performance: Cause, effect, or both? *Journal of International Economics 47*(1), 1–25.

Bernard, A. B., J. B. Jensen, and R. Z. Lawrence (1995). Exporters, jobs, and wages in US manufacturing: 1976-1987. *Brookings Papers on Economic Activity. Microeconomics 1995*, 67–119.

Bloom, N., F. Guvenen, B. S. Smith, J. Song, and T. von Wachter (2018). The disappearing large-firm wage premium. *AEA Papers and Proceedings 108*, 317–22.

Bloom, N. and J. Van Reenen (2006). Measuring and explaining management practices across firms and countries. CEP Discussion Paper 716, Centre for Economic Performance.

Bloom, N. and J. Van Reenen (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics 122*(4), 1351–1408.

Bonhomme, S. and L. Hospido (2017). The cycle of earnings inequality: Evidence from Spanish social security data. *The Economic Journal*.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework for matched employer employee data. *Econometrica 87*(3), 699–739.

Borusyak, K., P. Hull, and X. Jaravel (2018). Quasi-experimental shift-share research designs. Working Paper 24997, National Bureau of Economic Research.

Brown, C. and J. Medoff (1989). The employer size-wage effect. *Journal of Political Economy 97*(5), 1027–1059.

Brunner, B. and A. Kuhn (2014). The impact of labor market entry conditions on initial job assignment and wages. *Journal of Population Economics 27*(3), 705–738.

Buchinsky, M., D. Fougere, F. Kramarz, and R. Tchernis (2010). Interfirm mobility, wages and the returns to seniority and experience in the United States. *The Review of Economic Studies 77*(3), 972–1001.

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 257–273.

Caicedo, S., R. E. Lucas Jr, and E. Rossi-Hansberg (2019). Learning, career paths, and the distribution of wages. *American Economic Journal: Macroeconomics 11*(1), 49–88.

Caliendo, L., F. Monte, and E. Rossi-Hansberg (2015). The anatomy of French production hierarchies. *Journal of Political Economy 123*(4), 809–852.

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

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics 128*(3), 967–1015.

Card, D. and T. Lemieux (2001). Dropout and enrollment trends in the postwar period. In J. Gruber (Ed.), *Risky Behavior among Youths: An Economic Analysis*, pp. 439–482. University of Chicago Press.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

Chislett, W. (2002). *The Internationalization of the Spanish Economy*. Real Instituto Elcano de Estudios Internacionales y Estratégicos.

Christensen, B. J., R. Lentz, D. T. Mortensen, G. R. Neumann, and A. Werwatz (2005). On-the-job search and the wage distribution. *Journal of Labor Economics 23*(1), 31–58.

Cingano, F. and A. Rosolia (2012). People I know: Job search and social networks. *Journal of Labor Economics 30*(2), 291–332.

Colonnelli, E., J. Tåg, M. Webb, and S. Wolter (2018). A cross-country comparison of dynamics in the large firm wage premium. *AEA Papers and Proceedings 108*, 323–27.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education 1*, 697–812.

Davis, S. J. and B. Samaniego de la Parra (2017). Application flows. Mimeo, University of Chicago.

De La Roca, J. and D. Puga (2017). Learning by working in big cities. *The Review of Economic Studies 84*(1), 106–142.

Devereux, P. J. (2002). The importance of obtaining a high-paying job. Mimeo, University of California Los Angeles.

Dolado, J. J., C. García-Serrano, and J. F. Jimeno (2002). Drawing lessons from the boom of temporary jobs in Spain. *The Economic Journal 112*(480), F270–F295.

Eeckhout, J. and P. Kircher (2018). Assortative matching with large firms. *Econometrica 86*(1), 85–132.

Eriksson, T. and J. Ortega (2006). The adoption of job rotation: Testing the theories. *ILR Review 59*(4), 653–666.

Fabiani, S., F. Schivardi, and S. Trento (2005). ICT adoption in Italian manufacturing: firm-level evidence. *Industrial and Corporate Change 14*(2), 225–249.

Fackler, D., S. Müller, and J. Stegmaier (2017). Explaining wage losses after job displacement: Employer size and lost firm rents. Discussion Paper 32/2017, Halle Institute for Economic Research (IWH).

Fackler, D., T. Schank, and C. Schnabel (2015). Does the plant size–wage differential increase with tenure? Affirming evidence from German panel data. *Economics Letters 135*, 9–11.

Farber, H. S. and R. Gibbons (1996). Learning and wage dynamics. *The Quarterly Journal of Economics 111*(4), 1007–1047.

Fernández-Kranz, D. and N. Rodríguez-Planas (2018). The perfect storm: Graduating in a recession in a segmented labor market. *ILR Review 72*(2), 492–524.

Florens, J.-P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica 76*(5), 1191–1206.

Fox, J. T. (2009). Firm-size wage gaps, job responsibility, and hierarchical matching. *Journal of Labor Economics 27*(1), 83–126.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), 733–772.

García Pérez, J. I. (2008). La muestra continua de vidas laborales: Una guía de uso para el análisis de transiciones. *Revista de Economía Aplicada 16*(1).

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy 108*(5), 874–904.

Garicano, L. and E. Rossi-Hansberg (2015). Knowledge-based hierarchies: Using organizations to understand the economy. *Annual Review of Economics 7*(1), 1–30.

Gibbons, R. and M. Waldman (1999). Careers in organizations: Theory and evidence. Volume 3 of *Handbook of Labor Economics*, pp. 2373 – 2437. Elsevier.

Gibbons, R. and M. Waldman (2006). Enriching a theory of wage and promotion dynamics inside firms. *Journal of Labor Economics 24*(1), 59–107.

Gittleman, M., M. Horrigan, and M. Joyce (1998). "Flexible" workplace practices: Evidence from a nationally representative survey. *ILR Review 52*(1), 99–115.

Goerlich Gisbert, F. J. and I. Cantarino Martí (2015). Estimaciones de la población rural y urbana a nivel municipal. *Estadística Española 57*(186), 5–28.

Goldschmidt, D. and J. F. Schmieder (2017). The rise of domestic outsourcing and the evolution of the German wage structure. *The Quarterly Journal of Economics 132*(3), 1165–1217.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2018). Bartik instruments: What, when, why and how. Mimeo, Yale School of Management.

Greenstone, M., R. Hornbeck, and E. Moretti (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy 118*(3), 536–598.

Guner, N., G. Ventura, and Y. Xu (2007). Macroeconomic implications of size-dependent policies. Working Paper 6138, Centre for Economic Policy Research.

Guner, N., G. Ventura, and Y. Xu (2008). Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics 11*(4), 721–744.

Guvenen, F., G. Kaplan, J. Song, and J. Weidner (2017). Lifetime incomes in the United States over six decades. Working Paper 23371, National Bureau of Economic Research.

Hagedorn, M. and I. Manovskii (2013). Job selection and wages over the business cycle. *American Economic Review 103*(2), 771–803.

Haltiwanger, J. C., H. R. Hyatt, L. B. Kahn, and E. McEntarfer (2018). Cyclical job ladders by firm size and firm wage. *American Economic Journal: Macroeconomics 10*(2), 52–85.

Ham, J. C. and R. J. LaLonde (1996). The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training. *Econometrica 64*(1), 175–205.

Idson, T. L. and D. J. Feaster (1990). A selectivity model of employer-size wage differentials. *Journal of Labor Economics 8*(1, Part 1), 99–122.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

IMF (2015). Obstacles to firm growth in Spain. IMF country report 15/233.

Jarosch, G. (2015). Searching for job security and the consequences of job loss. Mimeo, Princeton University and Stanford University.

Jarosch, G., E. Oberfield, and E. Rossi-Hansberg (2018). Learning from coworkers. Mimeo, Princeton University.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy 87*(5, Part 1), 972–990.

Jovanovic, B. and Y. Nyarko (1997). Stepping-stone mobility. *Carnegie-Rochester Conference Series on Public Policy 46*, 289–325.

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics 17*(2), 303–316.

Kramarz, F. and O. N. Skans (2014, 01). When Strong Ties are Strong: Networks and Youth Labour Market Entry. *The Review of Economic Studies 81*(3), 1164–1200.

Krolikowski, P. (2017). Job ladders and earnings of displaced workers. *American Economic Journal: Macroeconomics 9*(2), 1–31.

Kugler, M. and E. Verhoogen (2012). Prices, plant size, and product quality. *The Review of Economic Studies 79*(1), 307–339.

Kwon, I., E. Meyersson Milgrom, and S. Hwang (2010). Cohort effects in promotions and wages evidence from Sweden and the United States. *Journal of Human Resources 45*(3), 772–808.

Lallemand, T., R. Plasman, and F. Rycx (2007). The establishment-size wage premium: Evidence from European countries. *Empirica 34*(5), 427–451.

Lazear, E. P. (2009). Firm-specific human capital: A skill-weights approach. *Journal of Political Economy 117*(5), 914–940.

Leung, D., C. Meh, and Y. Terajima (2008). Firm size and productivity. Working Paper 2008–45, Bank of Canada.

Liu, K., K. G. Salvanes, and E. Ø. Sørensen (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review 84*, 3–17.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics 9*(2), 508–523.

Lynch, L. M. and S. E. Black (1998). Beyond the incidence of employer-provided training. *Industrial and Labor Relations Review 52*(1), 64–81.

Macis, M. and F. Schivardi (2016). Exports and wages: Rent sharing, workforce composition, or returns to skills? *Journal of Labor Economics 34*(4), 945–978.

Main, B. G. M. and B. Reilly (1993). The employer size-wage gap: Evidence for Britain. *Economica 60*(238), 125–142.

Moore, H. L. (1911). *Laws of Wages: An Essay in Statistical Economics*. The MacMillan Company.

Moral-Benito, E. (2018). Growing by learning: Firm-level evidence on the size-productivity nexus. *SERIEs 9*(1), 65–90.

Moscarini, G. and F. Postel-Vinay (2012). The contribution of large and small employers to job creation in times of high and low unemployment. *The American Economic Review 102*(6), 2509–2539.

Müller, S. and R. Neubaeumer (2018). Size of training firms: The role of firms, luck, and ability in young workers' careers. *International Journal of Manpower 39*(5), 658–673.

Neal, D. (1999). The complexity of job mobility among young men. *Journal of Labor Economics 17*(2), 237–261.

Nix, E. (2017). Learning spillovers in the firm. Mimeo, University of Southern California.

Oi, W. Y. and T. L. Idson (1999). Firm size and wages. *Handbook of Labor Economics 3*, 2165–2214.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics 4*(1), 1–29.

Oyer, P. (2006). Initial labor market conditions and long-term outcomes for economists. *The Journal of Economic Perspectives 20*(3), 143–160.

Oyer, P. (2008). The making of an investment banker: Stock market shocks, career choice, and lifetime income. *The Journal of Finance 63*(6), 2601–2628.

Petrongolo, B. and M. J. San Segundo (2002). Staying-on at school at 16: the impact of labor market conditions in Spain. *Economics of Education Review 21*(4), 353–365.

Porcher, C., H. Rubinton, and C. Santamaría (2019). The role of establishment size in the city-size premium in Spain. Mimeo, Princeton University and Universidad Carlos III de Madrid.

Rosen, S. (1972). Learning and experience in the labor market. *Journal of Human Resources*, 326–342.

Rubinstein, Y. and Y. Weiss (2006). Post schooling wage growth: Investment, search and learning. *Handbook of the Economics of Education 1*, 1–67.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging 30*(4), 507–514.

Schoar, A. and L. Zuo (2017). Shaped by booms and busts: How the economy impacts CEO careers and management styles. *The Review of Financial Studies 30*(5), 1425–1456.

Schwandt, H. and T. Von Wachter (2019). Unlucky cohorts: Estimating the long-term effects of entering the labor market in a recession in large cross-sectional data sets. *Journal of Labor Economics 37*(S1), S161–S198.

Sievertsen, H. H. (2016). Local unemployment and the timing of post-secondary schooling. *Economics of Education Review 50*, 17–28.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019). Firming up inequality. *The Quarterly Journal of Economics 134*(1), 1–50.

Sorkin, I. (2018). Ranking firms using revealed preference. *The Quarterly Journal of Economics 133*(3), 1331–1393.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature 49*(2), 326–65.

Topel, R. (1991). Specific capital, mobility, and wages: Wages rise with job seniority. *Journal of Political Economy 99*(1), 145–176.

Topel, R. and M. P. Ward (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics 107*(2), 439–479.

Van Maanen, J. (1978). People processing: Strategies of organizational socialization. *Organizational dynamics 7*(1), 19–36.

von Wachter, T. and S. Bender (2006). In the right place at the wrong time: The role of firms and luck in young workers' careers. *American Economic Review 96*(5), 1679–1705.

von Wachter, T. and S. Bender (2008). Do initial conditions persist between firms? An analysis of firm-entry cohort effects and job losers using matched employer-employee data. In S. Bender, J. Lane, K. Shaw, F. Andersson, and T. von Wachter (Eds.), *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches*, pp. 135–162. University of Chicago Press.

Wee, S. L. (2016). Delayed learning and human capital accumulation: The cost of entering the job market during a recession. Mimeo, Carnegie Mellon University.
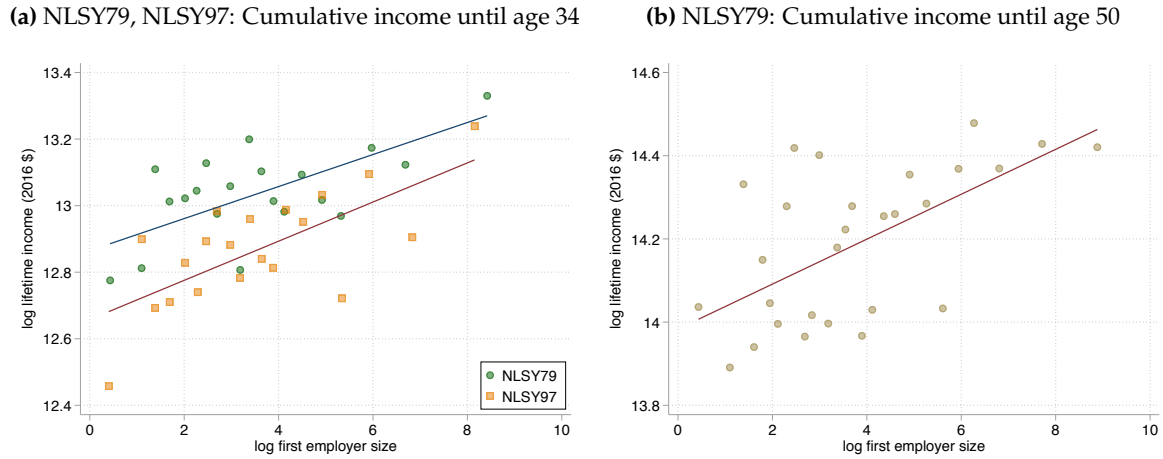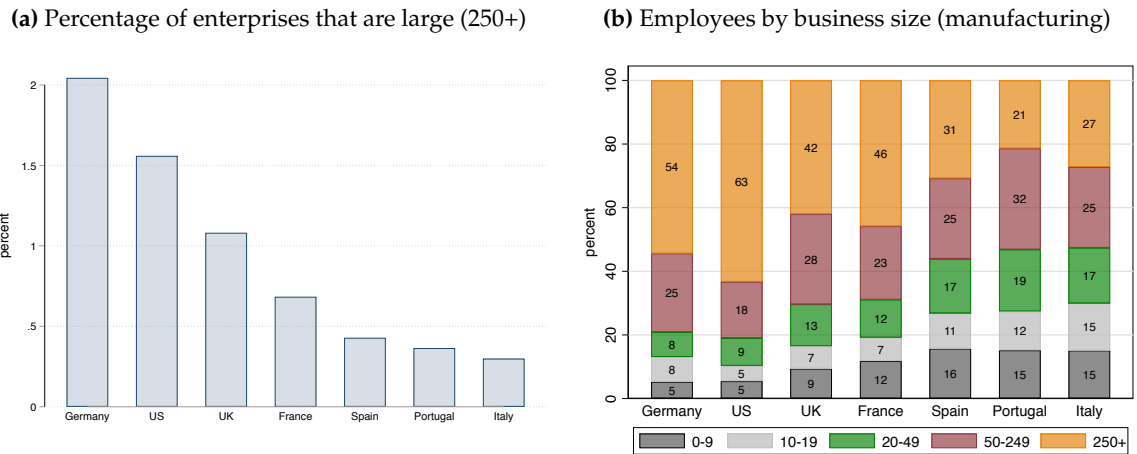
- SUPPLEMENTARY APPENDICES -

A   Additional Figures and Tables

Figure A1: Lifetime income and first-employer size in U.S. panel survey data

(a) NLSY79, NLSY97: Cumulative income until age 34   **(b)** NLSY79: Cumulative income until age 50


Notes: Sources are NLSY79 and NLSY97. Binned scatterplots. Horizontal axis plots the (log) establishment size of a respondents' first job after formal education. Vertical axis plots (log) lifetime income. Lifetime income is the cumulative sum of wage and salary income and unemployment benefits since labor market entry up until and including the year a respondent turns 34 (Panel (a)) or turns 50 (Panel (b)). Sample weights are used. Sample composed of male respondents who hold their first job before age 27, and appear in the survey in all years in which income contribute to lifetime income measures. Income in non-survey years is imputed as the midpoint of income in adjacent survey years. Panel (a): Correlation(79) =.15, N(79)=1,001, Correlation(97)=.11, N(97)=1,280. Panel (b): Correlation=.16, N=697.

Figure A2: Relatively few large firms in Spain: Firm size across countries

(a) Percentage of enterprises that are large (250+)   **(b)** Employees by business size (manufacturing)


Notes: Source is OECD. Data refer to year 2013. Panel (a): Percentage of total number of enterprises (excludes self-employed) in each country that have 250 employees or more. Panel (b): Percentage of manufacturing workers working in each employer size category. Categories might not add up to 100 due to rounding.

Figure A3: First-employer size distribution and second-employer transition

(a) First employer

(b) Second-employer transition
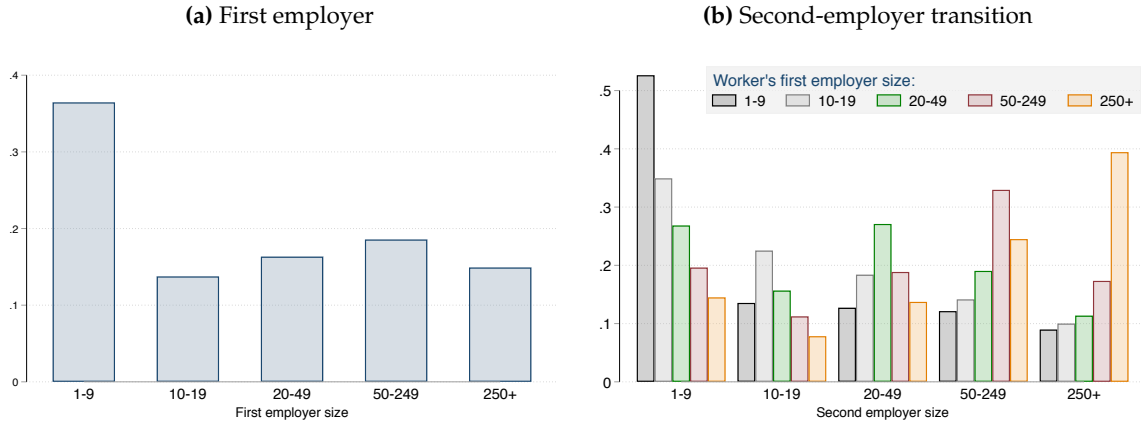
Notes: Panel (a): Distribution of first-employer size. Panel (b): Distribution of second-employer size, separately by first-employer size groups. Sample of male workers of all education levels, born in Spain between 1968–1980.

Figure A4: Income stabilizes by age 35: Annual income and growth age profiles (2006–2015)

(a) Log income age profile

(b) Income growth age profile
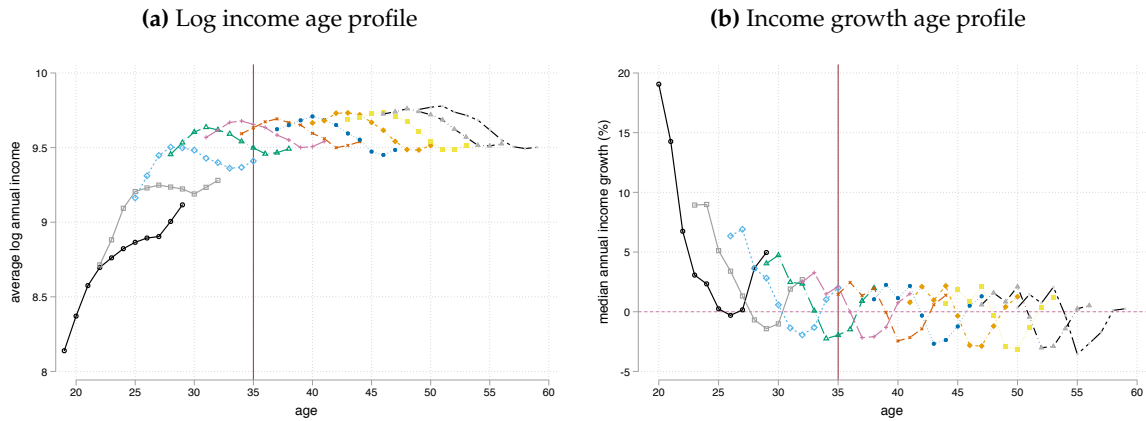
Notes: Age profiles for different cohorts in log annual income and annual income growth rates. Left panel: average log annual income by cohort and year. Right panel: median annual income growth rate by cohort and year. Growth rate g_t between annual income Y_{t-1} and Y_t computed as $100 \times \frac{Y_t - Y_{t-1}}{\frac{1}{2} \cdot (Y_t + Y_{t-1})}$ using longitudinal tax data on annual earnings for the years 2006–2015. Sample of Spain-born individuals who in a given year earn at least 2,400 Euro (2016 Euro). Each series represents a different birth cohort.

Figure A5: Lifetime income and first-employer size correlation: controlling for sector
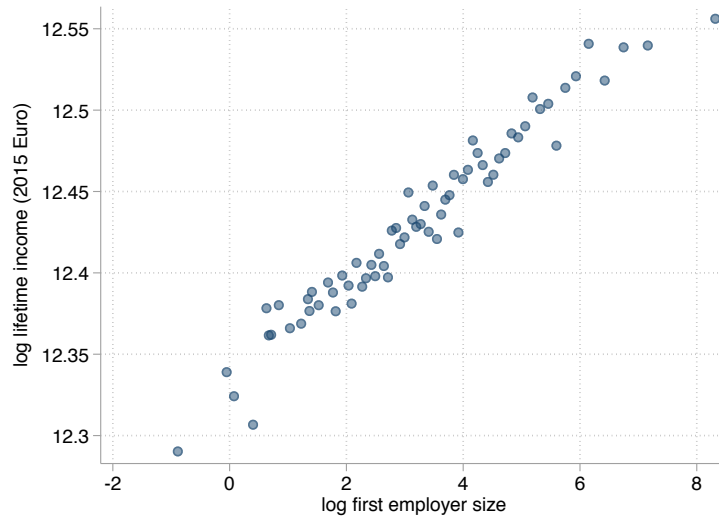
Notes: Conditional expectation of lifetime income as a function of first-employer size, adjusting for sector of first employer. Binned scatterplot. Log lifetime income (as defined in the text) on the vertical axis. Log size of worker's first employer on the horizontal axis. Both variables net out of 58 first-employer sector fixed effects. Sample of male workers of all education levels, born in Spain between 1968–1980.

Figure A6: Daily wages and unemployment trajectories by first-employer size category

(a) Average daily wages

(b) Fraction experiencing unemployment
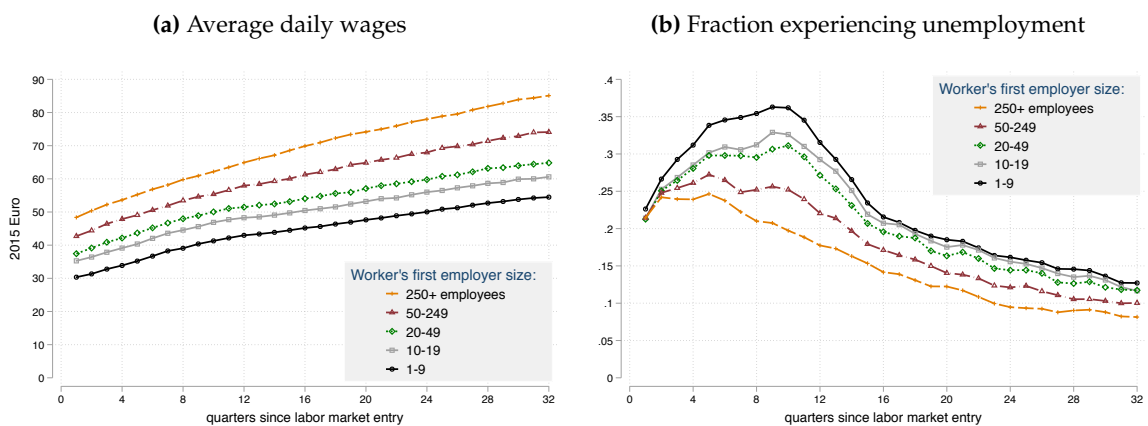
Notes: Panel (a): Evolution of average daily wages since labor market entry. Panel (b): Fraction of workers experiencing unemployment since labor market entry. Both panels categorize workers based on the size of their first employer. Sample of male workers of all education levels, born in Spain between 1968–1980.

Figure A7: Additional first labor market semester summary statistics, by educational attainment

(a) Year of first labor market experience
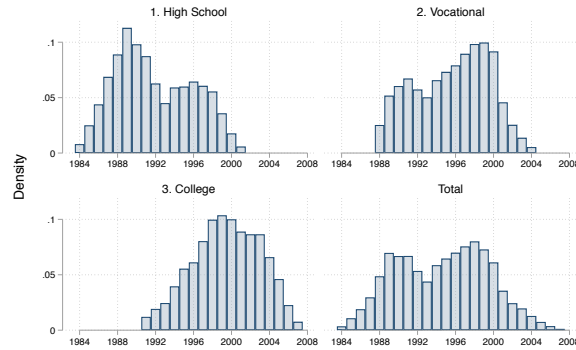
(b) Age at first labor market experience
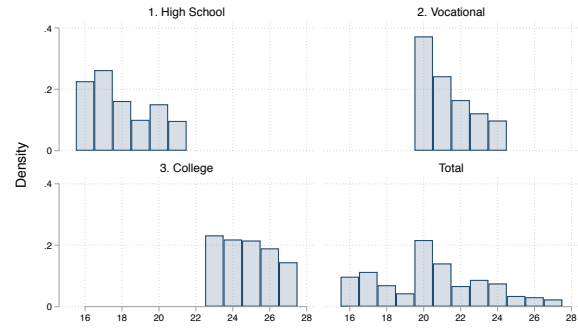
(c) Days worked during first labor market experience
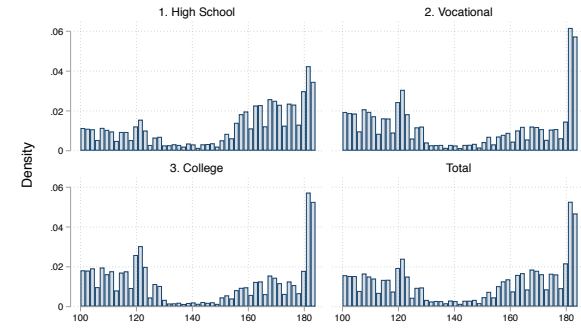
(d) First employer log size
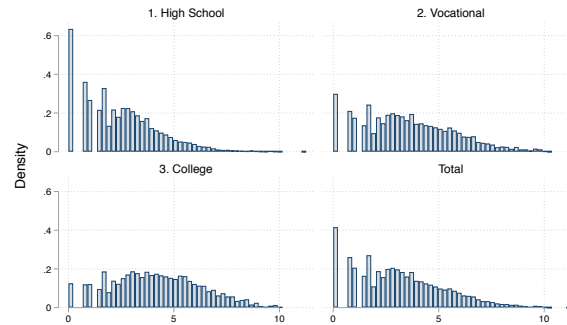
(e) Number of employers during first labor market experience
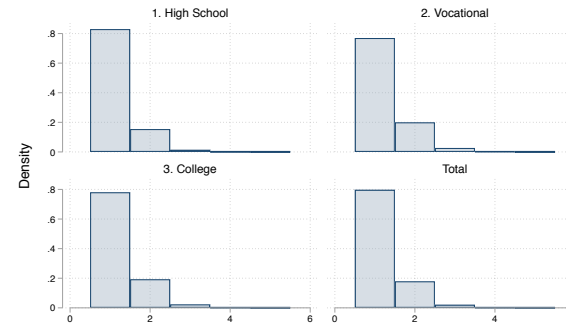
Notes: First labor market experience defined as the first six continuous months after predicted graduation (defined in text) that a person works for 100 days or more. All figures plotted separately by education level and overall. Panel (a): Calendar year of first labor market experience. Panel (b): Age distribution during first labor market experience. Panel (c): Distribution of days worked during the first labor market experience. Panel(d): Distribution of (log) first employer size. Panel (e): Distribution of number of employers during the first labor market experience. Sample of male workers of all education levels, born in Spain between 1968–1980.

**Figure A8:** IV residual variation and business cycle variation in Catalunya region
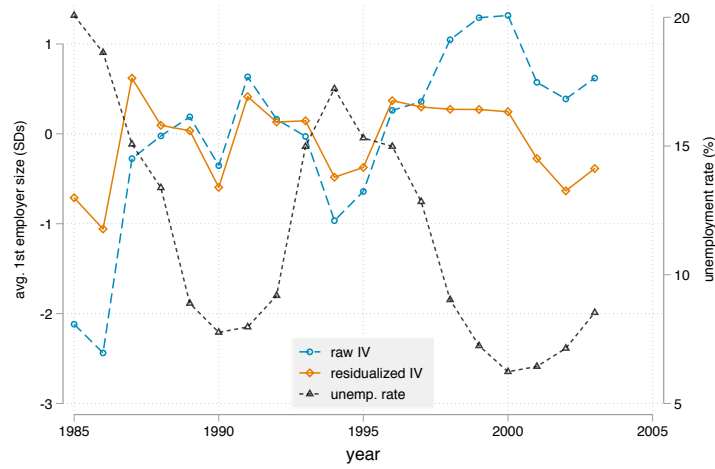
Notes: Time series evolution of the unemployment rate in Catalunya (black triangles), the instrument $\bar{s}_{-i}^{rec}$ described in the text (blue dots), and residuals from a regression of $\bar{s}_{-i}^{rec}$ on region of birth, education, and cohort fixed effects and a flexible of the regional unemployment rate at the worker's region of birth in his predicted graduation year (orange diamonds).

Figure A9: Labor demand instrument: residual variation
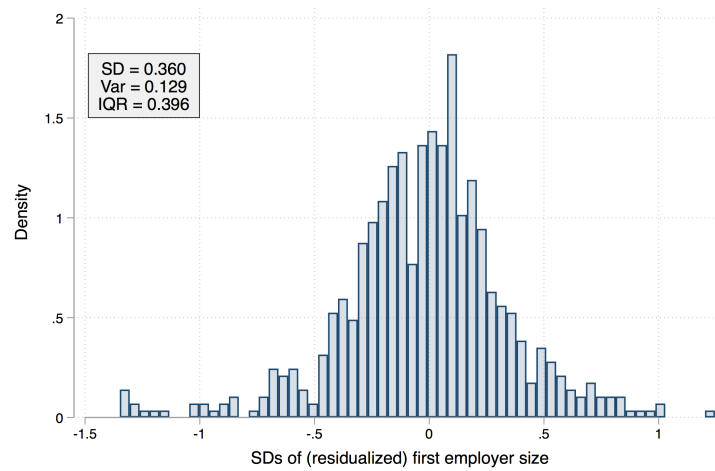
Notes: Histogram of (residualized) labor demand instrument across region of birth $\times$ education $\times$ year of birth bins. Expressed in units of standard deviations of (residualized) first employer size. In both cases residuals from a regression on a flexible function of unemployment rate at predicted graduation year, education fixed effects, region fixed effects, and birth cohort fixed effects.

Figure A10: IV-TSLS elasticity of lifetime income w.r.t. first-employer size

(a) First stage
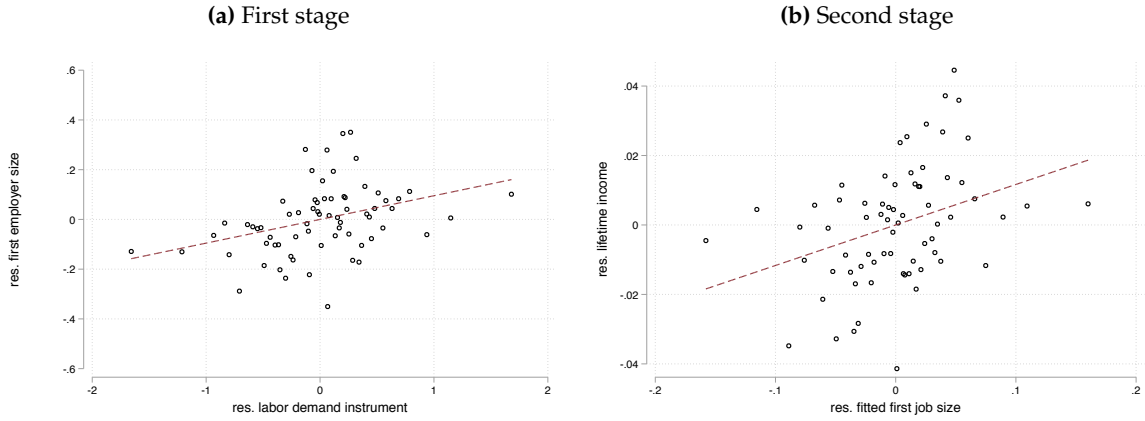
(b) Second stage

Notes: Binned scatterplots of first stage and second stage residual variation from equation (2) in the text, instrumenting for log first job size $s_{J(i)}$ using the instrument $\bar{s}_{-i}^{rec}$ described in the text. The outcome variable is log total income after first job semester (described in text) up until age 35. Sample of male workers of all education levels, born in Spain between 1968–1980.

Figure A11: Time spent and income earned at the first job: Subsample of "likely compliers"

(a) Time spent at first job

(b) Fraction of lifetime income earned at first job

Notes: Panel (a): Distribution of time spent at first job. Panel (b): Distribution of the fraction of lifetime income earned at the first job. Lifetime income defined in text. First job is that held during the first continuous six months after predicted graduation in which a person works for 100 days or more. Subsample of workers without a college degree, and born in less urban parts of Spain (i.e. "likely compliers", as explained in the text). These workers amount to 37% percent of original sample.

Figure A12: Time spent at first job by employer size, and IV results by time spent at first job

(a) Time spent at first job

(b) Fraction of workers at first employer

(c) Density of time spent at first job

(d) IV elasticity by time spent at first job
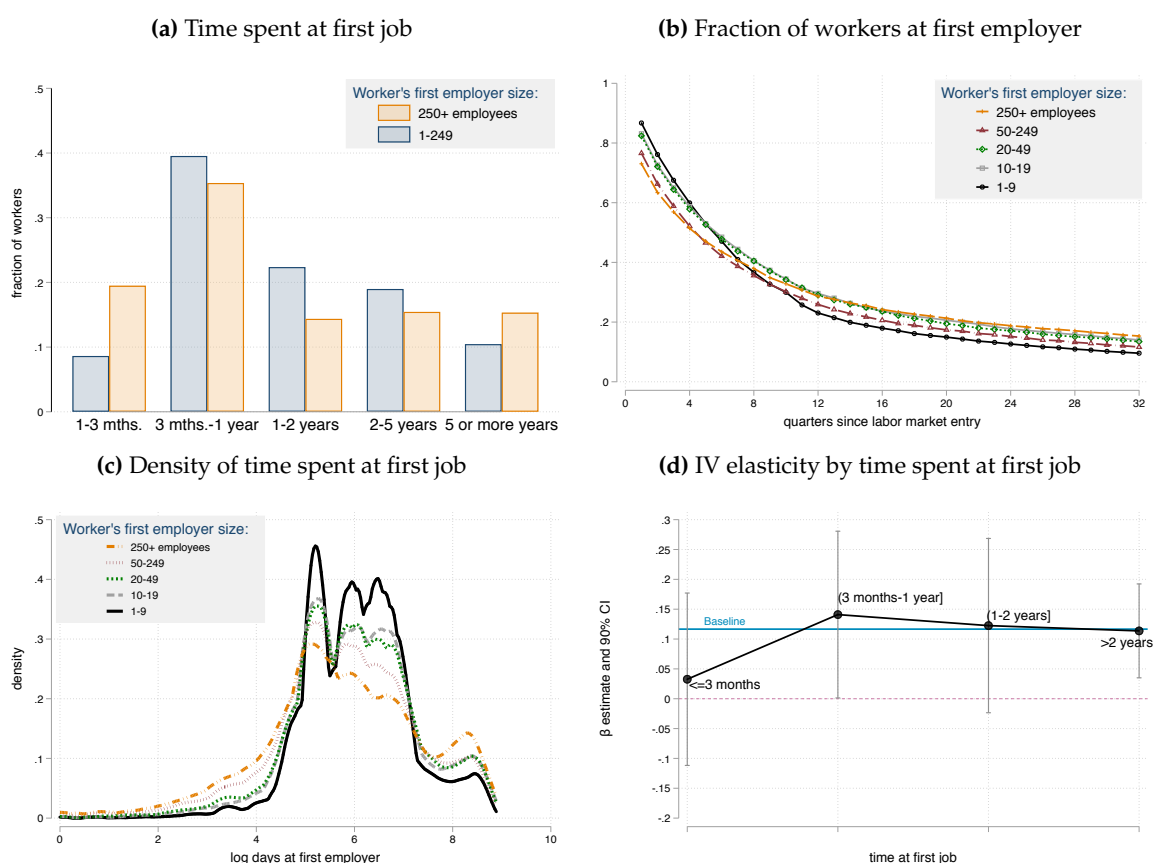
Notes: Panel (a): Distribution of time spent at first job, separately for workers starting at large employers (250+ employees) and everyone else. Panel (b): Fraction of workers who are currently working at their first employer, separately by first-employer size category. Panel (c): Kernel density estimates of (log) days spent at first job, by first-employer size. Panel (d): Elasticity of lifetime income with respect to first-employer size (TSLS estimates of equation (2)), estimated separately for four groups of workers based on the time spent at the first employer. Group $\leq$ 3 months: N=7,455. Group (3 months–1 year]: N=29,405. Group (1–2 years]: N=18,138. Group >2 years: N=24,943. Sample of male workers of all education levels, born in Spain between 1968–1980.

Table A1: IV residual variation uncorrelated with the business cycle

	Dependent variable = $\bar{s}^{rec}$				
	(1)	(2)	(3)	(4)	(5)
unemployment rate	-0.0828***	-0.0523***			
	(0.0076)	(0.0081)			
GDP growth			0.0707**	0.0263*	0.0055
			(0.0300)	(0.0151)	(0.0153)
SE Clusters	661	661	661	661	661
Fixed effects	no	yes	no	yes	yes
f(unemployment)	no	no	no	no	yes
Observations	79941	79941	79941	79941	79941

Notes: OLS relationship between the labor-demand composition instrument $\bar{s}^{rec}_{-i}$ (defined in the text) and business cycle conditions at workers' region of birth during predicted graduation year. Business-cycle conditions measured by the regional unemployment rate or regional GDP growth. Columns (2), (4), and (5) control for region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Column (5) additionally controls for a flexible function of regional unemployment during predicted graduation year. Regressions at the worker level. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Table A2: Lifetime income and first-employer size, varying discount factor

	OLS				First Stage	IV–TSLS			
	lifetime inc. 0%	lifetime inc. 1%	lifetime inc. 2%	lifetime inc. 3%	first job size	lifetime inc. 0%	lifetime inc. 1%	lifetime inc. 2%	lifetime inc. 3%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
first employer size	0.0276***	0.0274***	0.0271***	0.0269***		0.1166**	0.1195**	0.1225**	0.1255**
	(0.0011)	(0.0011)	(0.0011)	(0.0011)		(0.0481)	(0.0486)	(0.0492)	(0.0497)
labor demand instrument					0.0953***				
					(0.0193)				
F-stat excl. instr.					24.33				
SE Clusters	661	661	661	661	661	661	661	661	661
Observations	79941	79941	79941	79941	79941	79941	79941	79941	79941

Notes: OLS and IV-TSLS estimates of the elasticity of lifetime income with respect to first employer size. Lifetime income defined as sum of total income after first job semester (defined in text) until age 35, using 0, 1, 2, and 3 percent annual discounting since age 16. Lifetime income, first employer size, and labor demand instrument in logs. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Columns (1)–(4) show the OLS estimates, by varying annual discount rate. Column (5) shows the first stage. Columns (6)–(9) show the IV-TSLS estimates, by varying annual discount rate. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Table A3: Lifetime income and first-employer size, controlling for sector of first employer

	OLS		First Stage		IV-TSLS	
	lifetime income (1)	lifetime income (2)	first employer size (3)	first employer size (4)	lifetime income (5)	lifetime income (6)
first employer size	0.0276*** (0.0011)	0.0224*** (0.0009)			0.1166** (0.0481)	0.1269** (0.0581)
labor demand instrument			0.0953*** (0.0193)	0.0766*** (0.0166)		
F-stat excl. instr.			24.31	21.39		
SE Clusters	661	661	661	661	661	661
Sector FEs	no	yes	no	yes	no	yes
Observations	79941	78538	79941	78538	79941	78538

Notes: OLS and IV-TSLS estimates of the elasticity of lifetime income with respect to first employer size. Lifetime income defined as sum of total income after first job semester (defined in text) until age 35. Lifetime income, first-employer size, and labor-demand composition instrument in logs. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Even-numbered columns also include fixed effects for the two-digit sector of a worker's first employer (58 sectors). Odd-numbered columns correspond to the baseline estimation in Table 2. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Table A4: Predicting lifetime income: Size and sector of first employer

included regressors	F-stat: Size	F-stat: Sector
size	642.57	-
sector	-	61.78
size + sector	562.79	50.64

Notes: Predictive power of a worker's first-employer size and/or sector towards explaining lifetime income. First row shows the F-statistic from the employer size coefficient of the OLS estimation of equation (2) in the text. Second row refers to the estimation of an equation similar to (2) that excludes first employer size, but includes fixed effects for the 2-digit sector of a worker's first employer (58 sectors). It reports the F-statistic from the joint test of significance for the sector fixed effects. Third row is based on estimating the same regression including both first-employer size and sector. It reports the F-statistics from the size coefficient and from the joint test of the sector fixed effects.

Table A5: Career outcomes and first-employer size: 1st-2nd job unemployment gap sample, OLS estimates

	OLS estimates						
	lifetime income (1)	average daily wage (2)	lifetime earnings (3)	days worked (4)	second employer size (5)	annual income age 35 (6)	employer size age 35 (7)
first employer size	0.0186*** (0.0011)	0.0227*** (0.0009)	0.0186*** (0.0012)	-0.0040*** (0.0006)	0.2992*** (0.0070)	0.0266*** (0.0022)	0.1757*** (0.0074)
SE Clusters	654	654	654	654	654	654	653
Observations	34507	34507	34507	34507	32965	33817	27881

Notes: OLS estimates of the elasticity of different long-term outcomes with respect to first-employer size. Estimated for the sample of workers who experience an unemployment gap between their first and second jobs (43% of lifetime sample). I count as unemployment gaps those that are at least 2 months long. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. All measures are in logs. Columns (1)–(7) show the elasticity for different long-term outcomes measured between labor market entry and the year a worker turns 35: (1) lifetime income as defined in equation (1), (2) average daily wage, (3) lifetime earnings (lifetime income excluding unemployment benefits), (4) total days worked, (5) size of second employer, (6) annual income during the year worker turns 35, (7) size of worker's employer during year he turns 35. Column (7) excludes workers who worked for less than half the days of the year they turn 35. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Table A6: Differential returns to experience at large employers: Monthly wages

	(1)	(2)	(3)	(4)
$bigExp$	113.8394***	82.3169***	114.3346***	81.3516***
	(4.1303)	(4.0479)	(4.1326)	(4.0685)
$bigExp \cdot Exp$	-0.0115***	-0.0058***	-0.0069***	0.0002
	(0.0008)	(0.0008)	(0.0010)	(0.0010)
$bigExp \cdot Tenure$			-0.0090***	-0.0113***
			(0.0008)	(0.0008)
Exp	194.5721***	201.2274***	193.3354***	199.8320***
	(3.5292)	(3.4946)	(3.5283)	(3.4933)
Exp^2	-0.0283***	-0.0297***	-0.0286***	-0.0300***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
$Tenure$	129.1841***	117.2113***	130.8840***	119.3581***
	(1.7242)	(1.7243)	(1.7204)	(1.7171)
$Tenure^2$	-0.0218***	-0.0195***	-0.0210***	-0.0184***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Current employer size category FE	no	yes	no	yes
Clusters (workers)	125232	125232	125232	125232
N (worker × month)	16198308	16198308	16198308	16198308

Notes: Dependent variable is log monthly wage. Experience and tenure measured in days. $bigExp$ is experienced acquired in employers with 250+ employees. Exp is overall experience (including $bigExp$). $Tenure$ equals days worked in current employer. Point estimates and standard errors displayed multiplied times 10^6 for readability. All specifications include worker fixed effects, age (quadratic), unemployment rate (quadratic), 21 sector fixed effects, fixed effects for 6 municipality/urban area size categories, fixed-term contract fixed effects, and month fixed effects. Municipality/urban area size categories group employers into a) municipalities with pop. less than 40,000, b) urban areas with pop. less than 125,000, c) 125,000–250,000, d) 250,000–500,000, e) 500,000–1,500,000, and f) 1,500,000+. Current employer size category fixed effects groups employers into a) missing size, b) 1–5 employees, c) 6–19, d) 20–49, e) 50–249, and f) 250+. Standard errors clustered at the worker level in parentheses. * 0.10 ** 0.05 *** 0.01.

Table A7: Career outcomes: Reduced-form estimates

	Lifetime income				Other outcomes		
	lifetime inc. 0%	lifetime inc. 1%	lifetime inc. 2%	lifetime inc. 3%	average daily wage	days worked	lifetime earn. 0%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
labor demand instrument	0.0111***	0.0114***	0.0117***	0.0120***	0.0078**	0.0027	0.0105**
	(0.0040)	(0.0040)	(0.0040)	(0.0041)	(0.0036)	(0.0016)	(0.0042)
SE Clusters	661	661	661	661	661	661	661
Observations	79941	79941	79941	79941	79941	79941	79941

Notes: OLS estimates of the elasticity of lifetime income and other outcomes with respect to the labor-demand composition instrument. Columns (1)–(4): Lifetime income defined as sum of total income after first job semester (defined in text) until age 35, using 0, 1, 2, and 3 percent annual discounting. Lifetime income, and labor demand instrument in logs. Column (5): Average daily wage defined as sum of total income over total days worked after first job semester (defined in text) until age 35. Column (6): Total days worked after first job semester (defined in text) until age 35. Column (7): Lifetime earnings defined as sum of total earnings after first job semester (defined in text) until age 35, using 0 percent annual discounting. Average daily wage, total days worked, lifetime earnings, and labor demand instrument in logs. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

B   Additional Results, Extensions, and Robustness Tests

B.1   Additional data sources

Throughout the paper I use additional data sources that complement the social security data. I compute the time series of regional unemployment rates using the Spanish Labor Force Survey (*Encuesta de Población Activa*, or EPA). Throughout the paper I use the male unemployment rate from each year's second-quarter wave. In some specifications I also use regional GDP growth rates, which come from the Spanish Regional Accounts provided by the Spanish National Statistics Institute. This same entity keeps the Central Business Register (*Directorio Central de Empresas*, or DIRCE). I use this data together with OECD data to provide descriptive statistics on the firm size distribution of Spain and other countries.

The EU Household Panel (*Panel de Hogares de la UE*, or PHOGUE) allows me to observe characteristics of workers' households at age 17, for a subset of the cohorts I study. This is something I take advantage of in an specification test for my IV approach.

I use the 2011 Survey on the Involvement of the Adult Population in Learning Activities (*Encuesta sobre la participación de la población adulta en las actividades de aprendizaje*, or EADA) to document the relationship between employer size and employer-provided training and education.

The World Management Survey allows me to document the relationship between managerial quality and firm size for a sample of Spanish manufacturing firms.

Finally, I use survey data collected by the Bank of Spain to study the relationship between firm size, R&D, and technology adoption. This survey is the Central Balance Sheet Data (*Central de Balances Anual*, or CBA). I use a sample of around 2,000 medium and large firms over the years 1991–2007 who agreed to share their survey responses with researchers.

B.2   IV specification check: No correlation with household characteristics at 17

I study the relationship between the labor-demand composition IV, $\bar{s}_{-i}^{rec}$, and the characteristics of workers' households *before* labor market entry, when they are 17 years old. A correlation between household characteristics and $\bar{s}_{-i}^{rec}$ would be consistent with violations of the exclusion restriction. Reassuringly, I find no evidence of such a relationship when looking at household income, parents' employment, parents' education, and type of father's employer. I carry out this test in the following way.

Using the EU Households Panel (*Panel de Hogares de la UE*, or PHOGUE) allows me to observe the relevant information for four birth cohorts from my sample (1977–1980). Collapsing the data to the $\{rec\}$ cell level I estimate the following regression:

$$\bar{s}^{rec} = Z'_{rec}\psi + f(u_{r,t_0(e,c)}) + \iota_r + \iota_e + \iota_c + \nu_{rec}. \tag{B1}$$

Where Z_{rec} includes (cell averages of) workers' household income, parents' employment, parents' education, whether father works for a large employer, and whether father works for public sector, all measured when the worker is 17 years old.[1] Appendix Figure B1 shows that estimates of ψ are not significantly different from zero at conventional levels and that I fail to reject the joint test $\psi = 0$.

---

[1] I use a more aggregate geographical region of birth (NUTS-1) since the NUTS-2 regions I use in the main analysis (*Comunidad Autónoma*) is not observed in PHOGUE. I also assign 4.9% of workers for whom I do not observe region of birth (those who are living outside it throughout the years I observe them) to cells based on region of residence at age 17.

Figure B1: IV specification check: Instrumental Variable and Cohort Household Characteristics
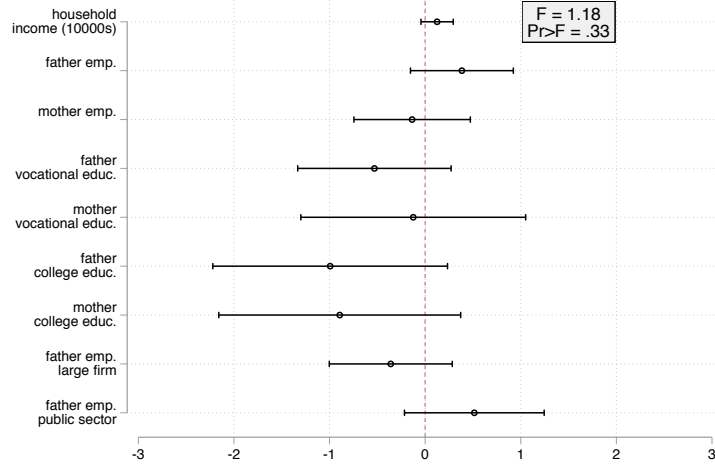
Notes: Point estimates and 95% confidence intervals of a regression of the cell-level instrument $\bar{s}_{r,e,t_0(e,c)}$ on workers' household characteristics when they are age 17 (shown in the figure), a flexible function of regional unemployment rate on predicted graduation year, region of birth fixed effects, cohort fixed effects, and educational attainment fixed effects. F-statistic and p-value for the joint test of non-significance for the nine coefficients above. Region of birth r aggregated to the NUTS-1 level (as opposed to NUTS-2 in main analysis). N=82 cells, observations weighted by number of workers in each MCVL cell. Data source for household characteristics is the EU Households Panel (*Panel de Hogares de la UE*).

B.3 IV specification check: No relationship between IV and educational investment decisions

I test for potential endogenous responses of educational investment decisions to the large-firm demand shocks that the IV leverages. I check for this possibility studying whether, after controlling for unemployment rates, regional labor demand composition influences education investment decisions. To do this, I follow the logic behind the index $\bar{s}_{-i}^{rec}$, and construct indices reflecting the labor demand composition that each worker would face at age 17 (high school predicted graduation), and at age 20 (vocational predicted graduation) in his region of birth. I then test whether these indices predict further educational investments estimating the following linear probability models:

$$\mathbb{1}\{educ_i > HS\} = \gamma \bar{s}_{-i}^{r,t_{17}} + f(u_{r,t_{17}(c)}) + \iota_r + \iota_c + \eta_i \tag{B2}$$

$$\mathbb{1}\{educ_i > Voc\} = \psi_1 \bar{s}_{-i}^{r,t_{20}} + \psi_2 \bar{s}_{-i}^{r,t_{17}} + f(u_{r,t_{20}(c)}) + f(u_{r,t_{17}(c)}) + \kappa_r + \kappa_c + \nu_i. \tag{B3}$$

Where $\mathbb{1}\{educ_i > HS\}$ and $\mathbb{1}\{educ_i > Voc\}$ are dummy variables that equal one if person i holds a vocational or college degree, or a college degree, respectively. $\bar{s}_{-i}^{r,t_{17}}$ is the (log) average first-employer size of workers with high school educational attainment, who are getting their first job in the year person i turns 17, in his region of birth. Similarly, $\bar{s}_{-i}^{r,t_{20}}$ is the (log) average first-employer size of workers with vocational educational attainment, who are getting their first job in the year person i turns 20, in his region of birth. Both indices, again, follow a leave-one-out approach. $u_{r,t_{17}(c)}$ and $u_{r,t_{20}(c)}$ are the regional unemployment rates at i's region of birth in the years he turns 17 and 20, respectively. The ιs and κs are birth region and cohort fixed effects.

Large and statistically significant estimates of γ, ψ_1, and/or ψ_2 would be worrying, indicating an endogenous labor supply response (in the form of educational investments) to the variation the IV approach uses. Appendix Table B1 shows the parameter estimates

for different specifications of equations (B2) and (B3). Reassuringly, the three coefficient estimates, across different specifications, are small and insignificant. Thus, I fail to reject the null hypothesis of no educational investment responses to the IV residual variation.

Table B1: IV residual variation does not predict educational investments: OLS estimates

	$Pr(educ > HS)$		$Pr(educ > Voc)$			
	(1)	(2)	(3)	(4)	(5)	(6)
labor demand	-0.0053	-0.0028			-0.0023	-0.0044
composition at 17	(0.0043)	(0.0052)			(0.0031)	(0.0055)
labor demand			-0.0036	-0.0071	-0.0033	-0.0065
composition at 20			(0.0028)	(0.0051)	(0.0027)	(0.0050)
Baseline prob.	0.569	0.486	0.162	0.285	0.162	0.285
SE Clusters	221	221	221	221	221	221
Sample (educ.)	all	HS & Voc.	all	Voc. & college	all	Voc. & college
Observations	79941	66998	79941	45486	79941	45486

Notes: OLS estimates of different specifications of equations (B2) and (B3) in the text. Dependent variable in Columns (1)–(2) is a dummy that equals 1 if a worker has an educational attainment higher than high school (i.e. vocational or college). Dependent variable in Columns (3)–(6) is a dummy that equals 1 if a worker has an educational attainment higher than vocational (i.e. college). All specifications include region-of-birth and birth-cohort fixed effects, and a quartic in the unemployment rate in the worker's region of birth when he is 17 years old. Columns (3)–(6) control in the same way for unemployment at age 20. Labor demand composition at 17 (20) is an index capturing the prevalence of large firms' labor demand in a worker's region of birth when he is age 17 (20), further described in the text. Column (2) excludes from the sample workers who eventually achieve a college degree. Columns (4) and (6) exclude from the sample workers whose highest educational attainment is high school. Standard errors clustered at the level of region of birth $\times$ birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

B.4 Lifetime income IV result: Additional robustness tests

In this section I show that the IV estimate of the elasticity of lifetime income with respect to first-employer size, discussed in Section 4.4, is robust. Appendix Figure B2 gathers the resulting IV elasticity estimates when using alternative specifications. Additionally, it shows results when discounting the measure of lifetime income (baseline estimates use the measure with no discounting). The black and round marker shows the baseline results from column (6) in Table 2.

Alternative flexible unemployment rate function. I change the way that I control for the regional unemployment rate during predicted graduation year. The white markers in Appendix Figure B2 show the estimates when I control for unemployment rate using a categorical piece-wise function (as opposed to the baseline quartic).[2] The estimates are very similar to baseline.

Additional cyclical indicator. The next robustness check involves a specification where in addition to flexibly controlling for regional unemployment rates, I also control flexibly for regional GDP growth during a worker's predicted graduation year in his region of birth. This is meant to address the fact that the unemployment rate is a single indicator that could imperfectly capture business cycle variation. Including a second indicator should diminish related concerns. The gray markers in Appendix Figure B2 show the estimates under this specification, which are almost identical to baseline.

Past business cycle conditions. One could worry that educational attainment, which I control for and use in the IV strategy, could endogenously be related to past business cycle conditions. The baseline specification simply controls for business cycle conditions at the time of predicted graduation. I estimate an alternative specification which controls for past

---

[2]I bin the unemployment rate into 3 categories (low, medium, high), include fixed effects for each of these categories, and allow the fixed effects to vary by educational attainment. The cutoffs for the three categories are 11% and 16% and are based on the worker-level distribution of regional unemployment rates at the time of graduation, roughly dividing workers equally between the three categories.

unemployment at a workers' region of birth, controlling for the unemployment rates at the years in which college (vocational) workers would have graduated from high school and vocational (high school) education. This is meant to capture unemployment conditions not only at the time of actual labor market entry, but at times when workers were potentially making educational investment decisions. The green markers in Appendix Figure B2 show the estimates under this specification. The results are very similar to baseline.

Business cycle conditions during year of labor market entry. The main specification controls for business cycle conditions during a workers' year of predicted graduation. This is meant to avoid the endogenous entry decisions that factor into the actual year of labor market entry. However, one might worry that a large-firm labor demand shock during the year of predicted graduation (captured by the IV) could impact business cycle conditions in following years and, through that channel, have a direct impact on workers' outcomes other than through their first employer. To allay this concern, I estimate a specification that, in addition to flexibly controlling for the unemployment rate during the year of predicted graduation, it also controls flexibly for the unemployment rate during the year of labor market entry. The red markers in Appendix Figure B2 show the estimates using this additional control. These elasticities are very similar to baseline.

Large employers and growing employers. Measuring employer size at the time the worker joins the firm could conflate having a first job at a large employer with having a first job at an employer which is doing well and growing in size. To address this distinction I estimate an alternative specification using a different measure of first-employer size. Instead of using employer size at the time of joining the firm, I use an average over the four years prior to the year the worker joined.[3] The orange markers in Appendix Figure B2 show the estimates using this measure. These elasticities are also very similar to baseline.

Sector of first employer. Firm sizes differ across sectors of activity. One might worry that the first-employer size effect is conflated with the effect of holding a first job at one or another sector. I address this concern by estimating a specification of equation (2) that explicitly controls for the sector of a worker's first employer. I use a two-digit definition, with 58 different sectors. The blue markers in Appendix Figure B2 show the first employer size elasticity estimates under this specification. The results are very similar to baseline.[4]

Finer geographical control. The baseline specification in equation (2) includes region-of-birth fixed effects (17 regions). Regions in Spain are further divided into 50 provinces (which is also the geographical level in which an employer—firm-times-province—is defined in the data). To check that my results are not driven by persistent differences of workers and employers across provinces within regions, I estimate equation (2) with birth province fixed effects rather than region. The pink markers in Appendix Figure B2 show the first employer size elasticity estimates under this specification and the results are practically identical to baseline.

Provincial size of first job. Larger employers are typically located in more populated areas. One could argue that this is part of the set of attributes defining large firms. However, we would like to know if the first-employer size premium is simply driven by geographical effects of more populated areas. I test for this possibility by estimating equation (2) with an additional control: (log) population of the province where the worker held his first job. The results from this specification are represented by the brown markers in Appendix Figure B2

---

[3]In a small number of cases the data for a given firm does not go back enough. When this happens I average over the amount of prior years of data available.

[4]Appendix Table A3 shows the OLS, first stage, and TSLS results when using sector fixed effects, together with the baseline results for comparison. Interestingly, when focusing on the OLS I find that the predictive power (predicting lifetime income) of first employer size is an order of magnitude larger than that of first employer sector. Appendix Table A4 shows this. It displays the F-statistic of the first employer size coefficient, and that of the joint test of significance of the 58 sectors for OLS regressions of equation (2).

and are essentially identical to baseline.

   Region-specific linear time trends. I estimate an expanded version of equation (2) which in addition to region and cohort fixed effects includes region-specific linear time (cohort) trends. The results from this specification are represented by the yellow markers in Appendix Figure B2. Point estimates are equal to baseline although standard errors are larger.

Figure B2: IV elasticity of lifetime income w.r.t. first-employer size: robustness
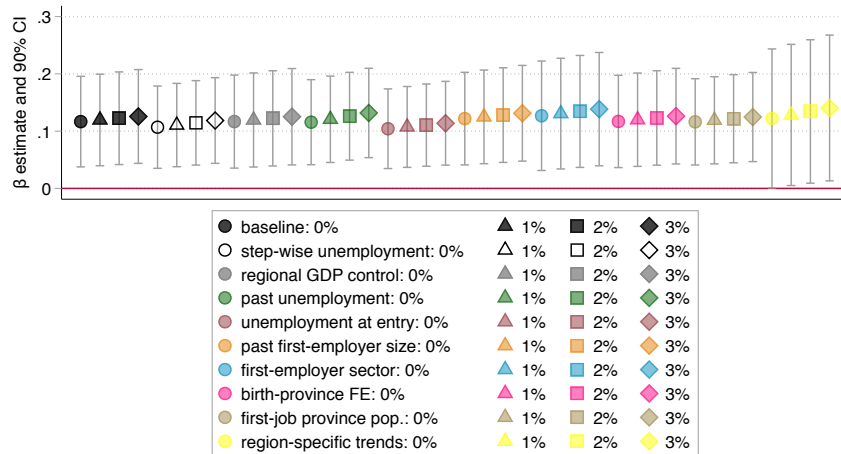
Notes: Point estimates and 95% confidence intervals of the IV TSLS elasticity of lifetime income with respect to first employer size using varying specifications of equation (2) in the text. Different marker shapes correspond to different annual discount factors in lifetime income computation. Black markers: baseline results coinciding with those in Table 2 columns (6)–(9). White markers: using a step-wise function of regional unemployment instead of the baseline quartic function. Gray markers: controlling for regional GDP growth during predicted graduation year at region of birth in addition to unemployment. Green markers: controlling for the unemployment rates in years previous to predicted graduation; for college (vocational) workers this includes the unemployment rate present when they would have graduated from high school and vocational (high school) education. Red markers: in addition to controlling for a flexible function of the unemployment rate at the time of predicted graduation, I control for a flexible function of the unemployment rate during the actual year of labor market entry. Orange markers: worker's first employer size measured as the average size over the four years prior to worker's hiring. Blue markers: controlling for sector of first employer (58 sector fixed effects). Pink markers: including province-of-birth fixed effects (instead of region-of-birth). Brown markers: controlling linearly for (log) population of province-year of first job. Yellow markers: including region-specific linear time trends.

B.5 Lifetime result robustness check: Uncensored income using tax data

As I discuss in Section 3, the monthly earnings measure in social security data is censored. I have followed a procedure similar to Bonhomme and Hospido (2017) to impute monthly earnings for censored observations.[5] While censored observations are few (8.7% and 3% of observations in the monthly panel are top- and bottom-coded respectively), one could wonder about the sensitivity of the main results to the imputation procedure.

A feature of the MCVL data is that social security records are also linked to tax data. The benefit of the tax data is that it provides measures of uncensored annual income. The downside is that, as opposed to social security earnings, tax data does not go back in time retrospectively. Tax earnings data are contemporaneous to each MCVL round, and thus available from 2005 onwards.[6]

Tax earnings data from 2005–2015 do not allow computing a lifetime income measure like the one in the main analysis. To test for robustness of the lifetime result using uncensored income data, I compute a measure of aggregate income earned during the eleven calendar years available in tax data:

$$Y_i^{05-15} = \sum_{t=2005}^{2015} y_{it}. \tag{B4}$$

Where y_{it} is the income person i earns in year t (in 2016 Euro).[7] The age at which this income is earned will vary across cohorts in my sample. The oldest (youngest) cohort, born in 1968 (1980), earns Y^{05-15} between the ages of 37 and 47 (25 and 35). I am able to compute this measure for 97% (77,754 workers) of my main analysis sample.

I estimate the elasticity of Y^{05-15} with respect to a worker's first employer size by estimating equation (2), using $\ln(Y_i^{05-15})$ as dependent variable.[8] Appendix Table B2 shows the OLS, first stage, and IV-TSLS results. It is reassuring to see that the estimated elasticities are very similar in magnitude to those in Table 2. OLS is equal to .0289 compared to .0269–.0276 in Table 2, IV is equal to .1408 compared to .1166–.1255.

The second check I carry out using uncensored tax data is to replicate the elasticity of income at age 35 with respect to first employer size (see Table 5). This replication is directly comparable since the tax data allows me to compute annual income at age 35 for 11 out of the 13 cohorts in my sample.[9] Appendix Table B3 shows that the results using tax data are very similar to those using social security data. The OLS is equal to .026 compared to .037 in Table 5. Reassuringly, the IV estimates are practically the same, .085 compared to .089.

---

[5]This involves grouping worker-month observations into 5,480 cells c { professional category $\times$ age $\times$ quarter} and parametrically model earnings within-cell while imposing no restrictions across cells. I assume log-normality within each cell and estimate the parameters μ_c and σ_c^2 using maximum likelihood. I then use these parameters to simulate earnings observations for bottom- and top-coded observations.

[6]They are also not available for residents of Navarre and the Basque Country since these regions have independent tax authorities.

[7]This measure of annual income includes labor earnings and unemployment benefits, as well as other sources of income such as business income and self-employed earnings.

[8]I exclude from the estimating sample 3,185 workers (4% of total) with $Y^{05-15} \leq 26,400$ Euro. 26,400 Euro amounts to average monthly earnings of 200 Euro, roughly half of the unemployment non-contributive subsidy. I am likely missing earnings data from these workers, who might either be working most of these years in Navarre, the Basque Country, or abroad.

[9]Again, I also exclude those with annual income at age 35 less than 2,400 Euro, equivalent to 200 Euro per month. These are 2.4% of total workers.

Table B2: Total 2005–15 tax income and first-employer size: OLS and IV-TSLS estimates

	OLS	First Stage	IV-TSLS
	income 2005-15 (1)	first employer size (2)	income 2005-15 (3)
first employer size	0.0289*** (0.0012)		0.1408** (0.0718)
labor demand instr.		0.0953*** (0.0205)	
F-stat excl. instr.		21.67	
SE Clusters	661	661	661
Observations	74569	74569	74569

Notes: OLS and IV-TSLS estimates of the elasticity of total 2005–15 tax data income with respect to first employer size. Income, first job size, and labor demand instrument in logs. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Standard errors clustered at the level of region of birth $\times$ education $\times$ birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Table B3: Annual tax income during age 35 and first-employer size: OLS and IV-TSLS estimates

	OLS	First Stage	IV-TSLS
	annual income age 35 (1)	first employer size (2)	annual income age 35 (3)
first employer size	0.0260*** (0.0012)		0.0853* (0.0436)
labor demand instr.		0.1434*** (0.0248)	
F-stat excl. instr.		33.39	
SE Clusters	561	561	561
Observations	60971	60971	60971

Notes: OLS and IV-TSLS estimates of the elasticity of tax data annual income at age 35 with respect to first employer size. Income, first job size, and labor demand instrument in logs. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Standard errors clustered at the level of region of birth $\times$ education $\times$ birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

B.6 Varying elasticity of lifetime income with respect to first-employer size

In this section I relax the constant elasticity assumption of Section 4, implicit in equation (2), and originally motivated by the linear-in-logs raw relationship between first-employer size and lifetime income (see Figure 1). Relaxing this assumption allows the possibility that increments in first-employer size are differentially valuable across the employer-size distribution. I estimate the following equation which allows a quartic polynomial:

$$y_i = \beta_1 s_{J(i)} + \beta_2 s_{J(i)}^2 + \beta_3 s_{J(i)}^3 + \beta_4 s_{J(i)}^4 + \delta' X_i + \varepsilon_i. \tag{B5}$$

Where y_i is log lifetime income and $s_{J(i)}$ is the log size of worker i's first employer. The covariates $X_i = [f(u_{r,t_0(e,c)}), \delta_r, \delta_c, \delta_e]'$ are, respectively, a flexible function of the regional unemployment rate at the time of predicted graduation, region of birth fixed effects, birth cohort fixed effects, and educational attainment fixed effects. They coincide with those in Section 4.

In this case, the elasticity of lifetime income with respect to first-employer size can vary across firm sizes and, for log size s, is equal to

$$\epsilon(s) = \beta_1 + 2\beta_2 s + 3\beta_3 s^2 + 4\beta_4 s^3. \tag{B6}$$

I follow Florens et al. (2008) and estimate (B5) using a control function approach. This involves first estimating the OLS first stage

$$s_{J(i)} = \gamma \bar{s}_{-i}^{rec} + \phi' X_i + \nu_i. \tag{B7}$$

Then, using the estimated residuals $\hat{\nu}_i$ for the control function approach.[10] Following Florens et al. (2008) I use a control function that interacts $\hat{\nu}_i$ with the polynomial of first-employer size $s_{J(i)}$. Thus, the elasticity parameters of interest in (B5) can be estimated by OLS in

$$y_i = \beta_1 s_{J(i)} + \beta_2 s_{J(i)}^2 + \beta_3 s_{J(i)}^3 + \beta_4 s_{J(i)}^4 + \delta' X_i + \sum_{l=0}^{4} \kappa_l \hat{\nu}_i s_{J(i)}^l + \varepsilon_i. \tag{B8}$$

In practice I estimate (B7) and (B8) jointly by non-linear least squares to obtain correct standard errors, clustered at the $\{rec\}$-cell level. Results are shown in Appendix Table B4. Using these estimates I compute the elasticity function (B6) and its standard error. This is shown in Appendix Figure B3. The estimated elasticity features an interesting non-linearity: it is small and statistically non-significant for the lower part of the firm size distribution; it increases up until it reaches its maximum around log size equal to 5 (80th percentile of the empirical first-employer size distribution), and decreases thereon. For the very high part of the firm size distribution (log size 7, 95th percentile of the empirical first-employer size distribution) it is again relatively small and non-significant.

The interpretation of this pattern seems intuitive. Conditional on starting out at a very small firm, differences in size do not matter that much. The same is true conditional on starting at very large firms. There is, however, a mid-high part of the firm-size distribution, where increments in the first-employer size seem to make a substantial difference. This could be capturing the difference between starting out in a middle-size employer or one of the, relatively few, large Spanish employers.

---

[10]The control function approach avoids using polynomials of the IV, $\bar{s}_{-i}^{rec}$, as if they were additional instruments. It comes, however, with additional assumptions relative to TSLS. In particular, it requires the instrument to be independent of unobservables rather than simply uncorrelated. It also imposes a linearity restriction on the conditional expectation $\mathbf{E}(\varepsilon|\nu)$.

Figure B3: Varying elasticity of lifetime income with respect to first-employer size
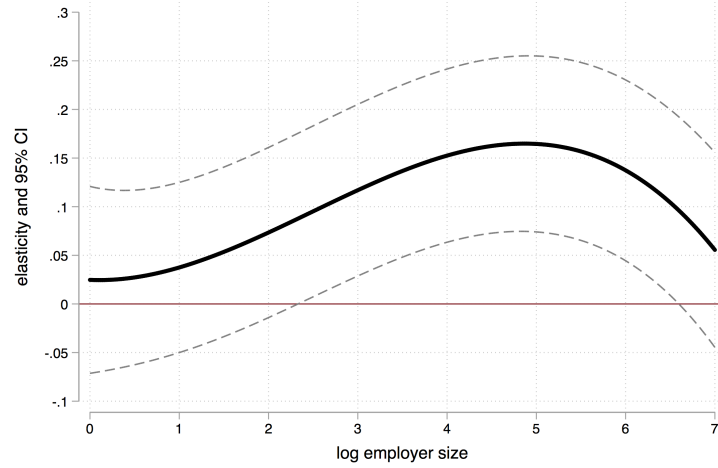
Notes: Varying elasticity of lifetime income with respect to first-employer size (defined in equation (B6)) and 95% confidence interval. Based on parameter estimates of equation (B5). Elasticity is equal to $\hat{\epsilon}(s) = \hat{\beta}_1 + 2\hat{\beta}_2 s + 3\hat{\beta}_3 s^2 + 4\hat{\beta}_4 s^3$. Function plotted until $s = 7$ which is the 95th percentile of the empirical distribution. Standard errors are clustered at the level of region of birth $\times$ education $\times$ birth cohort. Standard error of $\hat{\epsilon}(s)$ computed using the delta-method.

Table B4: Varying elasticity estimates: Control function approach

parameter	point estimate (std. error)
β_1	0.0248
	(0.0490)
β_2	-0.0020
	(0.0112)
β_3	0.0065***
	(0.0021)
β_4	-0.0006***
	(0.0001)
γ	0.0945***
	(0.0178)
SE Clusters	661
Observations	79941

Notes: Coefficient estimates and standard errors of the parameters of the varying elasticity of lifetime income with respect to first employer size. Estimated using a control function approach detailed in equations (B7) and (B8). Results obtained from estimating these two equations jointly using non-linear least squares. Standard errors clustered at the level of region of birth $\times$ education $\times$ birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

B.7 Job security: Temporary and permanent contracts

Like other European countries, Spain features a "dual" labor market, with a stark difference between permanent and temporary labor contracts (see Dolado et al., 2002). Given this, it is interesting to test for a relationship between first-employer size and job security later on in the working life. The interpretation of this type of analysis, however, requires some nuance. In particular, young workers could face a trade-off between a job offering high security and a job opening up future opportunities (getting "stuck" in a bad job).

The Spanish social security data include information on labor contract types, which allows me to investigate whether there is a link between the size of a worker's first employer and the subsequent prevalence of temporary vs. permanent contracts. Type of contract starts being recorded in my data in 1991 and it is missing in large proportions until 1998.[11] The oldest cohort in my sample was born in 1968, which motivates focusing on job security between the ages of 30 and 35.

Figure B4 shows the prevalence of temporary contracts for workers in my sample when they are between 30 and 35 years old.[12] 45% of workers never work under a temporary contract in this period. By contrast, 12% work exclusively under temporary contracts while aged 30–35. The remaining 43% of people work under both types of contract during this period.

I construct two indices capturing aspects of the job security a worker experiences between the ages of 30 and 35. The first one simply characterizes the extensive margin of temporary employment. This index is a dummy variable that equals one if a person ever worked (between ages 30–35) under a temporary contract, and zero otherwise.

The second index combines information on type of contract and employment. It captures whether the worker experiences, between the ages of 30 and 35, what I call *total job security*. I encapsulate this with a dummy variable that equals one if a worker, between 30–35, never works under a temporary contract and experiences non-employment for no more than 30 days. 33% of workers in my sample experience *total job security*.

To test for the link between first-employer size and these two indices, I use them as outcome variables in OLS and IV estimations of equation (2). Table B5 shows the results from this exercise. Columns (1) and (2) show that in OLS first-employer size does a good job at predicting job security experienced between ages 30–35. Starting the working life in a larger employer is significantly correlated with a lower probability of working under temporary contracts during the 30s (column (1)), and a higher probability of experiencing *total job security* more broadly (column (2)). Columns (4) and (5) show the equivalent IV results. The message is the similar as in OLS, although the estimates are somewhat imprecise. Column (4) indicates a negative causal effect between having a larger first employer and the probability of working later on under temporary contracts. Equivalently, column (5) shows a positive IV effect of first-employer size on the probability of achieving *total job security*, although the estimate is not statistically significant at conventional levels.

To the extent that job security is uncorrelated or positively correlated with employer quality, results from Table B5 suggest an additional channel through which the characteristics of a young worker's first employer can positively impact her later career trajectory. However, the interpretation of this result is less clear if there is a negative correlation between employer quality and job security.

---

[11] By contrast my earnings panel underlying lifetime income measures starts in 1984.

[12] Given that I pay attention to the interval between ages 30 and 35, in this section I focus on those who work for at least half the days in these six years. I also require that information on type of contract is missing for no more than one third of their days worked during these six years. These restrictions result in a sample of 68,614 workers, 86% of the original lifetime sample.

Figure B4: Fraction of days worked under temporary contract between ages 30–35
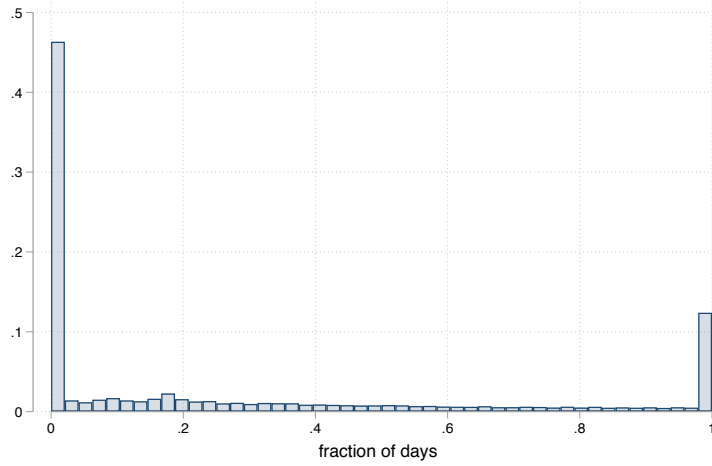
Notes: Distribution of the fraction of days worked under a temporary contract between the ages of 30 and 35. Workers in the lifetime analysis sample who, between the ages of 30 and 35, work for at least half the days and are missing information on type of contract for no more than one third of their days worked. $N = 68,614$ workers.

Table B5: Job security between ages 30–35 and first-employer size: OLS and IV-TSLS estimates

	OLS		First Stage	IV-TSLS	
	temporary contract (=1) (1)	total job security (=1) (2)	first job size (3)	temporary contract (=1) (4)	total job security (=1) (5)
first job size	-0.0134***	0.0098***		-0.0640*	0.0543
	(0.0011)	(0.0009)		(0.0372)	(0.0351)
labor demand instr.			0.0967***		
			(0.0189)		
F-stat excl. instr.			26.21		
LHS var. average	0.55	0.33		0.55	0.33
SE Clusters	661	661	661	661	661
Observations	68614	68614	68614	68614	68614

Notes: OLS and IV-TSLS estimates of β in equation (2), using two indices of job security as outcome variable. Outcome variable in columns (1) and (4) is a dummy variable that equals one if a person ever worked under a temporary contract between ages 30–35. Outcome variable in columns (2) and (5) is a dummy variable that equals one if a worker, between 30–35, never works under a temporary contract and experiences non-employment for no more than 30 days. Regressions includes 86% of workers from main sample who, between ages 30–35, were (i) employed for at least half the days, and (ii) no more than one third of their type-of-contract information is missing. First job size, and labor demand instrument in logs. Regressions at the worker level. All regressions control for a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

B.8   Additional evidence on persistence of first-employer effects

This section provides additional evidence on the persistent of first-employer effects, complementing the findings of Section 5.1 in the main text.

Time-varying elasticity of income with respect to first-employer size

I estimate a time-varying analogue of the elasticity of lifetime income with respect to first-employer size. Using the data in a quarterly panel format, and using quarterly income as dependent variable, I allow the elasticity of a worker's first employer's size to follow a

time trend by estimating the following equation:

$$y_{iq} = (\beta_1 + \beta_2 \cdot q + \beta_3 \cdot q^2) \cdot s_{J(i)} + X'_{iq}\gamma + \varepsilon_{iq}. \tag{B9}$$

Where y_{iq} is the log of quarterly income of worker i, q quarters after labor market entry. The β coefficients allow the elasticity with respect to first employer size, $s_{J(i)}$, to follow a quadratic trend. The vector X_{iq} includes a series of controls whose coefficients are also allowed to vary across time.[13] Appendix Table B6 shows the implied elasticities at different points in time (Appendix Table B7 shows the underlying β estimates). Appendix Table B6 allows a quadratic trend as in equation (B9), or imposing a linear trend (assuming $\beta_3 = 0$).

Table B6: Time-varying elasticity of income and first-employer size: Values at different points in time

Years after entry	Elasticity: Linear trend	Elasticity: Quadratic trend
3	0.0262 (0.0357)	0.0205 (0.0360)
6	0.0564 (0.0367)	0.0357 (0.0382)
9	0.0866** (0.0389)	0.0825** (0.0393)
12	0.1167*** (0.0419)	0.1608*** (0.0410)

Notes: Elasticity of quarterly income with respect to first-employer size at different points in time after labor market entry. Based on IV-TSLS estimates of equation (B9) in the text, shown in Appendix Table B7. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Table B7: Quarterly income and time-varying first-employer size elasticity

	OLS		IV-TSLS	
	(1)	(2)	(3)	(4)
first employer size	0.0509*** (0.0019)	0.0640*** (0.0026)	-0.0040 (0.0359)	0.0369 (0.0372)
first employer size $\times q$	-0.0006*** (0.0001)	-0.0023*** (0.0002)	0.0025*** (0.0005)	-0.0027 (0.0017)
first employer size $\times q^2$		0.0000*** (0.0000)		0.0001*** (0.0000)
Trend	linear	quadratic	linear	quadratic
SE Clusters	661	661	661	661
N (worker × quarter)	3569662	3569662	3569662	3569662

Notes: OLS and IV-TSLS estimates of the time-varying elasticity of quarterly income with respect to first-employer size outlined in equation (B9). Regressions at the worker×quarter level. Dependent variable is log total quarterly income, and q is the number of quarters passed since labor market entry. All regressions control for a flexible function of the regional unemployment at predicted graduation year, birth-cohort fixed effects, and education fixed effects. All these controls are allowed to vary across quarters. Also control for region-of-birth fixed effects, and quarter fixed effects. Columns (1) and (3) allow a linear time trend while Columns (2) and (4) allow a quadratic one. TSLS estimates in Columns (2)–(3) use as instrument the labor demand instrument described in the text and the same instrument interacted with q and q^2. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

The β estimates in Appendix Table B7 indicate an increasing first-employer size effect. This is true for both the linear and quadratic time trends, and it implies that a larger first

---

[13]It includes a quartic function of the regional unemployment at predicted graduation year, birth cohort fixed effects, and education fixed effects. All these controls are allowed to vary across quarters. Finally, I also include region of birth fixed effects, and quarter fixed effects.

employer results in higher earnings growth. Focusing on the linear trend, Appendix Table B6 shows that the time-varying elasticity three and six years after labor market entry is 0.026 and 0.056 although imprecisely estimated. Nine years after labor market entry this value is 0.087, and 12 years after it is the same value as the baseline lifetime elasticity, 0.117. The quadratic time trend delivers qualitatively similar results, although the implied elasticity twelve years after entry is somewhat larger, equal to 0.161.

Wage growth between the first and second job

One could still wonder whether persistence results are driven by the small fraction of people who stay with their first employer throughout this time period.[14] To address this, I test whether persistent first employer effects still arise when explicitly taking into account job mobility and initial wages at different jobs. I test whether workers with larger first employers experience greater wage growth when moving to their second job, holding constant first job tenure and second employer size. I do this by estimating

$$g_i^{1,2} = \beta_1 s_{J_1(i)} + \beta_2 s_{J_2(i)} + \rho \ln(\bar{w}_{i1}) + f_1(tenure_i^1) + f_2(tenure_i^2) + g(unemp_i^{1,2}) + X_i'\gamma + \varepsilon_i. \quad \text{(B10)}$$

Where $g_i^{1,2} \equiv \ln(\bar{w}_{i2}) - \ln(\bar{w}_{i1})$ is the growth rate between the average daily wage worker i earned in his second job ($\bar{w}_{i2}$) and the one he earned in his first job ($\bar{w}_{i1}$). $s_{J_1(i)}$ and $s_{J_2(i)}$ are log employer size for the first and second employers, $tenure_i^j$ is the amount of days i worked at his jth employer, $unemp^{1,2}$ controls for the existence and length of an unemployment spell between the first and second jobs, and X_i includes the same controls as equation (2) in addition to start of second job year dummies.[15]

I estimate different specifications of equation (B10). Appendix Table B8 shows OLS and IV-TSLS estimates of β_1. The OLS estimates are small, negative, and close to zero (though precisely estimated). The IV estimates are positive indicating an elasticity of between-job wage growth and first employer size of .09–.11. Thus, it seems that returns to a larger first employer already arise in the form of higher wage growth when moving from the first to the second job.

---

[14]6.8% of the workers in the sample remain in their first job until the year in which they reach age 35.

[15]While results from this regression are informative, they are somewhat descriptive in nature. This is because in spite of having a valid instrument providing exogenous variation in first-employer size, I lack additional instruments for (i) if and when a worker separates from his first employer, and (ii) second-employer size. Controlling for $\bar{w}_{i1}$ addresses at some level unobserved worker heterogeneity, but concerns related to selection and bad controls still remain.

Table B8: Between-job wage growth and first-employer size: OLS and IV-TSLS estimates

	OLS			First Stage			IV-TSLS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
first employer size	-0.0038***	-0.0056***	-0.0007				0.1072**	0.1021**	0.0938**
	(0.0011)	(0.0011)	(0.0010)				(0.0496)	(0.0480)	(0.0425)
labor demand instr.				0.0824***	0.0847***	0.0856***			
				(0.0152)	(0.0155)	(0.0153)			
F-stat excl. instr.				29.48	29.96	31.2			
U-E transition	no	yes	yes	no	yes	yes	no	yes	yes
Tenure 2nd job	no	no	yes	no	no	yes	no	no	yes
SE Clusters	661	661	661	661	661	661	661	661	661
Observations	72742	72742	72742	72742	72742	72742	72742	72742	72742

Notes: Dependent variable is the growth rate between the average daily wage a worker receives in his second job and that from his first job. All regressions control for second employer size, log average daily wage in first job, tenure (in days) at first job, start year at second job, a flexible function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. All employer size variables (first, second, instrument) are in logs. Columns (1)–(3) show the OLS estimates. Columns (7)–(9) show IV-TSLS estimates, instrumenting for first-employer size using the labor-demand composition index defined in the text. Columns (4)–(6) show the respective first stage. U-E transition controls for the existence and (cubic) length of an unemployment spell between the first and second jobs. Tenure 2nd job is a cubic of tenure at second job and a dummy variable capturing whether this tenure is censored or not. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

B.9 Mechanisms: Job search and human capital in a simple framework

This section complements the discussion in Section 5.2. I provide a simple framework that illustrates how first-employer persistent effects can arise through job search and human capital channels. I first focus on pure search and then add an on-the-job skill component.

Search

Consider workers who are matched to firms with varying desirability u, drawn from the distribution $F(u)$ with support $[\underline{u}, \bar{u}]$. The desirability index u could be the wage the worker receives in a given firm, or more generally capture additional traits of the firm workers' value. Search frictions imply that workers receive offers each period with probability λ. Then, the value of employment in period t at a firm with desirability u_t is given by

$$V_t(u_t) = u_t + \beta\Big[\lambda\mathbf{E}\big[\max\{V_{t+1}(u_t), V_{t+1}(u)\}\big] + (1 - \lambda)V_{t+1}(u_t)\Big], \tag{B11}$$

where the expectation is taken with respect to $F(u)$. Since search opportunities are common across firms, a worker will accept an offer u if $u > u_t$. Hence:

$$\mathbf{E}\big[\max\{V_{t+1}(u_t), V_{t+1}(u)\}\big] = F(u_t)V_{t+1}(u_t) + \int_{u_t}^{\bar{u}} V_{t+1}(u)f(u)du. \tag{B12}$$

It is straightforward to see that job desirability in a given period will be positively related to past desirability. First, the expected value of tomorrow's desirability as a function of today's is given by:

$$\mathbf{E}(u_{t+1}|u_t) = \big[(1 - \lambda) + \lambda F(u_t)\big] \cdot u_t + \lambda\big(1 - F(u_t)\big) \cdot \mathbf{E}(u|u > u_t). \tag{B13}$$

It follows that:[16]

$$\frac{\partial}{\partial u_t}\mathbf{E}(u_{t+1}|u_t) = (1 - \lambda) + \lambda F(u_t) > 0. \tag{B14}$$

An important point is that involuntary unemployment cuts this job-ladder persistence. Consider the same framework, augmented to allow for involuntary job separation. Each period, a match is dissolved with exogenous probability δ. In this case, the value of employment in period t at a firm with desirability u_t is given by

$$V_t(u_t) = u_t + \beta\Big[(1-\delta)\lambda\mathbf{E}\big[\max\{V_{t+1}(u_t), V_{t+1}(u)\}\big] + (1-\delta)(1-\lambda)V_{t+1}(u_t) + \delta D_{t+1}\Big]. \tag{B15}$$

Where D_t is the value of being unemployed. Normalizing the flow value of unemployment to zero,

$$D_t = \beta\Big[\lambda\mathbf{E}\big[V_{t+1}(u)\big] + (1 - \lambda)D_{t+1}\Big]. \tag{B16}$$

This illustrates that when an unemployed worker finds a job, she samples from the unconditional distribution of desirability $F(u)$. Thus, the desirability of subsequent jobs after the unemployment spell will be unrelated the desirability of previous jobs.

Human capital

Now consider that instead of a general desirability index, workers simply value earnings. Worker earnings in period t are given by $Y_t = RK_t$, where K_t is human capital at time t and R is the rental rate, assumed to be the same across employers. Firms differ in the

---

[16]Using the fact that $\frac{\partial}{\partial u_t}\mathbf{E}(u|u > u_t) = \frac{f(u_t)}{1-F(u_t)}\big[\mathbf{E}(u|u > u_t) - u_t\big]$.

opportunities for human capital development they offer to workers. In particular, consider the following human capital law of motion:

$$K_{t+1} = K_t + A_t K_t, \tag{B17}$$

where A captures the productivity of on-the-job human capital development and varies across firms following the distribution $F(A)$. Thus, while firms pay similar wages for a given amount of human capital, they differ in the productivity of human capital development they offer.[17] Under this setup, the value of employment in period t at a firm with human capital productivity A_t is given by

$$
\begin{aligned}
V_t(K_t, A_t) = RK_t + \beta \Big[(1-\delta)\lambda \mathbf{E}\big[\max\{ V_{t+1}(K_{t+1}, A_t), V_{t+1}(K_{t+1}, A) \} \big] \\
+ (1-\delta)(1-\lambda)V_{t+1}(K_{t+1}, A_t) + \delta D_{t+1}(K_{t+1}) \Big].
\end{aligned}
\tag{B18}
$$

A worker will accept a new offer A if $A > A_t$, since R and λ are common across firms. Assuming that $A = 0$ when unemployed (human capital stock stays constant) the value of unemployment is

$$D_t(K_t) = \beta \Big[\lambda \mathbf{E}\big[V_{t+1}(K_t, A) \big] + (1-\lambda)D_{t+1}(K_t) \Big]. \tag{B19}$$

After unemployment, subsequent jobs' attribute A will be unrelated to A at previous jobs since workers sample from the unconditional distribution $F(A)$. This result is similar to that above. However, this human capital model has an important distinction to the pure search model. After an unemployment spell, subsequent wages $Y_t = RK_t$ will still be directly related to the human capital productivity of previous employers. This is because a worker's human capital stock K_t does not disappear during unemployment, and it is a function of initial human capital and the human-capital productivity of all previous employers,

$$K_t = g\big(K_0, \{A_\tau\}_{\tau=0}^{t-1} \big). \tag{B20}$$

Finally, note that the human capital accumulation function (B17) implies that K_t increases proportionally, an example where initial investments (and thus initial draws of A) can be particularly relevant for long-term human capital accumulation. An example of an alternative law of motion explicitly capturing the idea that formative years could be more fruitful for human capital development is

$$K_{t+1} = K_t + A_t f(a_t) K_t, \tag{B21}$$

where a_t is the age of the worker and $f'(\cdot) < 0$.

B.10 Differential returns to experience at large employers: Additional checks

I address two potential concerns that could bias the estimates of differential return to experience from Section 6, or threaten their interpretation as return to skills. First, the possibility of large-firm experience working as a signal of (preexisting) high unobserved productivity. Second, possible bias arising from the additive separability assumption of worker and firm-size effects.

---

[17]In Appendix D I lay out a version of an imperfect competition wage-setting framework (Card et al., 2018) which delivers the result that larger firms are larger precisely because they offer better human-capital development opportunities.

Signaling. I have interpreted the differential wage return to large-employer experience as evidence of differential human capital acquisition at large employers. Consider an alternative interpretation. Working at a large or small employer makes no difference in terms of human capital development. However, big-firm experience serves as a signal of high unobserved ability for subsequent employers.[18] Then, workers with big-firm experience are paid more not because of what they have learnt at these jobs, but because employers believe these workers are of high productivity.

I test for this possibility following the logic of Altonji and Pierret (2001). The idea is that under the pure signal interpretation, the importance of large-employer experience should diminish over time as the worker's true ability is revealed to the employer.[19]

I estimate specifications of equation (5) that allow for the differential value of large-employer experience to vary by current employer tenure. In particular I augment equation (5) by estimating

$$\ln w_{it} = \alpha_i + \psi_{s(i,t)} + \gamma_1 bigExp_{it} + \gamma_2(bigExp_{it} \cdot Exp_{it}) + \gamma_3(bigExp_{it} \cdot Tenure_{it}) + X'_{it}\delta + \varepsilon_{it}$$
(B22)

This specification allows a differential return to experience in large employers that can vary by experience and tenure. That is, letting Z_{it} be equation (B22) regressors,

$$\frac{\partial \mathbf{E}(\ln w_{it}|Z_{it})}{\partial bigExp_{it}} - \frac{\partial \mathbf{E}(\ln w_{it}|Z_{it})}{\partial smallExp_{it}} = \gamma_1 + \gamma_2 Exp_{it} + \gamma_3 Tenure_{it}$$
(B23)

A large and negative $\hat{\gamma}_3$ would be consistent with the idea of large-employer experience serving as a signal for unobserved ability. Columns (3) and (4) of Table A6 show estimates of equation (B22) without and with $\psi_{s(i,t)}$, respectively. Focusing on column (4), the table shows that $\hat{\gamma}_1$ is essentially unchanged with respect to that of column (2). $\hat{\gamma}_3$ is negative, consistent with signaling playing some role. Understanding the magnitude of the implied decay by $\hat{\gamma}_3$ will be informative of the extent to which pure signaling drives the differential return to big-firm experience.

Appendix Figure B5 shows the rate of decay as tenure increases, holding constant experience at five years. The data is consistent with large-employer experience having some signaling value, but far from explaining all of the differential return. Given the estimates of $\{\gamma_1, \gamma_2, \gamma_3\}$, a worker should stay at the same employer for over 20 years before the large-employer experience differential vanishes, which is a level of tenure not present in this sample of relatively young, mobile workers.[20]

Additive separability. Another concern that could introduce bias in the differential experience return estimates is model misspecification arising in the form of employer-size premia that vary across worker types. This would mean that the assumption of common proportional employer-size premia for all workers (additive separability of $\psi_{s(i,t)}$ and α_i) is violated. If this is the case, there could be selection based on heterogeneous employer-size premia and those with higher large-employer match quality could have more large-employer experience. In that case, I could misattribute the returns to a match-specific component to the experience coefficient.

---

[18]Under an assumption of private information. Previous work such as Farber and Gibbons (1996) and Altonji and Pierret (2001) assume that information about workers' unobserved ability is shared across employers.

[19]One caveat of my approach is that the original test of Altonji and Pierret (2001) requires that the wage return to unobserved ability also varies over time in order to load the effect of learning about employer ability. In my case, I rely on the worker fixed effect as capturing unobserved ability. Since this effect is fixed over time, I might be underestimating the rate of decay of the return of large-employer experience.

[20]To arrive at the minimum number of 20 years take into account that tenure has to be less than or equal to experience. Then, $\frac{-\hat{\gamma}_1}{\hat{\gamma}_2 + \hat{\gamma}_3} = 7341.8$ days or 20.1 years.

Card et al. (2018) discuss how the violation of additive separability in firm and worker effects is a common concern in the AKM literature and provide specification tests that support this assumption in their context. I follow Card et al. (2018) and check the plausibility of the additive specification in equation (5) by checking the distribution of mean residuals for different employer-size categories and worker types. The logic is that if the additive model is correct, residuals should have mean close to zero for all employer size/worker type combinations. On the other hand if the employer size premiums vary systematically across worker types we should see systematic departures from zero.

Appendix Figure B6 plots the mean residual for each cell based on the six employer size categories and ten deciles of estimated worker effects. Mean residuals are relatively close to zero. The largest mean residuals are those corresponding to the lowest paid (1st decile) workers, a finding consistent with Card et al. (2018) which could be explained by minimum wage policies.[21]

Figure B5: Differential wage return to one year of large employer experience, by current employer tenure
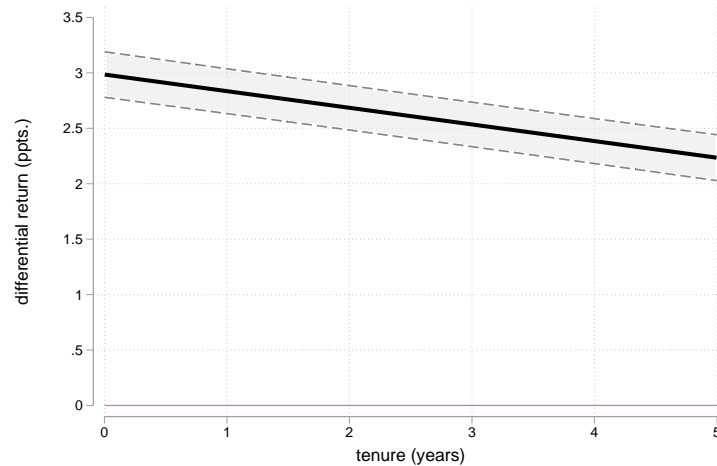
Notes: This figure plots the monthly wage differential return to one year of experience at a large employer (250+ employees) with respect to a year of experience elsewhere (<250 employees) for different current employer tenure levels, holding overall experience fixed. Uses estimates of equation (B22) (in Table A6, column (4)) and plots $365 \cdot 100(\hat{\gamma}_1 + \hat{\gamma}_2\overline{Exp} + \hat{\gamma}_3 Tenure)$ and a 95% level confidence interval computed using the delta method. $\overline{Exp}$ set at 1825 days (5 years). $Tenure$ is measured in days, x-axis re-scaled for readability. Standard errors are clustered at the worker level.

---

[21]Mean residuals also depart from zero more substantially for the "missing" employer size category. This is understandable since this is a built-in form of model misspecification arising from data limitations.

Figure B6: Mean residuals by worker effect decile/employer size
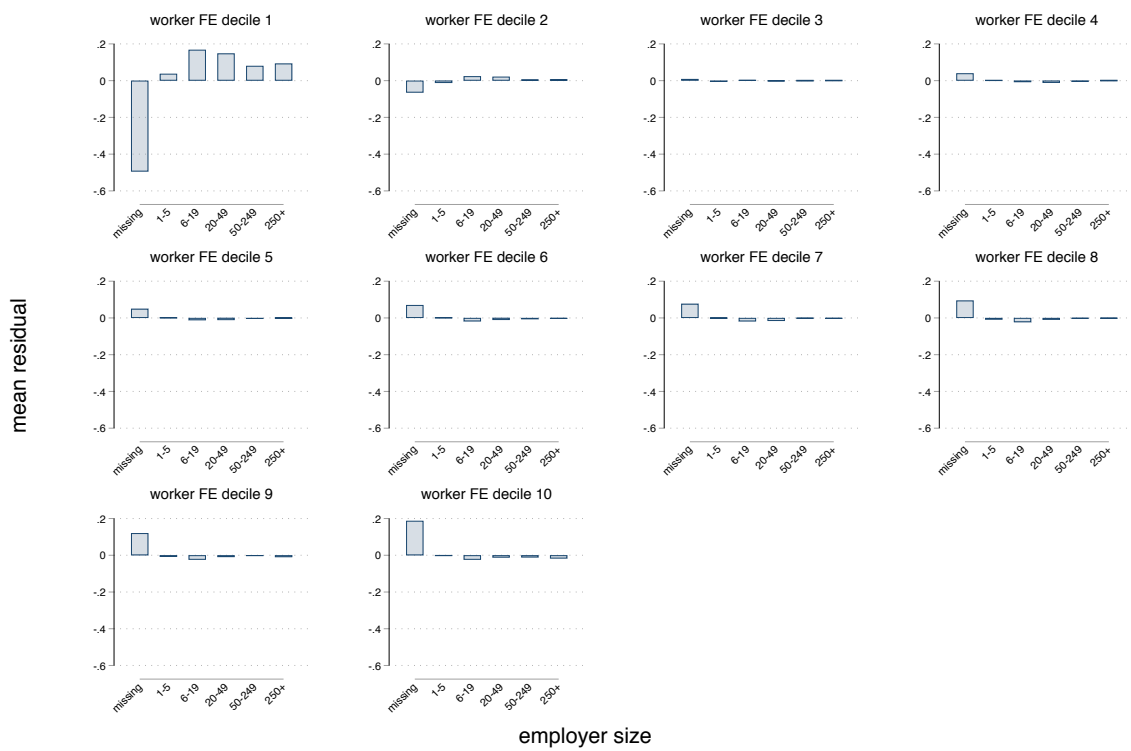
Notes: Figure shows mean residuals from estimated equation (5) with cells defined by decile of estimated worker effects (α_i) interacted with employer size category.

B.11   Promotions and differential returns to large-employer experience

Having found a differential wage premium for large-employer experience in Section 6, I study its relationship to career progression through promotions. The literature has emphasized the connection between promotions and workers' ability or human capital (see Gibbons and Waldman, 1999). A differential return to experience in terms of an increased arrival rate of promotions would further support the hypothesis that skills learned at large employers are more valuable over the working life.[22]

Social security data include information on professional categories, which I use to construct a proxy for promotions. Below, I describe the construction of this variable and provide evidence supporting its interpretation as promotions. Using this variable I estimate linear probability promotion (hazard) regressions of the following type:

$$Prom_{it} = \alpha_i + \psi_{s(i,t)} + \phi_{p(i,t-1)} + \lambda_1 bigExp_{it} + \lambda_2(bigExp_{it} \cdot Exp_{it}) + X'_{it}\delta + \varepsilon_{it}. \quad \text{(B24)}$$

Where $Prom_{it}$ is a dummy variable that equals one if worker i experienced a promotion on month t, α_i are worker fixed effects, $\psi_{s(i,t)}$ are current-employer size category fixed effects, $\phi_{p(i,t-1)}$ are indicators for the professional category worker i was holding on month $t-1$, $bigExp_{it}$ is the amount of actual experience (in days) that worker i has accumulated up until month t at employers with 250 or more employees, and Exp_{it} is the amount of total actual experience (in days, including large and small employers). X_{it} includes time-varying controls: a quadratic term for duration in current professional category, total experience (quadratic), tenure at current employer (quadratic), age (quadratic), regional unemployment level (quadratic), type of labor contract (permanent or fixed-term), sector fixed effects, and time (month) fixed effects.

In an analogous way to γ_1 and γ_2 in equation (5) in the text, λ_1 and λ_2 capture the differential impact of large-employer experience in the promotion probability, and how it varies over the working life. Let $Exp_{it} = bigExp_{it} + smallExp_{it}$ and Z_{it} be equation (B24) regressors, then

$$\frac{\partial Pr(Prom_{it} = 1 | Z_{it})}{\partial bigExp_{it}} - \frac{\partial Pr(Prom_{it} = 1 | Z_{it})}{\partial smallExp_{it}} = \lambda_1 + \lambda_2 Exp_{it}. \quad \text{(B25)}$$

Columns (1) and (2) of Appendix Table B9 show estimates from equation (B24). Column (1) does not include current employer size category fixed effects $\psi_{s(i,t)}$, while column (2) does. In both cases $\hat{\lambda}_1$ and $\hat{\lambda}_2$ indicate that large-employer experience has higher returns in terms of promotion probability that slowly decrease over time. Figure B7 helps understand the relevant magnitude implied by the coefficients and its evolution over time. On the left y-axis, it plots the differential change in the probability of promotion from one year of large-employer experience vs. one year of experience elsewhere together with a 95% confidence interval.[23] To interpret the magnitude of this differential, the right y-axis plots the relevant baseline: the monthly probability of promotion conditional on experience. It ranges from .023 when workers have one year of (actual) experience to .003 when they have twelve. The figure implies that the differential return to one year of large-employer experience amounts to 2.6% of the baseline probability when workers have one year of experience, 8.3% when they have six, and 11.6% when they have twelve.

The promotion results suggest that time spent at a large employer is more valuable than that spent elsewhere in terms of future career progression. I interpret this as further

---

[22]This would be consistent with model predictions in Gibbons and Waldman (2006), where sufficient time spent in a low-level job decreases to zero the probability of promotion.

[23]In particular, it plots $365 \cdot (\hat{\lambda}_1 + \hat{\lambda}_2 Exp)$.

Table B9: Differential returns to experience at large employers: Promotion arrival rate

	(1)	(2)
$bigExp$	1.4458***	1.6964***
	(0.2675)	(0.2847)
$bigExp \cdot Exp$	-0.0001**	-0.0002***
	(0.0001)	(0.0001)
Exp	-6.6838***	-6.7633***
	(0.2043)	(0.2046)
Exp^2	0.0013***	0.0013***
	(0.0000)	(0.0000)
$Prof.cat. - duration$	16.7451***	16.7498***
	(0.1343)	(0.1343)
$Prof.cat. - duration^2$	-0.0028***	-0.0028***
	(0.0000)	(0.0000)
Current employer size category FE	no	yes
Clusters (workers)	124872	124872
N (worker × month)	15953745	15953745

Notes: Dependent variable is a dummy that equals one if a worker experiences a promotion in that month. Experience and professional category duration measured in days. $bigExp$ is experienced acquired in employers with 250+ employees. Exp is overall experience (including $bigExp$). $Prof.cat. - duration$ equals the amount of days worked in the current professional category. Point estimates and standard errors displayed multiplied times 10^6 for readability. All specifications include worker fixed effects, current professional category fixed effects, age (quadratic), unemployment rate (quadratic), 21 sector fixed effects, fixed-term contract fixed effects, and month fixed effects. Current employer size category fixed effects groups employers into a) missing size, b) 1–5 employees, c) 6–19, d) 20–49, e) 50–249, and f) 250+. Standard errors clustered at the worker level in parentheses. * 0.10 ** 0.05 *** 0.01.

Figure B7: Differential change in probability of promotion to one year of large employer experience, by total experience
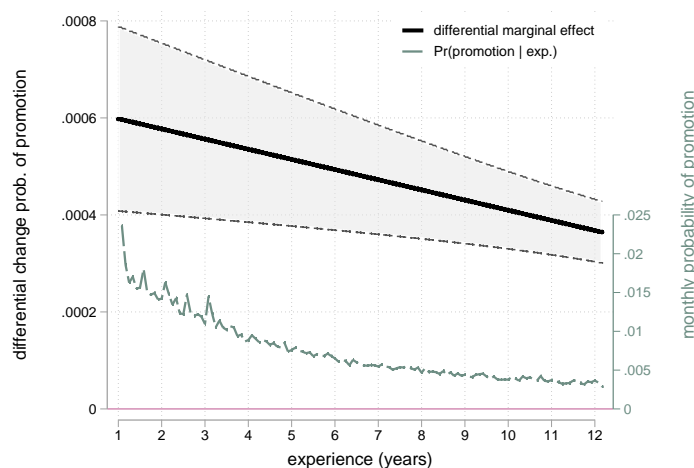
Notes: Differential increase in the monthly probability of promotion of one year of experience at a large employer (250+ employees) with respect to a year of experience elsewhere (<250 employees) (left y-axis), and the monthly probability of promotion (right y-axis) for different levels of experience. Left y-axis uses estimates of equation (B24) (Appendix Table B9, column (2)) and plots $365 \cdot (\hat{\lambda}_1 + \hat{\lambda}_2 Exp)$ and a 95% level confidence interval computed using the delta method. Exp is measured in days, x-axis re-scaled for readability. Standard errors are clustered at the worker level.

supportive evidence for the hypothesis that workers learn differentially valuable skills at large employers that pay off in terms of higher wages and faster career progression.

Construction of promotion variable

The data include a professional category variable ("*grupo de cotización*") that allows the creation of a promotion proxy. This variable is determined by the type of job a worker performs and not by her education level. There are originally 13 categories which I group into 10. I group together the three lower-ranked groups to which workers less than 18 years old belong. I further combine into a single group the original groups 6 and 7, based on wage data.

I interpret upward movements in professional categories as promotions and study its arrival rate in relationship to large-employer experience. My definition implies that a worker experiences a promotion in a given month if it is the first month he is employed in his highest-ranked category up to date (e.g. I assign a worker with the trajectory 6-5-4-4 as having promotions in months 2 and 3; I define a worker with the trajectory 6-4-5-4 as having a promotion only in month 2). I also do not count as promotions moves out from the lowest category (10), as these moves are mechanically related to workers' age.

C IV-TSLS Interpretation, Flexible First Stage Estimation, and Compliers' Characteristics

The goal of this appendix is to provide further insight into the instrumental-variable (IV) two-stage least squares (TSLS) estimation of the elasticity of lifetime income with respect to first-employer size (Section 4). In particular, I have argued (see Section 4.5) that heterogeneous treatment effects and compliers' characteristics likely explain the difference between the OLS and IV estimates. While this *local average treatment effect* (LATE) logic is well-known and well-understood for the case of binary treatments and binary instruments, it is less straightforward in settings such as mine where the treatment (first-employer size) and the IV (index of labor demand composition) take multiple values.

Here, I follow Angrist and Imbens (1995) to clarify what causal effect is TSLS estimating (which differences in potential outcomes, and for which subpopulations). I then build on these analytic results and, using a distribution regression framework (Chernozhukov et al., 2013), estimate weights from different parts of the first-employer size distribution that feed into the TSLS estimate. Finally, by carrying out this exercise across worker subgroups, I get a better understanding of what type of workers are driving the TSLS estimates. The findings of this exercise are that workers who are less educated and originally from less urban provinces are disproportionately likely to be "compliers", meaning that the size of their first employer is more sensitive to the demand variation my IV captures.

C.1 Analytical Framework

The goal is to explore the following questions in the presence of treatment effect heterogeneity, multivalued treatment, and multivalued instruments: (i) what causal effect is TSLS estimating? (which differences in potential outcomes, and for which subpopulations); (ii) from which treatment values (initial firm size) is it mostly coming from?; (iii) what are the characteristics of the relevant compliers for which the causal effect is estimated?

Setup

Potential outcomes (lifetime earnings) for worker i whose first employer (log) size is $s = 0, 1, 2, \ldots, J$ are denoted by Y_{si} .[24] The instrument (labor demand environment) is represented by Z_i. It could be binary $Z_i \in \{0, 1\}$, or multivalued $Z_i \in \{0, 1, 2, \ldots, K\}$. My empirical exercise uses the latter, but the former case is simpler to build intuition. Different values of the instrument induce different potential treatment values. S_{Zi} denotes first employer (log) size for each different instrument value. With a binary instrument, each worker i has two potential treatment values S_{1i} and S_{0i}.

Binary Instrument Case

Assumptions:

1. Independence: S_{1i}, S_{0i}, Y_{0i}, Y_{1i}, $\cdots$, Y_{Ji} are independent of Z_i.

2. Monotonicity: $S_{1i} \geq S_{0i}$ for all i.

What causal effect is TSLS estimating?

Angrist and Imbens (1995) show (in their Theorem 1) that TSLS identifies a weighted average of causal responses to a unit change in treatment, $Y_{si} - Y_{(s-1)i}$, for those whose

---

[24]Positive integers are not attractive for log size example, but units are immaterial in this discussion.

treatment status is affected by the instrument. Compliers in this case are characterized by (i) the base level at which they comply S_{0i}, and (ii) the intensity of compliance $S_{1i} - S_{0i}$.

More specifically their Theorem 1 shows that

$$\beta^{TSLS} = \sum_{s=1}^{J} \omega_s \cdot E[Y_{si} - Y_{(s-1)i} | S_{1i} \geq s > S_{0i}], \tag{C1}$$

where

$$\omega_s = \frac{Pr(S_{1i} \geq s > S_{0i})}{\sum_{m=1}^{J} Pr(S_{1i} \geq m > S_{0i})}.$$

Note that ω_s, the weight attached to the average of $Y_{si} - Y_{(s-1)i}$, is proportional to the number of people that the instrument induces to change first employer size from less than s to s or more. This weights are analogue to the proportion of compliers in the simple binary treatment case, and they are the stepping stone to answering the remaining two questions.

From which treatment values is β^{TSLS} mostly coming from?

The unit-response weights above can be estimated with observables S_i, Z_i since

$$Pr(S_{1i} \geq s > S_{0i}) = Pr(S_{1i} \geq s) - Pr(S_{0i} \geq s) = Pr(S_i \geq s | Z_i = 1) - Pr(S_i \geq s | Z_i = 0).$$

Plotting the weighting function

$$r(s) \equiv Pr(S_i \geq s | Z_i = 1) - Pr(S_i \geq s | Z_i = 0) \tag{C2}$$

would show which s values have higher weight in β^{TSLS}. Angrist and Imbens (1995) plot these types of weighting functions for their example of years of schooling (S_i) and quarter of birth (Z_i, first or last quarter).

What are characteristics of the relevant compliers?

It is useful to first see how this question would be answered in the simpler framework of a binary treatment. If one if interested in knowing wether for some covariate dummy X_i compliers are more or less likely to have $X_i = 1$:

$$\frac{Pr(X_i = 1 | C_i = 1)}{Pr(X_i = 1)} = \frac{Pr(C_i = 1 | X_i = 1)}{Pr(C_i = 1)} = \frac{E(S_i | Z_i = 1, X_i = 1) - E(S_i | Z_i = 0, X_i = 1)}{E(S_i | Z_i = 1) - E(S_i | Z_i = 0)}$$

where $C_i = 1$ if i complier (i.e. $S_{1i} - S_{0i} = 1$). Note that the above expression is based on objects that are observable in the data.

Back to the multivalued treatment, for some covariate dummy X_i one can see whether complier units (from a given treatment range) are more or less likely to have $X_i = 1$ than other units with:

$$r_X(s) = \frac{Pr(S_i \geq s | Z_i = 1, X_i = 1) - Pr(S_i \geq s | Z_i = 0, X_i = 1)}{Pr(S_i \geq s | Z_i = 1) - Pr(S_i \geq s | Z_i = 0)}. \tag{C3}$$

MULTIVALUED INSTRUMENT CASE

When both the treatment and the instrument are multivalued - as in my empirical implementation - the interpretation becomes more involved but the intuitions from above carry

forward.

Instrument Z_i can now take any of $k = 0, 1, \ldots K$ values. The **monotonicity assumption** now involves that $S_{ki} \geq S_{(k-1)i}$ for all k and i. Define the following for each pair of instrument values k and l:

$$\beta_{k,l} = \frac{E(Y_i|Z_i = k) - E(Y_i|Z_i = l)}{E(S_i|Z_i = k) - E(S_i|Z_i = l)}$$

Angrist and Imbens (1995) show that similarly as for their Theorem 1

$$\beta_{k,l} = \sum_{s=1}^{J} \omega_s^{kl} \cdot E[Y_{si} - Y_{(s-1)i}|S_{ki} \geq s > S_{li}], \tag{C4}$$

where

$$\omega_s^{kl} = \frac{Pr(S_{ki} \geq s > S_{li})}{\sum_{m=1}^{J} Pr(S_{ki} \geq m > S_{li})}. \tag{C5}$$

Their Theorem 2 concludes that in the multivalued instrument case

$$\beta^{TSLS} = \sum_{k=1}^{K} \mu_k \beta_{k,k-1}$$

$$= \sum_{k=1}^{K} \sum_{s=1}^{J} \mu_k \omega_s^{k,k-1} \cdot E[Y_{si} - Y_{(s-1)i}|S_{ki} \geq s > S_{k-1,i}] \tag{C6}$$

where

$$\mu_k \propto [E(S_i|Z_i = k) - E(S_i|Z_i = k - 1)] \cdot \psi_k,$$

and

$$\psi_k = [E(S_i|Z_i \geq k) - E(S_i|Z_i < k)]Pr(Z_i \geq k)[1 - Pr(Z_i \geq k)].$$

Note that the weights μ_k are arguably less interesting than the weights $\omega_s^{k,k-1}$; the first term that they are proportional to is constant under a first stage linearity assumption, and the second term simply gives more weight to the central part of the distribution of Z_i.

C.2 Flexible First Stage Estimation and Properties

In a general way, I model the first stage with the conditional distribution function

$$F(s|Z_i, X_i) = Pr(S_i \leq s|Z_i, X_i),$$

where S_i is log first employer size of worker i, Z_i is the labor demand instrument, and X_i are the remaining covariates from the first stage (unemployment controls, education fixed effects, birth cohort fixed effects, and region of birth fixed effects). I can estimate $F(s|Z_i, X_i)$ using the distribution regression framework outlined in Chernozhukov et al. (2013).

Once I estimate $F(s|Z_i, X_i)$ the first goal is to study properties of the weights $\omega_s^{k,k-1}$ in

equation (C6):

$$\omega_s^{k,k-1} = \frac{Pr(S_{ki} \geq s > S_{k-1,i})}{\sum_{m=1}^{J} Pr(S_{ki} \geq m > S_{k-1,i})}. \tag{C7}$$

This will help understand which are the values of first employer size and the instrument which are mostly driving the estimated coefficient.

The second goal will involve studying the heterogeneity of these weighting weights across different subpopulations (education, urban/rural).

Estimation of $F(s|Z_i, X_i)$ using distribution regression

Let $\mathcal{S}$ be the set of treatment values (log first employer size) I observe in the data. I follow Chernozhukov et al. (2013) and model $F(s|Z_i, X_i)$ separately for each threshold $s \in \mathcal{S}$. In particular

$$F(s|Z_i, X_i) = \Lambda\Big(g\big(Z_i, X_i; \theta(s)\big)\Big) \qquad \text{for all } s \in \mathcal{S} \tag{C8}$$

where Λ is a known link function and g is a function of Z_i, X_i whose parameters $\theta(s)$ vary for each different value of s. I set the link function to be logistic, $\Lambda(v) = \frac{e^v}{1+e^v}$, and $g\big(Z_i, X_i; \theta(s)\big)$ to be the same linear function of the instrument and controls used in the TSLS estimation,

$$g\big(Z_i, X_i; \theta(s)\big) = \gamma_0(s) + \gamma_1(s)Z_i + X_i'\delta(s),$$

where the controls X_i are the same as in the main IV specification from equation (2): a quartic of regional unemployment rate at predicted graduation interacted with educational attainment fixed effects, birth region fixed effects, and birth cohort fixed effects.

Estimating $\theta'(s) = \big[\gamma_0(s), \gamma_1(s), \delta(s)\big]$ for each $s \in \mathcal{S}$ involves running the following $|\mathcal{S}|$ logit regressions:

$$Pr(S_i \leq s|Z_i, X_i) = \Lambda\Big(\gamma_0(s) + \gamma_1(s)Z_i + X_i'\delta(s)\Big).$$

Continuous weighting function for different first employer size values

Using the parameter estimates $\hat{\theta}(s)$ from the procedure above I can compute objects that resemble the weights of of equation (C7). The key idea is to use the estimated distribution first stage and note that for an instrument Z_i that is close to continuous such as mine

$$Pr(S_{ki} \geq s > S_{k-1,i}) = Pr(S_i \geq s|Z_i = k) - Pr(S_i \geq s|Z_i = k-1) \approx \frac{\partial Pr(S_i > s|Z_i = k)}{\partial Z}, \tag{C9}$$

and that the distributional regression model readily provides an expression for the derivative of interest:

$$\frac{\partial Pr(S_i > s|Z_i, X_i)}{\partial Z} = -\gamma_1(s) \cdot \Lambda\Big(\gamma_0(s) + \gamma_1(s)Z_i + X_i'\delta(s)\Big) \cdot \Big[1 - \Lambda\Big(\gamma_0(s) + \gamma_1(s)Z_i + X_i'\delta(s)\Big)\Big]. \tag{C10}$$

Taking together equations (C7) and (C9), we can think of an estimable two-dimensional weighting function which averages across the distribution of covariates X_i:

$$r(s,k) = \int_{X_i} \phi(k) \cdot \frac{\partial Pr(S_i > s | Z_i = k, X_i = x)}{\partial Z} dF_{X_i}(x),$$

where:

$$\phi(k) = \left(\sum_{m=1}^{J} \frac{\partial Pr(S_i > m | Z_i = k, X_i = x)}{\partial Z} \right)^{-1}.$$

(C11)

Appendix Figure C1 plots an estimated function $\hat{r}(s,k)$ as a function of first employer size s, for different values k of the instrument:

$$\hat{r}(s,k) = \hat{\phi}(k) \cdot \frac{1}{N} \sum_{i=1}^{N} \left(-\hat{\gamma}_1(s) \cdot \Lambda\left(\hat{\gamma}_0(s) + \hat{\gamma}_1(s)k + X_i'\hat{\delta}(s)\right) \cdot \left[1 - \Lambda\left(\hat{\gamma}_0(s) + \hat{\gamma}_1(s)k + X_i'\hat{\delta}(s)\right)\right] \right)$$

where:

$$\hat{\phi}(k) = \left[\frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{J} \left(-\hat{\gamma}_1(s) \cdot \Lambda\left(\hat{\gamma}_0(s) + \hat{\gamma}_1(s)k + X_i'\hat{\delta}(s)\right) \cdot \left[1 - \Lambda\left(\hat{\gamma}_0(s) + \hat{\gamma}_1(s)k + X_i'\hat{\delta}(s)\right)\right] \right) \right]^{-1}$$

(C12)

From equation (C7) we can interpret these weights as putting higher values on the levels of s that, induced by the instrument, more people "jump over". Under this interpretation, Appendix Figure C1 suggests that marginal changes in the instrument induce people to avoid having a relatively small first employer (high estimated weights between the 25th and 40th percentiles).

C.3 Compliers' characteristics: Heterogeneity across workers

We can learn something about what are the characteristics of people more responsive to the instrument—characteristics of "compliers"—using the machinery developed above. In particular, we can use the logic from equation (C3): compute the weighting function (C11) for different covariate values and compare. This will tell us which subgroups is the TSLS estimate giving more weight to. Figure 5 in the main text plots this analysis across two different dummy covariates, together with the overall weighting function from Appendix Figure C1, holding the IV value constant in the 95th percentile.[25] The two characteristics I study are a dummy variable indicating urban or rural place of birth (top panel) and a dummy variable indicating college education or not (bottom panel).

The figure suggest that i) the instrument has a greater impact in first employer size for those born in rural places and for those without a college education, ii) this is specially so when shifting workers away from the bottom of the first employer size distribution, iii) the difference in both cases starts to diminish between the 70th and 80th percentiles of the first employer size distribution, and iv) the comparison reverses for very largest first employer sizes: The instrument seems to impact more the movements across this part of the distribution for urban-born and college workers.

Overall, the analysis carried out in this section supports the intuition laid out in the paper with respect to the comparison between OLS and TSLS elasticities: The "compliers" who are mostly driving the TSLS estimates are less educated and come from less urban

---

[25]Results are similar for other IV values. A high IV value represents a substantial large-firm hiring shock like the DuPont example I discuss in Section 4.

places of Spain. The "large" TSLS estimates (in comparison to OLS) seem to imply that these are the workers who are more sensitive to their first employer characteristics.

Figure C1: Estimated weight function from flexible first stage
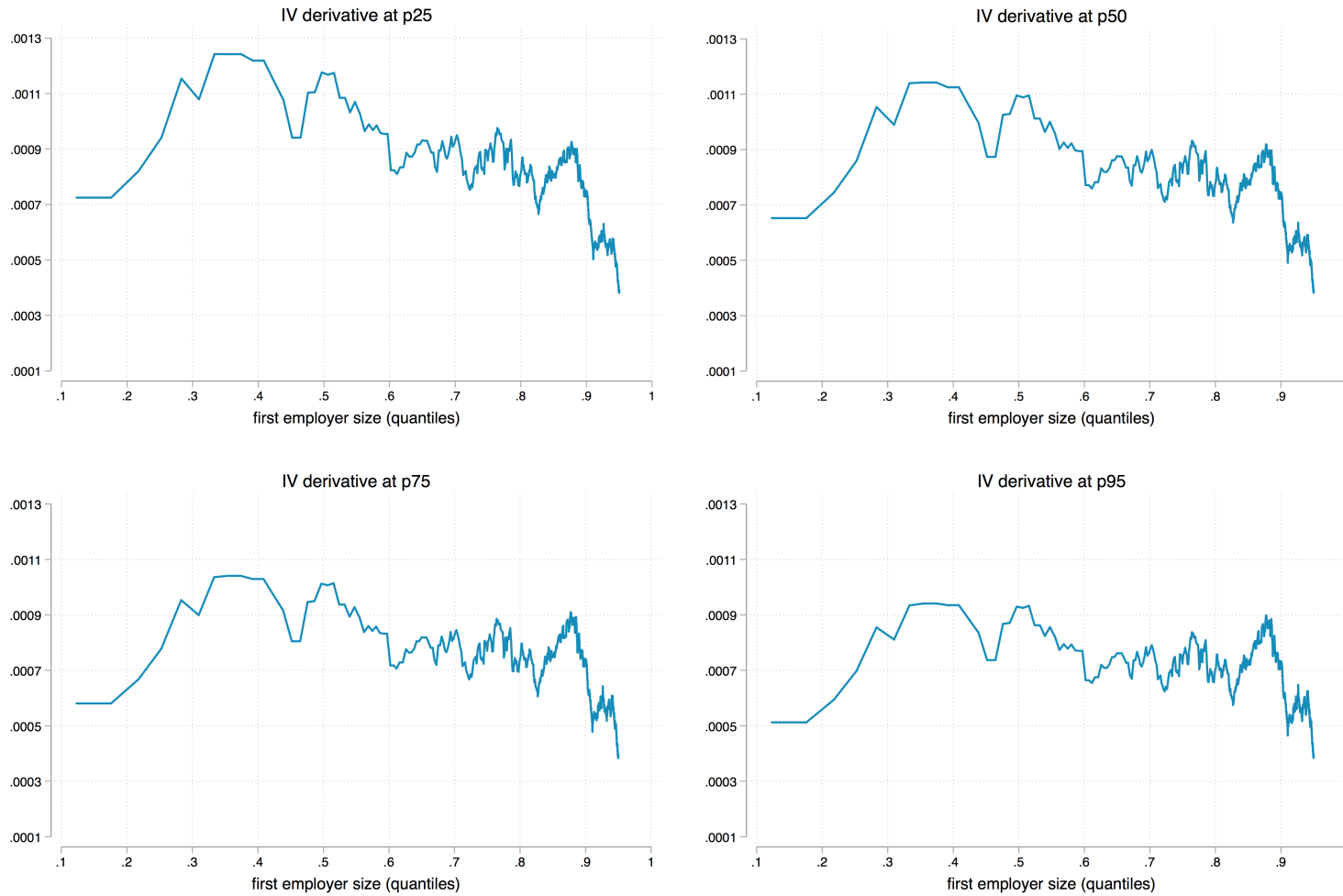
Notes: Estimated weight function from equation (C12) as a function of first employer size s. Plotted in different panels for different instrument values k.

D On-the-job Skills and Employer Size in an Imperfectly Competitive Labor Market

In this section I first discuss how a simple static model of an imperfectly competitive labor market (Card et al., 2018) can give rise to an equilibrium result in which firms with better training opportunities employ more workers in equilibrium, provided that workers value such training opportunities. I then provide a simple two-period extension of the previous model which rationalizes workers valuing such training. In these models, larger firms offer better training opportunities than smaller ones. Firms, however, are larger (in part) *because* they offer better training opportunities, and not the other way around.

D.1 Card et al. (2018) through the lens of training opportunities

Card et al. (2018) develop a static wage posting model of an imperfectly competitive labor market, with heterogeneous worker valuation from jobs at different employers. This heterogeneity in workers' employer valuations gives rise to firms setting wages to maximize profits in a monopsony-type of way.

A simplified version of their model features J firms and a mass one of workers of a single skill level. Each firm $j \in \{1, 2, \ldots, J\}$ posts a single wage w_j. All workers observe all wages, and firms hire any worker that chooses to work for them at the posted wage. Firms are heterogeneous, and workers have different preferences for working at different employers. Let the utility of worker i from working in firm j be given by

$$u_{ij} = \beta \ln w_j + \tilde{a}_j + \varepsilon_{ij}, \tag{D1}$$

where $\tilde{a}_j$ is defined as a firm-specific amenity with common value across workers, and ε_{ij} are idiosyncratic preference shocks (*e.g.* distance from home, or scheduling flexibility) which are independent draws from a type I extreme value distribution.[26]

For the sake of this example, think of $\tilde{a}_j$ as representing the quality of on-the-job skill development at firm j. This interpretation of $\tilde{a}_j$ is the key insight delivering the relationship between training opportunities and firm size.[27] Workers choose the firm j which provides the highest utility. Given the distributional assumption, the probability of choosing firm j is given by:

$$p_j = \frac{exp(\beta \ln w_j + \tilde{a}_j)}{\sum_{k=1}^{J} exp(\beta \ln w_k + \tilde{a}_k)}$$

which, if we assume that J is large enough so that there are no strategic interactions between firms, can be approximated by

$$p_j \approx \lambda \cdot exp(\beta \ln w_j + \tilde{a}_j),$$

where λ is a constant. This results in the firm-specific labor supply function

$$N_j(w_j) = a_j w_j^{\beta}, \qquad \text{where} \quad a_j \equiv \lambda e^{\tilde{a}_j} \tag{D2}$$

Firms have a linear production function where labor N_j is its only input, and have het-

---

[26] A more detailed specification would be given by $u_{ij} = \beta \ln(w_j - b) + \tilde{a}_j + \varepsilon_{ij}$ where b is the outside option. For simplicity in what follows I set the outside option to $b = 0$.

[27] It is not clear why workers in this static model would value skill acquisition per se if not rewarded for their skills. A simple two period extension in the following section deals with this issue.

erogeneous productivities A_j:[28]

$$Y_j = A_j N_j.$$

For simplicity, assume that firms face a constant unit price p in the product market. In that case, firm j sets wages by solving the profit maximization problem

$$\max_{w_j} pA_j N_j(w_j) - w_j N_j(w_j).$$

Using the labor supply function (D2) and taking the first order condition leads to equilibrium wage at firm j

$$w_j^* = pA_j \frac{\beta}{1+\beta},$$

and equilibrium employment level

$$N_j^* = a_j \left(pA_j \frac{\beta}{1+\beta} \right)^\beta.$$

We can see that the shift in labor supply driven by a_j in equation (D2) results in firms with higher quality training opportunities (or any other common value amenity represented in a_j) having higher equilibrium levels of employment.[29]

D.2 Heterogeneous training opportunities in a two period extension

I provide a simple extension featuring two periods. Firms set wages and workers choose firms in a frictionless environment in each of the two periods $t \in \{t_0, t_1\}$. Workers in t_0 all have the same level of skill but during employment in t_0 workers learn skills that will differentiate them in t_1. In particular, some workers achieve a high level of skill ($S_i = H$) while others only achieve a low level of skill ($S_i = L$).

Firms in t_0 are heterogeneous in the rate at which their workers achieve the high level of skill. Workers matched with firm j in t_0 reach $S_i = H$ with probability q_j, and $S_i = L$ with probability $(1 - q_j)$. Technology in the second period is such that the marginal product of high- and low-skill workers is different. Skills are fully portable across employers. When choosing employers in t_0, workers take into account the different probabilities across firms of reaching the high skill level, and any wage differential across skills in t_1.

Let J_t be the number of firms in period t. For simplicity we can think of $J_0 = J_1 = J$ but the identity of the firms is inconsequential since workers choose employers in period t_1 in a frictionless way. Firms' technology will, however, differ across periods. Since workers are homogeneous in the first period, each firm j posts a single wage w_j. In the second period, each firm j posts a pair of wages $\{w_{Lj}, w_{Hj}\}$, one for each type of worker. Let k_0 and k_1 index firms in the first and second periods respectively. When choosing employer j

---

[28]The argument relating training opportunities and equilibrium firm size still holds under a more general production function $Y_j = A_j N_j^\alpha$ where $\alpha \in (0, 1]$. Assuming linearity simplifies notation while delivering the same point.

[29]In the case where the production function features decreasing returns to scale with common parameter $\alpha \in (0, 1]$, the equilibrium wage and employment levels of firm j are

$$w_j^* = \left(pA_j \frac{\beta}{1+\beta} \alpha a_j^{-(1-\alpha)} \right)^{\frac{1}{1+\beta(1-\alpha)}}$$

and

$$N_j^* = \left(pA_j \frac{\beta}{1+\beta} \alpha \right)^{\frac{\beta}{1+\beta(1-\alpha)}} a_j^{\frac{1}{1+\beta(1-\alpha)}}.$$

In this case firms with higher a_j are still larger in equilibrium, like in the linear case. In this case, however, firms with higher a_j pay lower wages in equilibrium.

in period t_0 worker i faces an intertemporal utility function:

$$U_i(k_0 = j) = u_{ij}^0 + \delta \mathbf{E}(u_{ik_1}^1 | k_0 = j)$$

where

$$u_{ij}^0 = \beta \ln w_j + \varepsilon_{ij}^0,$$

$$u_{ij}^1 = \mathbb{1}\{S_i = H\} \cdot \left(\beta \ln w_{Hj} + \varepsilon_{iHj}^1\right) + \mathbb{1}\{S_i = L\} \cdot \left(\beta \ln w_{Lj} + \varepsilon_{iLj}^1\right),$$

δ is a discount factor, $\mathbb{1}\{\}$ is the indicator function, and $\{\varepsilon_{ij}^0\}$, $\{\varepsilon_{iHj}^1\}$,$\{\varepsilon_{iLj}^1\}$ are independent draws from type 1 extreme value distributions. The first implication of this utility representation is that workers have idiosyncratic preferences for firms that are independent across periods and across states of the world in the second period.[30] The second implication is that expected wages in t_1 as a function of first period employer acts as a common value firm component ($\tilde{a}_j$ in the static model) in t_0.

In period t_0 firms' production is linear in (homogeneous) labor:

$$Y_j^0 = A_j N_j.$$

In t_1 workers have been differentiated into low ability L and high ability H. Their marginal product of labor in this period is different, governed by the parameter $\theta \in (0.5, 1)$:

$$Y_j^1 = A_j\big((1 - \theta)L_j + \theta H_j\big).$$

For simplicity, assume that in both periods firms face a constant product price p.

The model is solved by backwards induction. In t_1, once the uncertainty about their skill is realized, workers see firms' wage postings and choose their preferred job. Thanks to the idiosyncratic preferences distributional assumptions, the same reasoning as in the static version (assuming no strategic interactions between firms) leads to firm-specific supply functions for each type of worker:

$$H_j(w_{Hj}) = \kappa_H w_{Hj}^\beta \tag{D3}$$

$$L_j(w_{Lj}) = \kappa_L w_{Lj}^\beta, \tag{D4}$$

where κ_S is a constant proportional to the fraction of workers of skill $S \in \{H, L\}$.

Firms take into account their firm- and skill-specific labor supply functions and set the pair of wages $\{w_{Lj}, w_{Hj}\}$ to maximize profits:

$$\max_{w_{Hj}, w_{Lj}} pA_j\big((1 - \theta)L_j(w_{Lj}) + \theta H_j(w_{Hj})\big) - w_{Lj}L_j(w_{Lj}) - w_{Hj}H_j(w_{Hj}).$$

Taking first order conditions and using the labor supply functions in (D3), (D4), leads to equilibrium wages in t_1

$$w_{Hj}^* = pA_j \frac{\beta}{1 + \beta}\theta, \qquad\qquad w_{Lj}^* = pA_j \frac{\beta}{1 + \beta}(1 - \theta).$$

Taking the above t_1 equilibrium wages as given in period t_0, and given the frictionless setting, we can get a simple expression for expected t_1 utility in the first period, as a function

---

[30]This could reflect the fact that younger and older workers value workplace characteristics differently, as well as the fact that job amenities and characteristics within a firm could be very different for its high versus low skill workers.

of firm j's probability of skill upgrading during t_0:

$$\mathbf{E}(u^1_{ik_1}|k_0 = j) = \mathbf{E}\Big[\mathbb{1}\{S_i = H\} \cdot \beta \ln w^*_{Hk_1}\big|k_0 = j\Big] + \mathbf{E}\Big[\mathbb{1}\{S_i = L\} \cdot \beta \ln w^*_{Lk_1}\big|k_0 = j\Big]$$

$$= \beta\Big[q_j \cdot \underbrace{\mathbf{E}(\ln w^*_{Hk_1})}_{\equiv \tilde{w}_H} + (1 - q_j) \cdot \underbrace{\mathbf{E}(\ln w^*_{Lk_1})}_{\equiv \tilde{w}_L}\Big]$$

$$= \underbrace{\beta\Big[\tilde{w}_L + q_j(\tilde{w}_H - \tilde{w}_L)\Big]}_{\equiv \psi_j}$$

This result then implies that the period t_0 equilibrium is one in which workers preferences are given by

$$u^0_{ij} = \beta \ln w_j + \psi_j + \varepsilon^0_{ij}, \qquad \text{where} \quad \psi_j \equiv \beta\Big[\tilde{w}_L + q_j(\tilde{w}_H - \tilde{w}_L)\Big].$$

This is analogous to (D1) in the static model above. In this case ψ_j is a function of the wage differential in t_1 and the probability of skill upgrading in firm j, acting as a common value firm-specific component.

The results from the static model then apply in this setting in t_0: In equilibrium firms with higher ψ_j - better training opportunities - will have a larger workforce. This extension rationalizes workers valuing training opportunities, and delivers the prediction that a larger skill wage gap will result in a larger elasticity of firm size with respect to training quality.

E Distinctive Large-Employer Attributes and Skill Accumulation

This appendix provides a discussion of firm characteristics that differ across large and small employers and could underlie more valuable development of on-the-job skills at larger firms. When possible, I provide descriptive evidence relating firm size and these attributes in the context of Spain.

E.1 Formal Training and Education

Large employers engage in higher amounts of training and in a more structured way. A reason for doing this might be the spreading of fixed costs associated with worker training; another reason might be the higher likelihood of large employers to benefit from training through internal labor markets. Lynch and Black (1998) show that training programs are more prevalent at larger employers, and that these include teaching of general skills such as computing and basic education.[31]

Appendix Table E1 uses survey data to show the positive relationship between firm size and employer-provided training in Spain. Workers at employers with 250+ employees are twice as likely to be engaged in informal workplace education than workers at employers with 1–10 employees (3.49% vs. 1.68%), around six times more likely to be engaged in formal workplace education (4.33% vs. 0.75%), and three times more likely to be engaged in either formal or informal workplace education (6.66% vs. 2.30%).[32]

Table E1: Workplace training and education across employer size

workers	percent of sample	percent informal ed.	percent formal ed.	percent informal or formal ed.
1-10	36.09	1.68	0.75	2.30
11-19	12.36	1.11	1.08	1.96
20-49	16.39	1.98	1.12	2.55
50-249	18.14	3.38	1.35	4.54
250+	17.02	3.49	4.33	6.66
N	2555			

Notes: Source is the 2011 Survey on the Involvement of the Adult Population in Learning Activities (*Encuesta sobre la participación de la población adulta en las actividades de aprendizaje*, or EADA). Sample restricted to those who are 18–35 years old and employed. Formal education is that which is expected to lead to a degree completion. Informal education is defined as practical activities oriented towards job preparation. I count formal or informal education as being workplace training and education if it is either financed by the respondent's employer, or if it mainly or exclusively takes place during working hours. Total sample size is 2,555 and percentages are computed using survey weights.

E.2 Organizational Structure

Learning the ropes. Other employer features different from formal task training could impact workers' general skill development. The organizations literature emphasizes how workers' outcomes can be impacted by internal structures and processes (see the discussion in Gibbons and Waldman, 1999). Significant attention has been devoted to "people

---

[31]The literature offers several reasons why employers would invest in training for their workers that might be valuable in other firms. While maintaining the traditional dichotomy between general and specific human capital, Acemoglu and Pischke (1999) point that in the absence of perfect labor market competition, common frictions that create monopsony rents will lead to employers finding it optimal to invest in the general human capital of their workers. Lazear (2009) proposes a model of firm-training in which all skills are general but used in different proportions by different employers. Such a model also leads to firms to pay for training that is valuable elsewhere.

[32]Formal education is that which is expected to lead to a degree completion. Informal education is defined as practical activities oriented towards job preparation. I count formal or informal education as being *workplace* education if it is either financed by the respondent's employer, or if it mainly or exclusively takes place during working hours.

procesing" or "organizational socialization" (Van Maanen, 1978)—how internal processes impact the way in which new workers learn the necessary skills at their new jobs. Many "people processing" practices that could impact a young worker's initial experiences in the firm can only be carried out successfully by firms with a large number of employees.

Large firms, with large batches of new workers, may be more likely to engage in the collective socialization of new employees by providing formal staff induction (Antona-copoulou and Güttel, 2010). Such processes may teach (especially inexperienced young workers) the necessary know-how and work culture to operate successfully in large organizations.

Job rotation. The practice of job rotation is related to the processing of newcomers. This can let workers develop diverse skills as well as helping them (and their employer) realize which are the tasks they are more productive at. While some workers might need to change employers in order to do so, large firms might offer the possibility of doing this internally. Larger employers have a wider set of tasks across which to rotate workers, and are more likely to do so (Gittleman et al., 1998; Eriksson and Ortega, 2006).

Managerial and coworker quality. The hierarchical production literature provides complementary theoretical and empirical evidence on the relationship between organizational structure, employer size, and skill-development opportunities for workers (Garicano, 2000; Garicano and Rossi-Hansberg, 2015; Caliendo et al., 2015). Robust predictions of these models are that the marginal return of a worker is linked to the the characteristics of other workers in her team, and that better managers lead better and larger teams (Lucas, 1978). This suggests an opportunity to learn from better peers and better managers at larger employers (see Caicedo et al., 2019; Nix, 2017; Jarosch et al., 2018). Bloom and Van Reenen (2006) show that larger firms tend to be better managed. Using data from the World Management Survey, Appendix Figure E1 shows that the correlation between size and management quality is present for Spanish employers.

Figure E1: Firm size and managerial quality in Spain
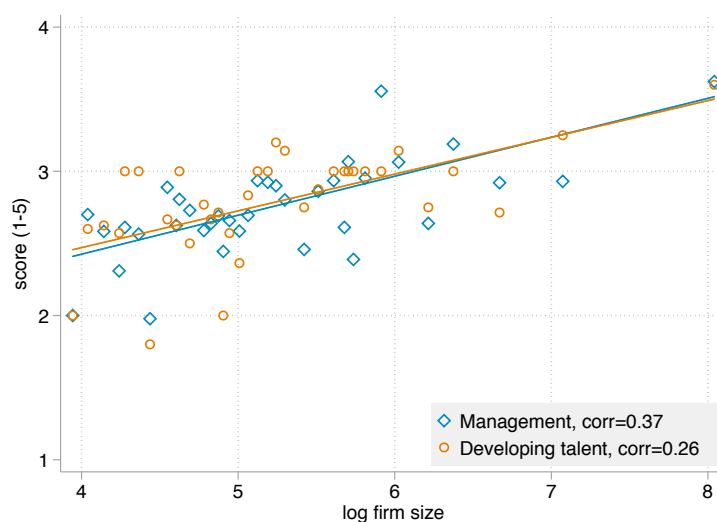
Notes: Source is World Management Survey, 2013 wave. Data on 214 manufacturing plants in Spain. Size refers to firm (not plant) size. Management is the average score of all survey management questions. Developing talent is the score of a single question (which is also included in the overall Management average).

E.3 Firm Production and Activities

Larger employers are more likely to be exporters and, similarly to size, this is a firm attribute the literature has associated with higher wages (Bernard et al., 1995; Bernard and Jensen, 1999).[33] Using data from Italy, Macis and Schivardi (2016) argue that export wage-premia are most important for workers with *previously existing* export-related experience. This is suggestive of a type of skill developed on the job, more likely to be acquired at large employers, and that could be valuable throughout workers' careers. Skills related to exporting activities could be particularly relevant in the context I study, given the undergoing modernization and internationalization of the Spanish economy at the time.

Kugler and Verhoogen (2012) document a strong correlation between manufacturing plants' size and the quality of their inputs and outputs. This complements the fact that larger employers tend to be more productive (e.g. Leung et al., 2008; Moral-Benito, 2018), and evidence suggesting that they are faster to adopt new technologies (e.g. Fabiani et al., 2005). Working with higher quality inputs, adhering to higher quality standards, being involved in more efficient processes, or using more sophisticated technology are channels through which workers might develop higher-value skills at large employers. Appendix Table E2 shows that during the 1990s and early 2000s, larger employers in Spain were more likely to invest on R&D and foreign technology transfers.

Table E2: R&D investment, foreign technology transfer payments, and firm size

	1(R&D investment > 0)		1(Foreign tech. payments>0)	
	(1)	(2)	(3)	(4)
log firm size	0.0655***	0.0537***	0.0312***	0.0331***
	(0.0044)	(0.0045)	(0.0034)	(0.0036)
Sector FE	no	yes	no	yes
Year FE	yes	yes	yes	yes
LHS var. average	0.1853	0.1853	0.0619	0.0619
Observations	3390	3390	3390	3390

Source: Central Balance Sheet Data Office, Bank of Spain (*Central de Balances Anual*, or CBA)
Notes: Linear probability models. Dependent variable is a dummy that equals one if a firm has positive R&D investments in a given year (Columns (1) and (2)) or a dummy that equals one if a firm has positive payments for foreign technology transfers in a given year (Columns (3) and (4)). A unit of observation is a firm-year. The sample includes 1,942 medium and large firms (average number of employees = 389) over the years 1991–2007, who agreed to share their survey answers with researchers. Sector fixed effects are for 19 distinct sectors. Explanatory variable is firm log number of employees. Robust standard errors in parentheses. * 0.10 ** 0.05 *** 0.01.

---

[33]The literature has considered explanations for this premium similar to those that the firm-size literature has focused on: worker composition vs. rent-sharing or other labor market frictions.