

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

Jaime Arellano-Bover [†]

November 2019

Abstract

I study the link between firm heterogeneity and young workers' long-term career outcomes. Using administrative (social security) data from Spain, which include workers' labor market histories and education, I investigate the long-term effects of landing a first job at a large firm versus a small one. Size could be a relevant employer attribute for inexperienced young workers because large firms are associated with greater training, higher wages, and enhanced productivity. The key empirical challenge is selection into larger firms—for instance, more able people may land jobs at these firms. To overcome this challenge, I develop an instrumental-variables approach that leverages large firms' year-to-year idiosyncratic hiring shocks within each region. These large-firm hiring episodes, in turn, generate variation in the composition of regional labor demand. I find that starting at a larger firm leads to substantially better career outcomes such as lifetime income. To shed light on the mechanisms driving this result, I test whether the effect is (i) due to workers staying with their first employer; (ii) driven by job search mechanisms favoring large-firm workers; (iii) present for those who lose their first job; (iv) explained by experience acquired at large firms being more valuable. These tests find support for two complementary channels. The first is a job search channel by which a larger first employer leads to subsequent jobs at other large firms. The second is a human capital channel by which on-the-job skills developed in formative years are more valuable if they are acquired at larger firms.

*I thank Caroline Hoxby, Ran Abramitzky, Luigi Pistaferri, and Isaac Sorkin for their advice and encouragement. I have benefited from useful conversations with Manuel Arellano, Barbara Biasi, Nick Bloom, Mark Duggan, Thomas Ginn, Pete Klenow, Guido Martirena, Paul Oyer, Santiago Pérez, Alessandra Peter, Nicola Pierri, and Tom Zohar. I also wish to thank workshop participants at Stanford University and seminar and conference attendants at various institutions. Spanish Social Security staff were very helpful in providing me with special data extracts. This project benefited from the Leonard W. Ely and Shirley R. Ely Graduate Student Fellowship through a grant to the Stanford Institute for Economic Policy Research. Their financial support is gratefully acknowledged.

[†]Yale University and IZA. E-mail: jaime.arellano-bover@yale.edu.

1 Introduction

Firms are heterogeneous along many dimensions including pay, productivity, workforce 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 very 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, low managerial quality, 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 one's career path: 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. Abowd et al., 1999; Card et al., 2013; Song et al., 2015; Card et al., 2018). 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 whether and how first-employer heterogeneity impacts young workers' careers. I focus on firm size (number of employees) and document a causal relationship between holding a first job at a larger or smaller employer and long-term labor market outcomes. Size is an appealing firm attribute since it works as a sufficient statistic for a variety of hard-to-observe characteristics. Previous literature suggests that larger firms have higher levels of worker training, higher wages and productivity, better management, and distinct organizational practices.²

The thought experiment I seek to replicate is random assignment of inexperienced workers to be hired by a larger or smaller firm. Firms that differ in size are heterogeneous in other attributes associated with size; all of which form part of the thought experiment of being hired by potential first employers of different sizes.³ I develop an instrumental-variable (IV) approach to address non-random sorting of workers and firms. The IV is based on the timing of large firms' idiosyncratic labor-demand shocks in relation to workers' labor market entry and provides plausibly exogenous variation in the size of a worker's first employer. This approach is designed to (i) address any bias arising from workers' unob-

¹See Card et al. (2018) for evidence on firm-driven pay differentials, Syverson (2011) for differences in productivity, Lynch and Black (1998) for workplace training, Bloom and Van Reenen (2007) for management quality, and Fabiani et al. (2005) for technology adoption

²A longstanding literature documents a positive correlation between employer size and wages (e.g. 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 theoretical link between managerial talent and size goes back to Lucas (1978). Bloom and Van Reenen (2006) document a positive correlation between management practices and size. The hierarchical production literature (e.g. Garicano, 2000; Fox, 2009; Garicano and Rossi-Hansberg, 2015) provides theoretical and empirical evidence on the relationship between organizational practices and size.

³Hence, the thought experiment is *not* to exogenously increase the size of a given firm. Size *per se* might not be that relevant; what likely makes it relevant are firm attributes and practices that come with it.

servable characteristics correlated with first-employer size,⁴ and (ii) hold constant business cycle conditions at the time of labor market entry. Under the IV assumptions, I estimate the causal impact of first-employer heterogeneity, as measured by size, on workers' long-term career outcomes. My results show that matching with a larger first employer persistently improves labor-market prospects. The estimated effect is substantial, with an elasticity between lifetime income and first-employer size equal to 0.12. Analyzing the career dynamics that underlie this effect, the evidence is consistent with two complementary channels: larger firms providing more valuable skill-development (human capital), and larger first employers leading to subsequent jobs at other large firms (job search).

The data, which include working histories since their onset, allow me to construct a measure of lifetime income that I use as main outcome of interest. This measure aggregates, into a single metric, many years of monthly labor earnings, which include wages and unemployment benefits. Lifetime income is well-suited to study long-term consequences of early work experiences since it mostly captures labor market outcomes that happen many years after labor market entry. At the same time, in contrast with other approaches in studies of long-term effects, it is not an only outcome measured long after the treatment of interest. Instead, it captures the full stream of earnings since the beginning of the working life.

The first part of the paper documents a novel empirical fact: the relationship between the size of a worker's first employer and her lifetime income, which I call the first-employer size effect. Pooling data on workers of different cohorts and education levels I find a strong positive correlation in first-employer size and lifetime income (see Figure 1). This correlation is still strong after controlling for a range of observables.⁵ Still, this relationship could be driven by workers' unobservable characteristics at the beginning of the working life that persist over time. For instance, large firms might hire more productive people who would earn more throughout their career irrespectively of their first employer. Accounting for such unobservables motivates my IV approach.

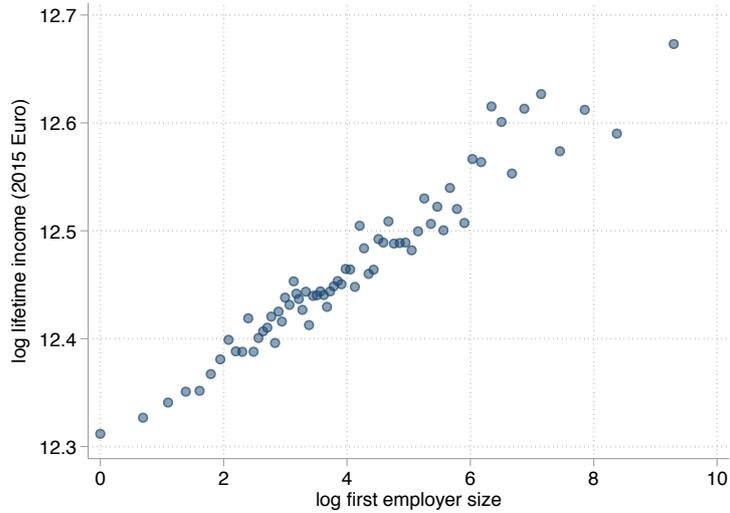
The IV strategy is based on variation in regional labor demand composition across time. The idea behind this approach is that idiosyncratic shocks in the hiring decisions of a *small* number of *large* firms can meaningfully affect regional labor demand composition. The IV aims to isolate changes in 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 occurs across years and Spanish regions. Depending on when and where a young person first enters the labor market, and who happens to be hiring, she will be exposed to different propensities to join larger or smaller firms.

A stylized example illustrates the IV approach. Consider two high school graduates who were both born in the region of Asturias (~1m population), one year apart from each other. The graduation year of the younger person is 1993, coinciding with the opening in the region of a large and modern plant of the U.S. multinational DuPont that hires around 1,000

⁴For example, more talented young people might be more likely to seek and find their first job at larger firms, which could be more desirable (Sorkin, 2018).

⁵These include workers' education, cohort, region of birth, or the business cycle conditions at labor market entry. This positive correlation also arises within two-digit first-employer industries.

Figure 1: Positive correlation between lifetime income and first-employer size



Notes: Raw correlation between workers' first-employer size and lifetime income. 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.

workers. The older worker's high school graduation was one year earlier. This timeline suggests that the worker from '93 will be more likely to have her 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.

The Spanish context provides rich variation in large-firm hiring shocks. During the sample years of labor market entry (1985–2003), Spain underwent an economic transformation following adhesion to the EU (1986). This period was characterized by an opening to trade, growth in foreign firms' investments, market reforms, and expansion of the country's regional infrastructures (Chislett, 2002). This led to great dynamism in large firms opening and expanding operations across different parts of the country. This variation allows me to keep constant effects that cyclical conditions during labor market entry might have on long-term outcomes, which has been the focus of previous work (e.g. Oyer, 2006; Kahn, 2010; Oreopoulos et al., 2012). I keep cyclical conditions constant by controlling for measures of regional unemployment rates and/or regional GDP growth, essentially using only the variation in large-firm hiring that is uncorrelated with business-cycle trends.

The IV aims to aggregate multiple episodes of large-firm hiring shocks into a single index. That is, to build an index measuring the labor demand composition inexperienced workers face across years and regions. I implement this 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 (predicted graduation year).

The exclusion assumption of the IV approach would be violated in the presence of unobservable labor supply factors at the birth cohort-birth region-education level that impact

lifetime outcomes and are correlated with the large-firm labor demand shocks the IV captures (Goldsmith-Pinkham et al., 2018; Borusyak et al., 2018). For instance, in the context of the DuPont example above, the IV assumption would not be satisfied if high-ability high school seniors in 1993 who *would have* gone to university do not do so as a result of DuPont's arrival. Then, the '93 high school cohort from Asturias would be of unobservable higher ability *and* more likely to start at a large firm. I conduct different checks to allay this type of sorting concerns. I show that the variation in labor demand isolated by the IV does not impact educational investments, nor does it correlate with the characteristics of workers' households and parents at age 17. Additionally, the empirical approach is based on *birth* region and *predicted* graduation year, as opposed to the region and year when a worker held her first job. That is, the IV assigns to each worker variation in large-firm hiring occurring in her region of birth in the year that, given typical completion times, a person of her birth cohort and educational achievement would usually graduate. This parses out potential workers' responses to local labor demand composition by moving to other regions or delaying graduation.

The results imply that first-employer characteristics can shape workers' career prospects. The estimated IV elasticity of lifetime income with respect to first-employer size is 0.12. The magnitudes are meaningful: a 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%. Thus, the first-employer size effect operates partly by shielding workers from unemployment.

The distinction between the IV and OLS estimates is of interest. The IV estimate of 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. I find evidence that compliers tend to be less educated and from less urban areas. It is plausible that these younger and less knowledgeable workers are the ones benefiting the most from a large first employer. They might have worse outside options or benefit more from on-the-job skills. This LATE result indicates that I capture the causal effect for the less advantaged young workers, who might be of particular interest.

I then study the economic forces underlying the link between first-employer size and career outcomes. I consider potential mechanisms and discuss characteristics of large firms that might explain them. First, I 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 mobility and rising wages), a time-varying elasticity showing first-employer size leading to higher wage *growth*, and first-employer effects that are still present at age 35 (an age at which incomes have stabilized and most people 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).⁶ Search frictions imply that wage growth arises from workers' "job-shopping", i.e., being able to find better matches over time. On-the-job human capital development raises workers' productivity, leading to wage growth.

First-employer characteristics could impact search for ensuing jobs. Search models indicate how frictions and 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; Lise, 2013; Krolkowski, 2017). If large firms are generally more desirable, a larger first employer would result in a "higher" starting point in the ladder.⁷ Workers would then only switch jobs voluntarily to move to an even more desirable firm. Persistent first-employer effects could then arise from subsequent jobs that are of differential quality based on first-employer size. I find that, consistent with this hypothesis, first-employer size has a positive causal effect on the size of ensuing employers. A larger first employer leads to larger subsequent ones.⁸

Young workers could acquire more valuable skills at large firms for reasons such as higher workforce training, learning from better peers and managers, conducting more productive tasks, or working with higher-quality inputs and outputs.⁹ Further, it could be especially productive to acquire these skills early in the working life, when workers are in a formative period. The skill formation literature emphasizes the complementarity in skills acquired at different points in time—"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).¹⁰

Differentiating a human capital channel from a job ladder channel is not straightforward, as both predict persistent effects. The key insight, present in models of on-the-job search, is that an unemployment spell disrupts a job ladder progression but has much less effect on human capital (unless the unemployment spell is long enough for skills to depreciate). Specifically, the unemployed lack a current employer as an option to weigh against new offers and are brought to the "bottom" of the ladder. Evidence of persistent first-

⁶Learning about a worker's unobserved abilities, or about job match quality is the third driver reviewed in Rubinstein and Weiss (2006) (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).

⁷Evidence from the U.S. supports, among mature firms, a job ladder in terms of firm size (Haltiwanger et al., 2018). Sorkin (2018) documents that, when taking into account both pay and non-pay amenities, there is a positive correlation between firm desirability and size.

⁸While consistent with a job ladder hypothesis, this could also be consistent with a human capital explanation if skills learned at a large first employer are more valuable at other large firms.

⁹Macis and Schivardi (2016) suggest export wage-premiums are larger for workers with previously existing export-related experience; at the same time large firms are more likely to be exporters. Kugler and Verhoogen (2012) document a positive correlation between plant size and the quality of their inputs and outputs. Larger employers tend to be more productive (Moral-Benito, 2018), and faster to adopt new technologies (Fabiani et al., 2005). For studies on learning from coworkers see Nix (2017) and Jarosch et al. (2018).

¹⁰While this literature focuses on childhood, their model can be applied to skill development of young adults. Evidence indicating that cognitive abilities start declining when people reach their 20s (see Salthouse, 2009, and literature cited therein) is indicative of the importance and productivity of skill acquisition during the *early* stages of the working life.

employer effects among those experiencing unemployment spells would be consistent with a human capital channel. The data support this hypothesis: the first-employer size effect is present for the group of workers in my sample that experience an unemployment spell between their first and second jobs.

If skills acquired at large employers are more valuable, we would expect higher returns to *experience* obtained in large firms than that obtained elsewhere. I test this prediction estimating monthly-panel wage equations that, while controlling for unobserved worker heterogeneity and *current* employer characteristics, allow differential returns to experience acquired at the larger employers.¹¹ 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.¹² Further, person fixed effects absorb unobservable time-invariant worker characteristics. The results show that one year of experience in the largest firm-size group is around 2.5 percent more valuable than one year of experience acquired elsewhere. The positive differential is present for all current-employer sizes, which is consistent with the portability of these skills. Finally, using data on professional categories, I show that the differential return to large-firm experience is also present on the arrival rate of promotions. Experience at large employers is associated with faster career progression.

Overall, my findings imply that working for different firm types as a young person can have effects that last throughout one's career. Size seems to be a relevant measure, capturing meaningful employer characteristics. Further, size has the advantage that is observable to all and, not being model-based, transparently measured. Human capital is a channel that is consistent with the evidence. Firm heterogeneity in the provision of on-the-job skills has interesting implications: In the presence of imperfect wage adjustment and worker mobility, firms that increase young workers' productivity in persistent ways might not fully internalize this in their operational decisions. Additionally, the efficiency losses some argue arise from size-dependent policies and regulations (IMF, 2015; Guner et al., 2008) could be magnified if larger firms provide more valuable skills.

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 as further evidence of a human capital channel. Section 7 discusses large-firm attributes and practices that could impact young workers' human capital development. Section 8 concludes. Different **appendices** provide supplementary analyses and robustness checks.

¹¹The data allow me to measure experience rather precisely, in number of days worked at each employer.

¹²Keeping constant current employer characteristics controls, on some level, for job-ladder benefits of large-firm experience. In fact, the returns to large-employer experience are larger when not controlling for current-employer size.

2 Relation and Contribution to the Literature

This paper is the first to establish a direct causal link between young workers' first-employer characteristics and long-term career outcomes.¹³ This relates to a growing literature on firm-driven wage differentials (e.g. [Abowd et al., 1999](#); [Card et al., 2013](#); [Song et al., 2015](#); [Sorkin, 2018](#); [Abowd et al., 2018](#); [Card et al., 2018](#)). These studies suggest that equally-skilled workers can earn substantially different wages at different employers.¹⁴

While this literature has studied contemporaneous worker-firm relationships, analyzing how a worker's current wage can vary depending on who her current employer is, it has not focused on persistent effects stemming from past employment at heterogeneous firms.¹⁵ This paper shows how and why early-career firm heterogeneity can have persistent implications for long-term inequality.

Additionally, this literature has generally lacked contexts in which workers are exogenously driven to leave or join particular types of firm. The typical empirical approach relies on workers' firm-switching to identify firm effects, considering the timing and destination of moves to be as good as random after controlling for worker effects. By contrast, I use plausibly exogenous variation in the probability of a young worker joining a larger or smaller employer. I can do this thanks to two features of my empirical design. First, I focus on inexperienced workers entering the labor market for the first time after graduation. This is a group of people who search for jobs at a well-defined point in time for external reasons (typical educational degree length), so focusing on them avoids studying job-switching of experienced workers, which are typically the result of endogenous moving or layoff decisions. Second, I develop an instrumental variable that captures fluctuations in large-firm labor demand, which avoids biases arising from the endogenous sorting of workers and firms.

A long tradition in labor economics, going back to [Moore \(1911\)](#), documents a robust positive correlation between a worker's earnings and the size of her employer, the employer-size premium ([Brown and Medoff, 1989](#); [Oi and Idson, 1999](#)). This is a pervasive labor market finding in different contexts.¹⁶ The literature, however, has not agreed on why the premium exists, nor determined whether it has a causal component.¹⁷

I document in a causal way that young workers' who match with larger first employers enjoy better long-term labor market outcomes. Additionally, I provide understanding of the

¹³Other papers studying this question 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](#)).

¹⁴A related strand of literature on firm-driven wage inequality studies rent-sharing, estimating spillovers of firm-level productivity shocks into workers' wages. See [Card et al. \(2018\)](#) for a review.

¹⁵[Abowd et al. \(2018\)](#) and [Bonhomme et al. \(2018\)](#) provide 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. \(2018\)](#) document how a worker's past firm-type may impact her earnings after changing jobs.

¹⁶[Lallemand et al. \(2007\)](#) document the size premium for Spain and other European countries. Recent work documents a declining firm-size premium in the U.S. but not in other countries ([Bloom et al., 2018](#); [Colonnelli et al., 2018](#)).

¹⁷Some 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 arguably endogenous worker characteristics (i.e. marital status or family composition).

channels generating this relationship. I show how *first*-employer size is relevant in the long term, even after changing jobs. A mechanism that is consistent with the evidence is large firms providing better skill-development opportunities for their young workers. I provide evidence indicating that the returns to experience acquired at large firms are differentially more valuable than experience acquired elsewhere.

Other work 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, 2018). The *first-market-tightness* effect documented in these papers is related to the first-employer size effect since inexperienced workers are more likely to be hired by large employers during booms, a pattern that arises in my data and has previously been documented in other contexts (Oreopoulos et al., 2012; Brunner and Kuhn, 2014). This is also consistent with the work of Moscarini and Postel-Vinay (2012) who find a negative correlation between firm-level net job creation and unemployment levels that is stronger for larger employers. Other work related to *first-market-tightness* includes Beaudry and DiNardo (1991), Devereux (2002), and Oyer (2006).¹⁸

This literature has several examples of credible empirical strategies, confidently establishing the *first-market-tightness* effect. However, we know less of the underlying mechanisms. Oreopoulos et al. (2012) do some work on this front by documenting that graduating during a recession leads to higher job mobility and initial and persistent matches with lower quality employers (measured by size and average wages).^{19 20}

This paper helps to improve our understanding of the mechanisms for why initial labor market entry is so important, finding explicit support for both job search and human capital explanations. More concretely, it also informs which are plausible mechanisms behind the *first-market-tightness* effect. This is because I study the effects of first-employer heterogeneity, one of the suggested mechanisms in this literature, but do so while keeping constant business-cycle fluctuations.²¹ Documenting the first-employer size effect in (i) a causal way, and (ii) net of cyclical variation, helps narrowing down likely mechanisms behind the broader *first-market-tightness* effect. The first-employer size effect, in turn driven by

¹⁸Beaudry and DiNardo (1991) find that the lowest unemployment rate realized since workers began working for their current employer is a better predictor of current wages than the initial unemployment rate. In unpublished work, Devereux (2002) tests whether starting wages at a new job impact the level of wages four years after. He uses the state unemployment rate at the time of job start as an instrumental variable and concludes that there is a positive effect, arguing that signaling and human capital explain this persistence. Oyer (2006) documents that Ph.D. economists graduating in times of lower unemployment get better initial matches and subsequent professional outcomes.

¹⁹However, 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.

²⁰Other work aims to better understand persistent first-market-tightness effects. Kwon et al. (2010) argue that those who entered the labor market during a boom get promoted faster. Liu et al. (2016) suggest that skill mismatch during entry is strongly countercyclical. Wee (2016) argues that labor market entry during a recession hinders learning about comparative advantage and human capital accumulation

²¹In previous papers cyclical variation was either the “treatment” of interest (e.g. Kahn, 2010; Oreopoulos et al., 2012), or used as an instrument driving workers into different jobs (e.g. Devereux, 2002; Oyer, 2006).

job search and human capital channels, seems to underpin part of the *first-market-tightness* effect.²²

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). Relevant 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;²³ Müller and Neubaeumer (2018) argue that training at a larger firm leads to lower long-term unemployment. This paper contributes to this literature by documenting a causal relationship between career outcomes and young workers' first job characteristics, and finding support for an early-career skill-development explanation.

By documenting differential wage growth paths induced by heterogeneous first employers, this paper contributes to a large 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.²⁴

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 Continua de Vidas Laborales*, or MCVL), a 4% non-stratified random sample of Spanish Social Security administrative records, extracted yearly from 2004 to 2015.²⁵ 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.²⁶

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

²²In Section 4.10 I perform a simple exercise that allows to quantify this relationship.

²³Both von Wachter and Bender (2006) and Fackler et al. (2017) find larger losses for German workers who involuntarily separated from larger firms.

²⁴I 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 are not likely to explain persistent entry-cohort effects. That is, different cohorts of new recruits *in a given firm* seem to enjoy similar skill-development opportunities. This contrasts with this paper, where I argue that differential skill investments *across different firms* seem to partly explain persistent first-employer effects.

²⁵Throughout the paper I use complementary data sources that I describe in Section B.1 of Appendix B.

²⁶I do not consider workers who were only sampled in 2004 since the 2004 MCVL did not include workers' education.

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,²⁷ and each month's payroll taxable base.

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.²⁸ I follow a procedure similar to [Bonhomme and Hospido \(2017\)](#) to impute monthly earnings for censored observations.²⁹ Censored observations are relatively few.³⁰

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 will include unemployment income in my main lifetime income measures.

These social security records are also matched with tax records recording uncensored annual earnings for the years 2006-2015. The downside from using this measure of uncensored earnings to study long-term effects is that the time series is significantly shorter. However, I later show how the main lifetime results are robust when using measures 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,³¹ each firm typically 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 (see Appendix Figure A2). This implies that an employer identifier in the data represents a firm-province combination. This notion of employer will be equivalent to a firm for single-establishment firms, and smaller than a firm for firms operating in different provinces. Firm-province is the employer definition that I use throughout the paper.³² Since I focus 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

²⁷Professional category (*grupo de cotización*) is determined by the type of job a worker performs and not by her education level. It is used administratively to determine certain things such as taxable base maximums. I use this variable later on when I study promotions.

²⁸Since the taxable base of the self-employed is not a function of their monthly income, I do not observe earnings for self-employed workers.

²⁹This 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.

³⁰In my monthly panel sample 8.7% and 3% of observations are top- and bottom-coded respectively.

³¹More than 95% of Spanish workers are in the general regime of Social Security ([Bonhomme and Hospido, 2017](#)). For example, certain civil servants and agricultural workers are excluded from the general regime.

³²This definition notwithstanding, I use the words *firm* and *employer* interchangeably in this paper.

whenever it differs from *firm-province*.

For each employer I observe its location, sector, age, and number of workers. Number of workers is my measure of employer size. I also observe a *firm* identifier which groups together employers belonging to the same firm. While this allows me to identify two sampled employers that belong to the same firm I still use firm-province as the employer unit. This is 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.

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 relevant employers in my sample going back to 1980. This 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).³³

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.³⁴

I apply several sample restrictions. I restrict 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 relying on a person's (Spanish) region of birth. The lifetime analysis requires me to observe each worker a sufficient number of years. The data imposes a tradeoff between how many cohorts I study and how many years I follow each worker. Balancing this tradeoff, I restrict attention to the 1968–1980 birth cohorts while they are aged 16–35. I only include those who, between 16 and 35 years of age, predominantly work as wage earners.³⁵ 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

³³While the information on workers' histories is provided retrospectively, the original MCVL extracts provide information on employer size (number of workers) *at the time of data extraction* (2004–2015). Since most of the workers in my analysis have their first job before 2004, I augmented the data with a special extract recording employers' sizes going back in time. This allows me to observe employer size at the time each worker in my sample joined his first employer. MCVL staff prepared this special extract. It contains employer size across time (back until 1980) for the set of employers who are the first or second employers of workers in my cross-sectional lifetime analysis sample. For the rest of employers, I observe employer size starting in 2004.

³⁴If a worker has more than one employer in a given month, I add earnings from the different employers while keeping the characteristics of the main employer (the one who provides higher earnings that month).

³⁵I exclude those who are self-employed for 40% of the time or more during this period

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. I include those who, between 16 and 35 years, have sufficient attachment to the formal labor market.³⁶ I exclude workers who have their first job in the public sector, have their first job very late (later than age 22 for high school graduates, 25 for vocational, and 28 for college),³⁷ or in a Social Security regime different than the general regime.³⁸ These restrictions result in a sample of around 80,000 people. Table 1 provides summary statistics for this 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.

³⁶This implies restricting attention to 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 (see [Guvenen et al., 2017](#)).

³⁷Those 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.

³⁸For the minority of workers outside the general regime, the number of workers in their same Social Security account might not be informative of the size of their employers. See footnote 31.

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.³⁹ This definition aims to capture the first relevant job after finishing formal education, while avoiding summer work or very temporary employment.⁴⁰

First-employer size. For each worker I assign as his first-employer size the size of his employer during his first labor market experience.⁴¹ For the small fraction of workers who have more than one employer during this semester, I assign the maximum size across these employers.⁴²

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 that a worker gets between labor market entry and some age T .⁴³ I include as labor income both earnings and unemployment benefits. A potential return to different first employers could come in the form of unemployment benefits. This is due to high unemployment rates in Spain and the fact that benefits are a function of previous earnings. The lifetime income measure takes the following form:

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

Where w_{it} are monthly earnings, u_{it} are monthly unemployment benefits, and δ is a discount rate. I do not discount the flow of income in the main analyses but I show how the main results are robust to different discount rates.

In practice 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 A5 for evidence for Spain on income profiles stabilizing during the mid 30s). Table 1 provides summary statistics for this measure.

³⁹Based on standard Spanish completion times I assign year of predicted graduation as the year in which people with a high school degree turn 17, 20 for vocational education, and 23 for college education. I do not observe actual graduation year in the data.

⁴⁰Panel (a) in Appendix Figure A9 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 Appendix Figure A9 plot the distributions of age and days worked during the first labor market experience.

⁴¹In robustness checks I also use average size during the four years prior to the worker joining the firm.

⁴²Panel (d) of Appendix Figure A9 plots the distribution of first-employer (log) size. Panel (e) shows that the vast majority of workers only have one employer during this semester.

⁴³The availability of administrative data spanning long periods of time has started to allow analysis that measure and analyze lifetime income. For recent examples see [Guvenen et al. \(2017\)](#) and [Nybom \(2017\)](#).

The median is 254,142 Euro (2015).^{44 45}

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 either zeroes or unemployment benefits) deals with this issue.

This measure is particularly interesting and well-suited for the study of the long-term consequences of a first employer. To a large extent this outcome is the result of what happens many years after labor market entry. Studies of long-term effects often rely on a single cross-section taken a long time after the "treatment" (a long "lag"). I avoid this by aggregating many years of data month by month. This is possible thanks to observing workers' labor market histories since their start and for a prolonged period of time.

3.4 Large and small firms in Spain

To inform the analysis in this paper, some remarks on the Spanish firm-size distribution are in order. Compared to other countries, Spain has very few large employers. According to OECD data from 2013, around 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 A1). 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).

Compared 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 Spanish worker who does not match with a large employer might disproportionately be a very small, 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 A1). 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 (see Appendix Figure A3).

⁴⁴In 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).

⁴⁵Appendix Figure A4 illustrates why incorporating unemployment benefits is potentially important. It shows the percentage of lifetime income coming from unemployment benefits, by deciles of lifetime income. For those in the lower part of the lifetime income distribution this percentage amounts to between 7.5 to 4.5 percent.

4 Size of First Employer and Career Outcomes

This section documents the relationship between the size of a worker’s first employer and her 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 small firm. Larger firms are characterized by attributes that could potentially be relevant young workers, likely driving any first-employer size effect.⁴⁶

4.1 Descriptive Facts

There is an unconditional positive relationship between the size of a worker’s first employer and her long-term career outcomes. Figure 1 above shows the relationship between my measure of lifetime income and first-employer size (pooling together workers of different cohorts, locations, and education levels). There is a strong positive relationship between the two which is rather linear in logs. The correlation coefficient is equal to 0.21.⁴⁷

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. Appendix Figure A7 shows that similar patterns arise when looking at employment and daily wages.

4.2 Accounting for Workers’ Observable and Unobservable Characteristics

The goal is to estimate the elasticity of a worker’s lifetime income with respect to the size of her first employer. The elasticity is given by β in the expression

$$y_i = \beta s_{J(i)} + X_i' \gamma + \varepsilon_i. \quad (2)$$

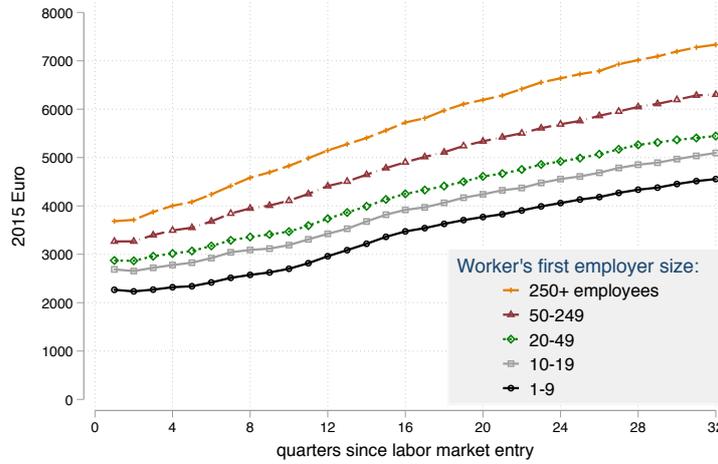
Where y_i is a measure of (log) lifetime income for worker i , $s_{J(i)}$ the (log) number of employees of firm J where i held her first job, X_i are a set of covariates, and ε_i is an error term.⁴⁸ All variation is cross-sectional since each worker only has one first job and one level

⁴⁶In Section 7 I discuss what these attributes and mechanisms 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 Section 7 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).

⁴⁷This relationship is not explained away by firms’ sector of activity. If this were true, using size as a relevant summary measure of firm attributes would still be a valid approach. It would, however, lead us to consider a different set of mechanisms explaining the lifetime premium. Appendix Figure A6 shows that the pattern from Figure 1 holds within sectors (controlling for 58 sector fixed effects).

⁴⁸I measure first-employer size at the year the worker joined the firm. The main results are robust to alternative size measures.

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.

of lifetime income.

If young people entering the labor market were randomly assigned to firms, β could be consistently estimated by OLS. In practice workers are not randomly assigned to firms and, possibly, unobserved determinants of lifetime income are 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 are some of the reasons that would bias OLS estimates of β and the motivation for the IV approach that I develop below.⁴⁹

4.3 Empirical Approach: Estimating Equation and IV

A version of equation (2) is the baseline estimating equation:

$$y_i = \beta s_{J(i)} + f(u_{r,t_0(e,c)}) + \delta_r + \delta_e + \delta_c + \varepsilon_i. \quad (3)$$

Where y_i , $s_{J(i)}$, and β have the same interpretation as above. c indexes birth cohorts, e refers to three educational attainment levels -high school, vocational, college- and r indexes regions of birth.⁵⁰ $t_0(e, c)$ 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 δ s represent region

⁴⁹ [Eeckhout and Kircher \(2018\)](#) develop a theoretical model in which firms jointly choose the amount and quality of labor, thus jointly determining the firm-size distribution and worker sorting.

⁵⁰ Spain is divided into 17 regions (Comunidades Autónomas) plus two North Africa enclaves (Ceuta and Melilla). Regions are further divided into provinces (see Appendix Figure A2).

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.

Note that 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 to partial out this channel from the effect of starting off 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. 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.⁵¹

The IV approach uses variation in the *composition* of regional labor demand 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.⁵² 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, she 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 (population of around 1 million), 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, who 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.

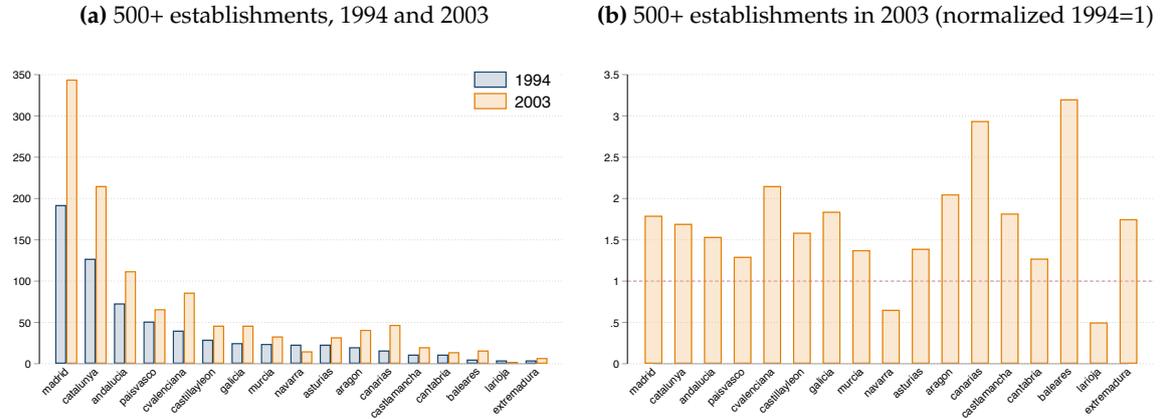
The goal of the IV is to aggregate and summarize this type of large-firm hiring shocks across years and regions in my sample. Ideally, the DuPont example would be just one of many large-firm labor-demand shocks. Fortunately, the institutional and historical context provides a setting of rich variation. 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 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 data on the population of establishments. For

⁵¹Later I show that using regional GDP growth as additional cyclical control does not affect results.

⁵²In relation to this idea, Gabaix (2011) shows how shocks to a small number of large firms can generate business cycles. However, as I explain below, I use variation in large/small firm labor-demand composition while holding business-cycle variation *constant*.

each region, the figure shows the number of establishments with 500+ employees in 1994 and in 2003.⁵³ In 15 out of 17 regions the number of large employers increased, and in most of them substantially so. This is true even for regions that had initially fewer large firms.

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



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 latter part of sample period of labor market entry. Sample period of labor market entry is 1985–2003, DIRCE only go as far back as 1994.

I construct an index that captures the variation in labor-demand composition a given 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 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.⁵⁴ 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 getting their first job at i ’s region of birth, and are doing so during i ’s predicted graduation year.⁵⁵ 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 $l = 1, \dots, N$ indexes workers in my sample and $\mathbb{1}\{\cdot\}$ is the indicator function. The IV approach predicts worker i ’s (log)

⁵³The 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.

⁵⁴An alternative approach would be to use data on vacancies to construct the IV. Systematic vacancies data does not exist in Spain.

⁵⁵I 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} .

first employer size, $s_{J(i)}$, with

$$\bar{s}_{-i}^{rec} = \ln \left(\frac{\sum_{l \neq i} \exp(s_{J(l)}) \cdot \mathbb{1}\{\tilde{r}_l = r_i, e_l = e_i, t_l = t_0(e_i, c_i)\}}{\sum_{l \neq i} \mathbb{1}\{\tilde{r}_l = r_i, e_l = e_i, t_l = t_0(e_i, c_i)\}} \right). \quad (4)$$

4.4 IV Discussion

IV variation and the business cycle

It is expected that the differential hiring patterns of large versus small firms will be correlated with business cycle variation (Moscarini and Postel-Vinay, 2012). Indeed, pooling variation across time and regions, there is a negative correlation in my data between regional unemployment rates and the average employer size of new hires (see Section 4.10).⁵⁶ However, the empirical approach summarized in equation (3) is aimed at shutting down any impacts that business cycle conditions at entry might have on long-run prospects. This comes both through the inclusion of region and cohort fixed effects, and through explicitly controlling for a flexible function of the unemployment rate a worker faces during labor market entry.

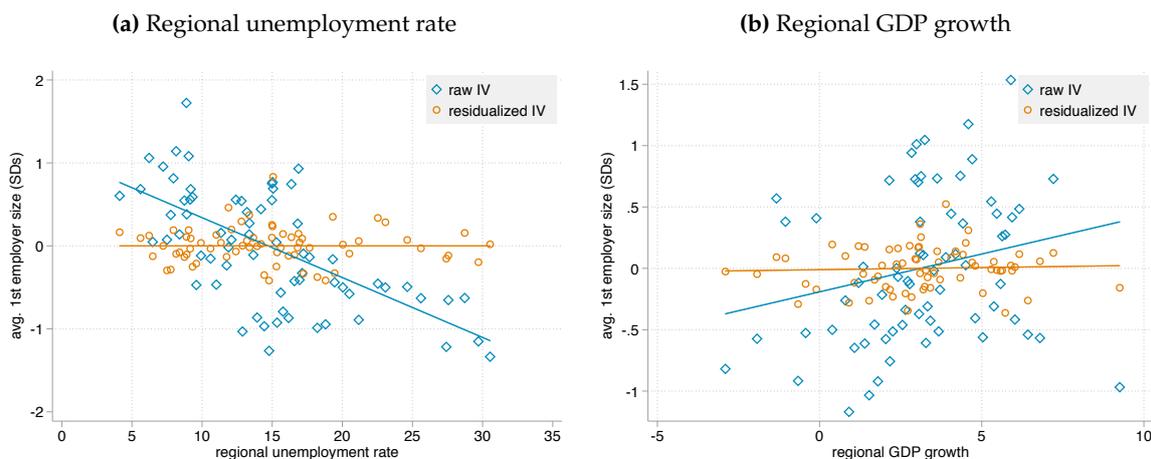
Figure 4 further illustrates how cyclical conditions are held constant and the residual variation the IV approach uses. Panel (a) plots the correlation between the unemployment rate at entry and i) the IV \bar{s}_{-i}^{rec} (*raw IV*), and ii) residuals from a regression of \bar{s}_{-i}^{rec} on δ_r , 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). 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.⁵⁷

While controlling for the unemployment rate is meant to capture cyclical variation one could worry that, being a single indicator, it might not do so completely. Panel (b) on Figure 4 allays these concerns using data on regional GDP growth rates. This is a different cyclical indicator and, since it is excluded from the specification in equation (3), it is not *mechanically* unrelated to the IV residual variation. Reassuringly, a similar pattern as with regional unemployment rates emerges. There is a positive correlation between the instrument and regional GDP growth (black dots). This is, again, consistent with the idea that, uncondition-

⁵⁶The fact that young workers who enter the labor market during a recession start at 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). In Section 4.10 I further study the relationship between cyclical conditions and first-employer size, and quantify the relationship between the effects of starting at a larger firm (holding unemployment constant) and the effect of unemployment at entry.

⁵⁷Appendix Figure A11 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.

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



Notes: Binned scatterplots of the IV \bar{s}_i^{rec} described in the text (*raw IV*) and residuals from a regression of \bar{s}_i^{rec} 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*). 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.

ally, large-firm hiring is more prevalent during good economic times. Controlling for fixed effects and the unemployment rate (blue diamonds) results in the IV residual variation having a flat relationship with cyclical conditions as measured by regional GDP growth.⁵⁸ The results from this test indicate that the empirical approach and the IV variation are making a good job in keeping business cycle conditions constant.⁵⁹

Region of birth and predicted graduation year

Note that the instrument and covariates depend on *region of birth* (not first job region), and *predicted* graduation year (not actual graduation year, or first job year). In this way I avoid implicitly controlling for potentially endogenous outcomes (“bad controls”) such as migration, or strategic timing of graduation and/or labor market entry.^{60 61}

Instrument's predictive power: First stage

For the IV approach to work sufficient large-firm hiring variation after controlling for cyclical conditions is needed. This seems to be the case, in accordance with the DuPont

⁵⁸ Appendix Table A1 shows the regression results underlying Figure 4.

⁵⁹ Additionally, as a robustness check, I estimate versions of equation (3) that in addition to controlling for the regional unemployment rate, they will also flexibly control for regional GDP growth.

⁶⁰ According to Spanish Education Ministry data, in 2014 86% of college graduates enrolled in their region of origin. Labor force survey data (EPA) from 1992–2015 shows 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.

⁶¹ Appendix Figure A10 suggests why this might be relevant. It reveals two sources of potential endogeneity. Unemployment rates are on average lower at the actual region and time of first job, than at the region of birth and predicted graduation year. This implies that people move across regions and delay their first job in ways that correlate with labor market conditions. These patterns support the choice of basing the empirical approach on region of birth and predicted graduation year: any endogenous migration or timing choices will not bias my estimates.

example and the description of the 1985–2003 institutional context. Appendix Figure A12 shows a histogram of (residualized) labor demand instrument values across region of birth \times education \times year of birth cells. Values are normalized so that they are expressed in units of standard deviations of (residualized) first employer size. The figure shows substantial dispersion across cells, with a standard deviation of 0.387 standard deviations of (residualized) first employer size.

The IV variation does a good job at predicting first-employer size. This is seen graphically in Appendix Figure A13. The F-statistic of the excluded instrument is equal to 24.3 (see Table 2 below).

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 *and* more likely to have a large first employer. This 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 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 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. Note that this type of test also works to allay 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

In my empirical approach 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 about this is that educational attainment could be endogenous in this setting, as opposed to predetermined like region and year of birth.⁶² This type of concern warrants consideration based on evidence on the countercyclicality of education enrollment decisions (e.g. Card and Lemieux, 2000; Petrongolo and San Segundo, 2002; Sievertsen, 2016).

There are features of my empirical approach that somewhat relax these worries. As described above, I make an effort to keep constant business cycle conditions. While in the main specification in equation (3) I control for the unemployment rate only at the predicted graduation year (of actual educational attainment), as a robustness check I also estimate specifications that control for past regional unemployment rates. Given this, 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.⁶³

I test for this possibility studying whether, after controlling for the unemployment rate, 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, such as the DuPont arrival, 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.

Similarly, these types of worries should be more pressing if the large-firm hiring shocks 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,

⁶²Whether education is predetermined or not is relative to the question one is studying. In a life-cycle sense it is not predetermined. If one is studying, say, labor supply decisions of people over 65 as response to contemporaneous policy, it might be reasonable to consider it predetermined. It might be less straightforward thinking about this when studying the characteristics of a workers' first employer and its later consequences, making it worthwhile discussing here.

⁶³Also 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.

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 0.15.⁶⁴ 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.⁶⁵

4.5 Lifetime Income: Results

Table 2 shows OLS, first stage, and second stage results of estimating β in equation (3) using the proposed instrumental-variables approach. 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 $\{\text{region of birth} \times \text{educational attainment} \times \text{birth cohort}\}$ -level since this is the level through which the IV operates (Abadie et al., 2017).⁶⁶ Column (5) shows first-stage results.⁶⁷ The instrument does a good job at predicting first job size, with an excluded instrument F-statistic of 24.3. 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. Below I discuss potential reasons behind the difference in OLS and IV estimates and explore this issue in more detail in Appendix C.

4.6 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. Also, 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 compliers' LATE argument (Imbens and Angrist, 1994) would explain the relatively-high IV magnitude. I now explain how this is a plausible scenario, and provide evidence consistent with it.⁶⁸

The first thing to ask is who, given the nature of my 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

⁶⁴0.15 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).

⁶⁵Davis 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.

⁶⁶Appendix Figure A13 provides graphical evidence of the IV result, showing the variation that identifies β through the first and second stage of TSLS.

⁶⁷Appendix Table A10 displays estimates of the reduced-form.

⁶⁸Unobserved ability bias by itself, by which more productive workers match with larger firms, would 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.

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

	OLS				First Stage	Second Stage			
	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.

compliers. 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 labor demand scenarios. 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.

This intuitively suggests that the less able or less knowledgeable young workers comprise the group of IV compliers, a possibility I explore and confirm more formally in Appendix C. 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. This exercise is useful towards formally thinking of how does a compliers' LATE argument translate into an empirical setting such as mine with a continuous treatment and a continuous instrument. This exercise confirm that workers who are less educated (and thus, younger) and originally from less urban areas are the ones mostly driving my TSLS estimate (see Appendix Figure C2).⁶⁹

Next, I consider whether it is likely that these younger and less knowledgeable workers have the highest long-term benefits of a larger first employer. Assessing this requires reviewing the mechanisms behind the first-employer size effect. I explore mechanisms in the following sections and the main conclusion is that both job search and human capital seem to be two main driving channels. They are both consistent with the first-employer size effect being larger for the younger less knowledgeable. Take human capital. If large employers offer better on-the-job skill development opportunities, those who start out with the least skills might benefit more from training and the day-to-day learning of working in a large organization. Job search points in a similar direction. A higher ability young adult who initially matches with a worse employer is more likely to successfully find a better match while employed, reducing the overall impact that the first employer might have over her career.

Overall, the evidence suggests that the less able and less knowledgeable are more influenced by the labor demand variation my IV uses. Further, there are good reasons to believe these are precisely the young workers who might benefit the most from matching with a large first employer.⁷⁰ This would be consistent with and explain the higher magnitude of the IV estimates with respect to OLS. Finally, the LATE result should be kept in mind when

⁶⁹The 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 these weights and, when allowing them to vary across groups, I find them to be consistently larger for those without a college degree and those born in less urban parts of Spain.

⁷⁰The evidence from Bonhomme et al. (2018), whose findings indicate that "lower-type" workers gain the most from employment at a "higher-type" firm, is consistent with this idea.

interpreting the results. However, even if the estimated magnitudes of the first-employer size effect are not representative for all workers, I seem to be capturing the causal effect for the less advantaged young workers who might be of particular policy interest.

How is this sustained in equilibrium? An interesting question outside the scope of this paper is to think of 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 the human-capital benefits a given worker derives are proportional to the costs she generates for the firm (either through explicit training, or from the time it takes her to learn the ropes of the job). Firms might not want to, or not be able to, discriminate wages based on this. Even if firms did lower wages to equalize long-term benefits, there are reasons why young workers could turn down such a deal. This would be the case if either i) they are not aware of the long-term benefits of such an initial job, or ii) they are aware of the benefits but not able to borrow against them and, being cash-constrained, it is not worth to accept a lower wage offer that factors in these benefits.

4.7 Lifetime Income: Robustness and Extensions

Robustness. I show that the IV elasticity of lifetime income with respect to first-employer size is robust to several alternative specifications in Section B.4 of Appendix B. I show that the results are stable when controlling for regional unemployment in an alternative way, 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, and when controlling for first-employer location (province) size. In Section B.5 I show that the first-employer size effect is robust to using uncensored measures of income found in the 2005–15 tax records merged with the social security data.

Varying elasticity across firm sizes. In Section B.6 of Appendix B I relax the constant elasticity assumption implicit in equation (3). 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 (3) 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.

4.8 Wages, Employment, and Earnings

The lifetime income effect of matching with a larger first employer could combine effects on different margins. Indeed, lifetime income is a combination of the 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 (3) replacing lifetime income y_i with three different outcomes (in logs): average daily wages, total days worked, and lifetime earnings (which do not count unemployment benefits). Table 2 shows OLS, first stage, and second stage 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, or 94% of the elasticity of lifetime income (equal to .117, see column (6)). 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%). These numbers are consistent with the first-employer size effect arising from employment and average wages. Further, the higher earnings stream seem to imply that whenever workers do experience unemployment, they are entitled to higher benefits.⁷¹

4.9 Magnitude of Estimated Effects

The baseline estimated elasticity of lifetime income with respect to first-employer size is .117. At least for the group of relevant compliers, then, underlying differences across larger and smaller firms result in that matching with a 10% larger first-employer leads to 1.17% higher lifetime income. A different way to interpret this magnitude is using the standard deviation of log first employer size, 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%.⁷² The corresponding number for average daily wages (estimated elasticity of 0.082) is 18.8%.

Consider now the difference between starting out at a first employer that is either i) a micro enterprise (1–9 workers), or ii) a large enterprise (250+ workers).⁷³ A first-employer size increase from 9 to 250 workers is substantial. In the empirical first-employer size distribution 9 and 250 correspond to the 34th and 85th percentiles respectively.⁷⁴ The estimated elasticities imply that this large increase in first-employer size leads to 31% and 47% higher average daily wages and lifetime income, respectively.⁷⁵

As I discuss in the next section, heterogeneous on-the-job skill acquisition during for-

⁷¹In Section B.7 of Appendix B I test for the relationship between first-employer size and job security later on in the working life.

⁷²From $100 \cdot [\exp(2.1 \times .1166) - 1] = 27.74$

⁷³OECD defines as micro enterprises those with fewer than 10 employees. Those with 10 to 49 employees are considered small, 50–249 medium, and 250+ large.

⁷⁴Put differently, it corresponds to a 1.6 standard deviations increase in log size.

⁷⁵From $100 \cdot [\exp((\ln(250) - \ln(9)) \times .082) - 1] = 31.34$, and $100 \cdot [\exp((\ln(250) - \ln(9)) \times .1166) - 1] = 47.35$

mative years seems to be one of the driving channels of the first-employer size effect. An illustrative benchmark when studying the consequences of young workers' first experiences in the labor market could thus be the returns to higher education. There is a vast literature on this topic (see [Oreopoulos and Petronijevic, 2013](#), for a review). Results from a recent paper by [Nybom \(2017\)](#) are particularly well-suited as a benchmark, since he uses measures of lifetime earnings similar to mine. Using administrative data from Sweden, he estimates the lifetime returns to college to be 24.51% on average (ATE) and 32.25% for those who do attend college (ATT).

This discussion suggests that the magnitudes behind first-employer size effect are economically meaningful. Starting as a young worker with one or another employer can lead to very different labor market experiences in the long run; particularly so for the younger and less knowledgeable workers whose first-employer match is more sensitive to the IV large-firm hiring shocks.

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

In this section I perform a simple exercise to 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, 2018](#)).

There are reasons to expect that the persistent positive effect of starting at a larger employer that I document in this paper partly explains the negative effects of entering the labor market during a recession. [Moscarini and Postel-Vinay \(2012\)](#) provide evidence indicating that, in aggregate terms, large-firm hiring is differentially prevalent during good economic times. [Oreopoulos et al. \(2012\)](#) and [Brunner and Kuhn \(2014\)](#) provide direct evidence that workers that enter the labor market during a recession hold their first job at smaller firms.

I begin by estimating the following regression in my sample:

$$s_{J(i)} = \gamma u_{r,t_0(e,c)} + \delta_r + \delta_e + \delta_c + \varepsilon_i. \quad (5)$$

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 δ s 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 shows OLS estimates of γ in equation (5) for the whole sample and different subgroups. The negative estimates of γ are consistent with previous literature and the evidence found on Figure 4 and Appendix Figure A10. The estimate for the full sample in column (1),

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

	first employer size		
	(1)	(2)	(3)
unemployment rate at entry	-0.0099*** (0.0035)	-0.0117*** (0.0044)	-0.0166*** (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) only the subsample of those with a high school or vocational education (excluding college workers). Column (3) includes workers with a high school or vocational education who were born in the less urban provinces of Spain. Standard errors clustered at the level of region of birth \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

equal to $-.0099$, is very similar to that found in [Oreopoulos et al. \(2012\)](#).⁷⁶ 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.⁷⁷ As discussed in Section 4.6 and in Appendix C, 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 $.117$, see Table 2), and (ii) results from [Fernández-Kranz and Rodríguez-Planas \(2018\)](#), who estimate the first-market-tightness effect in the Spanish context. They find that entering during a typical recession (in Spain, a 8 ppt increase in the unemployment rate) is associated with cumulative 10-year earnings losses of 9.6%, 12.5%, and 6.4% for high school, vocational, and college workers respectively. This allows me to benchmark how much of the first-market-tightness effect can be explained by the first-employer size effect. The calculation is described and its results shown in Table 4.

The conclusion of this simple exercise is 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 places of Spain, this fraction is between 12% and 15%. This is a group of workers that are likely compliers in the IV estimates of Section 4.⁷⁸

⁷⁶While [Brunner and Kuhn \(2014\)](#) also provide evidence on the negative relationship between first-employer size and the unemployment rate, they do so using three firm-size categories (small, medium, large) rather than a continuous measure of employer size like $s_{J(i)}$.

⁷⁷I classify workers as rural- or urban-born based on their province of birth. I use data from [Goerlich Gisbert and Cantarino Martí \(2015\)](#) who estimate the fraction of urban (living in urban clusters, based on population and population density) and rural population (living outside these clusters) at the municipality level using data from 2006. I use their province aggregates and I classify as less urban 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 it is close to the median in my sample.

⁷⁸A caveat of this simple calculation is that while [Fernández-Kranz and Rodríguez-Planas \(2018\)](#) report cumulative losses for the first ten years since labor-market entry, my lifetime income measure (used to estimate the elasticity 0.117) aggregates, on average, 15 years of income. The effects of [Fernández-Kranz and Rodríguez-](#)

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

(1)	(2)	(3)	(4)	(5)	(6)
Sample	$\hat{\gamma}$	$\hat{\gamma} \times 8$ (8 ppt = typical recession)	% change in first-employer size	% change in lifetime income ($(4) \times .117$)	% 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 (5) 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 Section 4 (see 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). They 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 provide evidence on the persistence of this effect. This means it is not solely mediated by the time a worker spends at his first job and, thus, workers' trajectories in subsequent jobs are affected by the nature of their first employer. An inherent caveat in the persistence analysis is lack of exogenous variation in the employment dynamics following a first-employer match.⁷⁹ 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. In any case, using the rich employment dynamics present in the data provides the opportunity to carry out useful tests that can inform discerning between mechanisms.

The persistence of the first-employer size effect is consistent with a large literature modeling workers' endogenous decisions to move across jobs, and the ensuing "job ladder" that arises in labor markets. Indeed, Section 5.2 discusses and provides evidence for job search as one of the underlying mechanisms. Additionally, I argue that the high prevalence of involuntary separations among young Spaniards (which cut job ladder progressions) leaves scope for other persistent mechanisms. Namely, I explain how human capital accumulation is a likely additional channel and provide evidence consistent with it.⁸⁰

Planas (2018) have mostly dissipated after 10 years, while I find that a larger first-employer puts workers on a persistently higher wage growth path (see Section 5).

⁷⁹Ham 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 afterwards.

⁸⁰While discussing and finding evidence consistent with job search and human capital, other channels I cannot test for might also play a role. In particular, the social networks that are developed in a large firm could be larger or more valuable than in a small employer. These social ties could be beneficial for labor market outcomes throughout the working life (Cingano and Rosolia, 2012). This is an interesting potential channel

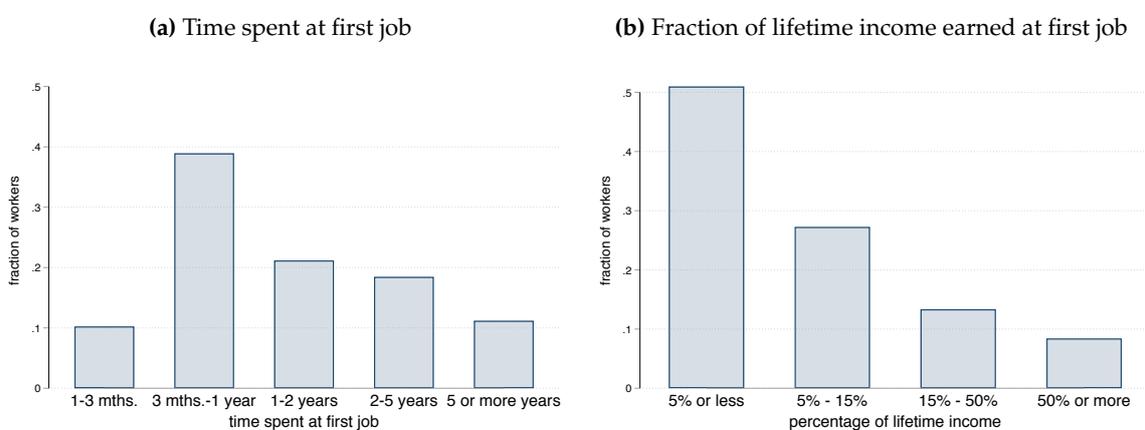
5.1 Persistence

If higher earnings at large firms arise solely due to rent-sharing, efficiency wages, or compensating differentials, we could expect the first-employer size effect to disappear as workers move on to other jobs since these explanations are linked to the employment relationship. Alternatively, there are reasons to expect persistent effects. Namely, human capital accumulation (heterogeneous skill development at heterogeneous employers) and search frictions (which can create a link between the quality of ensuing employers).

Descriptive evidence

Descriptive patterns in the data are consistent with persistence. The first thing to note is the high job mobility of young workers (Topel and Ward, 1992). Figure 5 shows that most workers do not stay at their first employer for a very 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. Figure 2 and Appendix Figure A7 show this for labor income, unemployment, and daily wages trajectories.

Figure 5: Time spent and income earned at the 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.

The combination of high mobility and earnings growth implies that a small fraction of lifetime income is directly earned at the first job. Figure 5 shows that this share is rather low for most workers.⁸¹ It represents 5% or less of lifetime income for half of the workers in my sample. 5–15% for 28% of workers, 15–50% for 12% of workers, and 50% or more for

outside the scope of this paper which future research could focus on.

⁸¹To compute this fraction I use the lifetime income measure used in Section 4 analysis, which is the 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).

less than 10% of the workers. Appendix Figure A14 shows that the patterns of Figure 5 are very similar when focusing on the subsample of workers I previously identified as “likely compliers” (non-college, and from less urban places).⁸²

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

I now provide more direct evidence on the persistence of the first-employer size effect. First, 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:

$$y_{iq} = (\beta_1 + \beta_2 \cdot q + \beta_3 \cdot q^2) \cdot s_{J(i)} + X'_{iq} \gamma + \varepsilon_{iq}. \quad (6)$$

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.⁸³ Table 5 shows the implied elasticities at different points in time (Appendix Table A5 shows the underlying β estimates). This is done allowing a quadratic trend as in equation (6), or imposing a linear trend (assuming $\beta_3 = 0$).

Table 5: Time-varying elasticity of income and first-employer size: Values at different points in time

Years after entry	Elasticity:	
	Linear trend	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 (6) in the text, shown in Appendix Table A5. Standard errors clustered at the level of region of birth \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

The β estimates in Appendix Table A5 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 employer results in higher earnings growth. Focusing on the linear trend, Table 5 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

⁸²Additional evidence on mobility and time spent at the first employer can be found in Appendix Figure A16.

⁸³In particular 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.

quadratic time trend delivers qualitatively similar results, although the implied elasticity twelve years after entry is somewhat larger, equal to 0.161.⁸⁴

These results, together with the high job mobility shown above, are evidence of a persistent effect of first employer size. Workers seem to reap returns from a larger first employer years after most of them have moved on to subsequent jobs.

Wage growth between the first and second job

While the above results strongly suggest persistent first-employer effects, one could still wonder if they are driven by the small fraction of people who stay with their first employer throughout this time period.⁸⁵ 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

$$g_i^{1,2} = \beta_1 s_{J_1(i)} + \beta_2 s_{J_2(i)} + \rho \ln(\bar{w}_{i1}) + f_1(\text{tenure}_i^1) + f_2(\text{tenure}_i^2) + g(\text{unemp}_i^{1,2}) + X_i' \gamma + \varepsilon_i. \quad (7)$$

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 j th employer, $\text{unemp}_i^{1,2}$ 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 (3) in addition to start of second job year dummies.⁸⁶

I estimate different specifications of equation (7). Table 6 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.

Income at age 35

Finally, it is informative to ask whether there is a detectable first-employer size effect on the income workers earn on the year they turn 35. This would be further evidence

⁸⁴Comparing these time-varying estimates with the results on lifetime income from Table 2 (elasticity of .117) is not straightforward since (i) the lifetime income measure is based on age and these estimates are based on potential experience, and (ii) mapping an average of these timing-varying effects into the lifetime elasticity should take into account the steep growth in wages that takes place during these years.

⁸⁵6.8% of the workers in the estimating sample remain in their first job until the year in which they reach age 35 (the final year in my panel).

⁸⁶While I believe 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 6: Between-job wage growth and first-employer size: OLS and IV-TSLS estimates

	OLS			First Stage			Second Stage		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
first employer size	-0.0038*** (0.0011)	-0.0056*** (0.0011)	-0.0007 (0.0010)				0.1072** (0.0496)	0.1021** (0.0480)	0.0938** (0.0425)
labor demand instr.				0.0824*** (0.0152)	0.0847*** (0.0155)	0.0856*** (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 \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

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 (from the U.S.) 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).⁸⁷ Thus, first-employer effects at age 35 (when the average person in the sample has been in the labor market for 15 years) would be suggestive of permanent effects past the actual years I consider in my lifetime income measure.

I estimate a version of equation (3) in which the outcome variable is income earned during the calendar year a worker turns 35. I do this using 88.3% of workers in my sample who work for at least half of the days in that year.⁸⁸ Table 7 shows the estimated elasticity, which is around 0.09.

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 then focus attention to drivers of life-cycle wage growth. Rubinstein and Weiss (2006) review such drivers and point to human capital accumulation and job search as two of the main ones.⁸⁹

⁸⁷I provide related evidence for Spain in Appendix Figure A5.

⁸⁸I 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.

⁸⁹Learning about a worker's unobserved abilities, or about job match quality is the third driver reviewed in Rubinstein and Weiss (2006) (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). For a paper trying to tease out job-specific human capital vs. learning about match quality see Nagypál (2007).

Table 7: Income during age 35 and first-employer size: OLS and IV-TSLS estimates

	OLS	First Stage	Second Stage
	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 \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Search

Search models show how on-the-job search and search frictions jointly result in a job ladder, where workers move across jobs (“up” the job ladder) only if a new offer is more desirable than the current job (e.g. [Burdett and Mortensen, 1998](#); [Christensen et al., 2005](#); [Lise, 2013](#); [Jarosch, 2015](#); [Krolikowski, 2017](#), among others). How could the first-employer size effect be mediated by search frictions? Existing evidence suggests that larger firms tend to be more desirable ([Haltiwanger et al., 2018](#); [Sorkin, 2018](#)).⁹⁰ A first job at a large firm could give a worker a “higher” starting point in the ladder, with him moving to more desirable subsequent jobs than if he had started at a smaller, potentially less desirable firm.

Section B.8 in Appendix B lays out a simple framework that complements this discussion. It illustrates the job search channel; first by itself and then adding a human capital channel too.

Human capital

Human capital acquired on the job, either through learning-by-doing or through formal training, is an important driver of wage growth.⁹¹ There are several reasons why workers could acquire more skills, and/or of higher value at large firms (see Section 7). Additionally, it could be particularly productive to acquire these skills early in the working life, when

⁹⁰[Haltiwanger et al. \(2018\)](#) argue that, conditioning on firm age, there is empirical evidence in the U.S. for a job ladder in terms of firm size. [Sorkin \(2018\)](#) documents that, when taking into account both pay and non-pay amenities, there is a positive correlation between a firm’s desirability and its size.

⁹¹The literature offers several reasons why would employers invest in training for their workers that will be valuable in other firms. While maintaining the traditional dichotomy between general and specific human capital, [Acemoglu and Pischke \(1999\)](#) argue 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. [Rosen \(1972\)](#) develops a model of the job market where on-the-job learning of general skills is heterogeneous across jobs.

workers are in a formative period.⁹² The skill formation literature emphasizes the complementarity in skills acquired at different points in time; the notion that “skill begets skill”, and that skills developed at one stage of the life cycle impact the productivity of further skill development in the next stage (see Cunha et al., 2006).⁹³ All this (together with the discussion in Section 7) suggests that heterogeneous skill acquisition at larger or smaller first employers could significantly impact young workers’ human capital development in persistent ways, paying off throughout the working life.⁹⁴

The framework of Section B.8 in Appendix B complements this discussion. I add to the framework on-the-job human capital development that is heterogeneous across employers and, I discuss ways in which the quality of early employers could be key in persistent ways.

Empirical evidence

Search frictions and a job ladder. I find evidence consistent with a job ladder channel. Using the IV framework from Section 4, I find that starting at a larger employer leads to larger subsequent employers. I estimate equation (3) using (log) size of a worker’s second employer and (log) size of his employer at age 35 as dependent variable. Results are in Table 8. 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 induced by search frictions (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 in such environments.⁹⁵

Human capital. Compared to search frictions, heterogeneous skill-development opportunities across employers leading to persistent consequences is a more novel mechanism.⁹⁶ It is important to assess this channel since it could carry interesting implications. We might ask ourselves whether firms offering better human-capital development opportunities internalize this in their decisions. Further, the efficiency implications some argue arise from size-dependent policies and regulations (IMF, 2015; Guner et al., 2008; Garicano et al., 2016) could be seen under a new light if large firms are generally better at making their young

⁹²Gibbons and Waldman (2006) develop a model where task-specific skill development generates persistent effects based on the mix of skills developed at a worker’s first job.

⁹³While this literature empirically focuses on childhood, their model can be applied to skill development of young adults. Further to the point of the importance and productivity of skill acquisition during the *early* stages of the working life is evidence indicating that cognitive abilities start declining when people reach their 20s (see Salthouse, 2009, and literature cited therein).

⁹⁴In 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.

⁹⁵Reduced job mobility resulting from a larger first-employer would also be consistent with job search models and a job ladder channel. However, I have estimated equation (3) using as outcome variable different measures of job mobility (e.g. total number of employers, or tenure at the first job) and the results are small and noisily estimated (results available upon request). The reason why this theoretical prediction does not seem to be borne out in the data could have to do with the existence of heterogeneous on-the-job search processes at different firms. This would balance two opposing forces: larger employers being more desirable (reducing mobility), and larger employers offering better on-the-job search opportunities (increasing mobility). In contrast, the canonical on-the-job search model assumes a common arrival rate of new offers across all jobs.

⁹⁶Different theoretical models going back to Rosen (1972) lay out how on-the-job learning can vary across jobs and its implications.

Table 8: Subsequent employers and first-employer size: OLS and IV-TSLS estimates

	OLS	First Stage	Second Stage	OLS	First Stage	Second Stage
	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 \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

workers persistently more productive.

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 (see the framework of Appendix B, Section B.8), 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.

The high frequency of the data allows me to have a good sense of who are the workers that experience unemployment between jobs. I can then assess heterogeneity in the first-employer size effect along this dimension. I categorize workers based on whether they experience an unemployment spell between their first and second job or not.⁹⁷ 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 no or 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 first and second stage IV results of equation (3) can be found in Table 9.⁹⁸ 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 .117.⁹⁹ Elasticities of comparable magnitudes to baseline also arise for

⁹⁷I categorize as having an unemployment spell workers who are not employed for at least two full months between the two jobs. This is to ensure I do not mis-categorize job-to-job transitions with a small break in between.

⁹⁸OLS results can be found in Appendix Table A6.

⁹⁹That 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 confidence intervals around these estimates are, however, somewhat wide.

Table 9: Career outcomes and first-employer size: 1st–2nd job unemployment gap sample

	First Stage	Second Stage						
	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 \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

average daily wages, lifetime earnings, subsequent employers' size, or income at age 35. The latter is noisily estimated but similar to baseline.¹⁰⁰ 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.¹⁰¹

Finally, some descriptive facts shed more light on the plausibility of a human capital channel in addition to a job search explanation. The first-to-second job transition is not the only time when young workers might experience unemployment, losing their search capital. The average worker has over 7 employers until age 35 (see Table 1), and many job transitions are induced by involuntary separations (youth unemployment in Spain is notoriously high). Appendix Figure A15 shows that the occasions in which young workers see their job ladder progressions severed abound. It shows the fraction of people who experience at least one unemployment spell between any two jobs. A majority of workers experience some unemployment, especially those holding three or more jobs until age 35.

So far I have argued that, after establishing persistence of the first-employer effect, search and human capital are the candidate channels. 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 test the notion that skills acquired at large employers are more valuable by studying the differential returns to experience from large vs. small employers.

6 Differential Returns to Large-Employer Experience

In this section I provide an additional test for the human capital mechanisms. If on-the-job skills developed at large employers are more valuable we would expect that experience acquired at large firms is more valuable than experience acquired elsewhere. I test for a differential return to large-firm experience in terms of wages and promotions.

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 allows me to precisely quantify actual experience at different employers measured in days. Experience at large employers could be correlated with unobserved worker characteristics and attributes of the current employer that influence wages.¹⁰² 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, type of contract). The remaining variation that I

¹⁰⁰I do not find an economically or statistically significant elasticity for total days worked for this group (who all experience unemployment after their first job).

¹⁰¹Similarly, 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 A16).

¹⁰²For the evolution of the literature estimating the returns to general experience and tenure (seniority) see Altonji and Shakotko (1987), Abraham and Farber (1987), Topel (1991), Altonji and Williams (2005), and Buchinsky et al. (2010). In relation to firm size, Fackler et al. (2015) document that wage growth of firm stayers is positively correlated with firm size.

use is to compare workers who work for observably similar employers and have the same amount of experience, but acquired this experience in different, large or small, firms. Later, I address plausible remaining concerns such as endogeneity arising from a match-specific component, or the interpretation of returns to big-firm experience as returns to skill.

6.1 Wages

I estimate the following monthly wage equation:

$$\ln w_{it} = \alpha_i + \psi_{s(i,t)} + \gamma_1 \text{bigExp}_{it} + \gamma_2 (\text{bigExp}_{it} \cdot \text{Exp}_{it}) + X'_{it} \delta + \varepsilon_{it}. \quad (8)$$

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).¹⁰³ X_{it} includes time-varying controls: a quadratic term for total experience, tenure at current employer (quadratic), age (quadratic), regional unemployment level (quadratic), type of labor contract (permanent or fixed-term), sector fixed effects, and time (year-month) fixed effects.¹⁰⁴

The parameters γ_1 and γ_2 capture the differential value of experience at large firms and how it varies over the working life. Let $\text{Exp}_{it} = \text{bigExp}_{it} + \text{smallExp}_{it}$ and Z_{it} be equation (8) regressors, then

$$\frac{\partial \mathbf{E}(\ln w_{it} | Z_{it})}{\partial \text{bigExp}_{it}} - \frac{\partial \mathbf{E}(\ln w_{it} | Z_{it})}{\partial \text{smallExp}_{it}} = \gamma_1 + \gamma_2 \text{Exp}_{it}. \quad (9)$$

Note that worker fixed effects, α_i , and current employer size category indicators, $\psi_{s(i,t)}$, imply that γ_1 and γ_2 are identified through workers' movements across employers of different sizes. The young age of the workers in this panel (16–35) and the high mobility across employers during these years enable this. Observing monthly data and being able to quantify actual experience in days provides additional useful variation.

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 fixed effects) imply that γ_1 and γ_2 are estimated by 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

¹⁰³As discussed in footnote 33, employer size is not observed prior to 2004 except for the firms for which I obtained a special extract (those which are first or second employers of the workers in the sample). 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 6 categories: missing size, 1–5 employees, 6–19, 20–49, 50–249, and 250+.

¹⁰⁴This type of specification is similar to that used in De La Roca and Puga (2017) to study worker learning in cities. To the extent that larger employers are located in larger cities my results will be related to theirs. The equation can also be seen as a version of the AKM wage model (Abowd et al., 1999) where firm effects are aggregated into firm-size category effects.

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.¹⁰⁵

Columns (1) and (2) of Appendix Table A7 show estimates of equation (8). 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 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 6 helps understand the magnitude implied by the coefficients and its evolution over time. It plots in percentage terms the differential return to one year of large-employer experience (specification including $\psi_{s(i,t)}$).¹⁰⁶ One year of large-employer experience results in between 3% and 2% higher monthly wages than a year of experience elsewhere. 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.¹⁰⁷

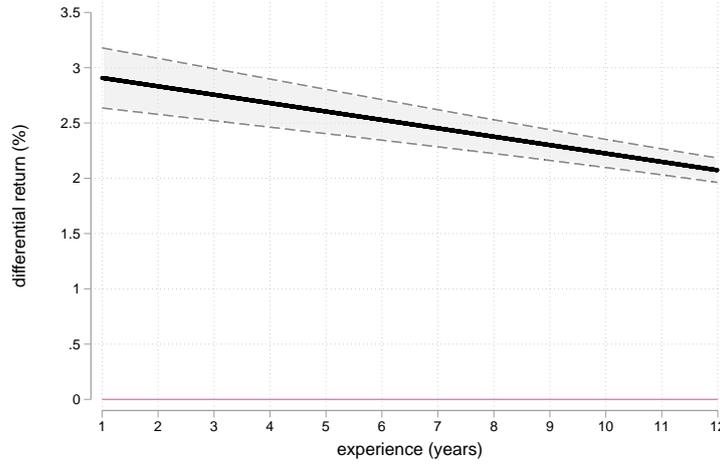
Next, I explicitly test whether large-employer experience is differentially more valuable across all *current* employer sizes. One possibility is that the differential skills that workers learn at large firms are only valuable at other large firms. Alternatively, these skills could be more general, and useful in many types of future jobs and firms. Understanding which is the case is informative towards grasping the level of portability of skills developed at large firms. Portability will, in turn, intuitively impact how valuable are these skills. Appendix Table A8 shows estimates of γ_1 and γ_2 from equation (8) which are allowed to vary by current-employer size category (interacted with $\psi_{s(i,t)}$). The table shows that the same qualitative pattern remains across all employer sizes with large, positive estimates of γ_1 and small estimates of γ_2 . Thus, it seems that experience from large employers is differentially more valuable across the firm-size distribution, supporting the idea that valuable skills acquired at large employers are portable across employer types.

¹⁰⁵ Abraham and Farber (1987) make a similar point about the distinction of returns to experience per se and the returns to job search. Haltiwanger et al. (2018) show that, when focusing on mature firms, there is empirical evidence in the U.S. for a job ladder in terms of firm size. On some level the fact that I focus on the return to big-firm experience *over and above* returns to experience at smaller firms reduces the scope for a search returns interpretation. A higher offer arrival rate or higher search intensity for workers in large firms would be required for the differential returns to arise from search.

¹⁰⁶ Concretely, it plots $365 \cdot 100(\hat{\gamma}_1 + \hat{\gamma}_2 Exp)$ for different levels of *Exp*, together with 95% confidence intervals.

¹⁰⁷ Figure 6 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 (8) ends.

Figure 6: 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 (8) (Table A7, 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.9 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 are reassuring and support the above interpretation of the findings.

6.2 Promotions

Having found a differential wage premium for large-employer experience, 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.¹⁰⁸

Social security data include information on professional categories, which I use to construct a proxy for promotions. Section B.10 in Appendix B describes how I construct this variable and provides evidence supporting its interpretation as promotions. Using this variable I estimate linear probability promotion (hazard) regressions of the following type:

$$Prom_{it} = \alpha_i + \psi_{s(i,t)} + \phi_{p(i,t-1)} + \lambda_1 bigExp_{it} + \lambda_2 (bigExp_{it} \cdot Exp_{it}) + X'_{it} \delta + \varepsilon_{it}. \quad (10)$$

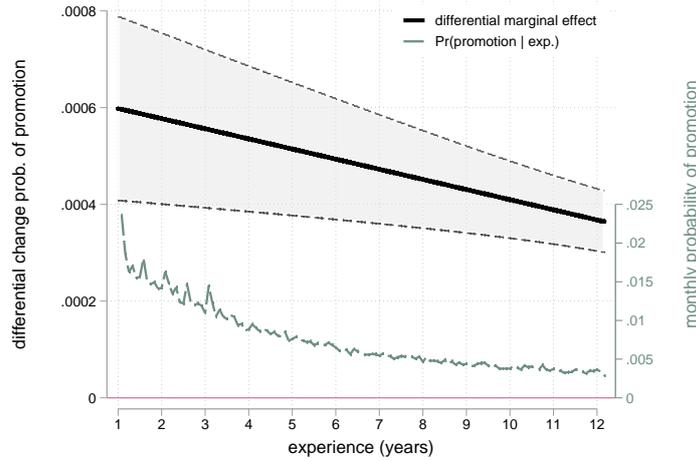
¹⁰⁸This would be consistent with model predictions in Gibbons and Waldman (2006), where sufficient time spent in a low-level job decreases to zero the probability of promotion.

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 (8), 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 (10) regressors, then

$$\frac{\partial Pr(Prom_{it} = 1|Z_{it})}{\partial bigExp_{it}} - \frac{\partial Pr(Prom_{it} = 1|Z_{it})}{\partial smallExp_{it}} = \lambda_1 + \lambda_2 Exp_{it}. \quad (11)$$

Figure 7: 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 (10) (Appendix Table A9, 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.

Columns (1) and (2) of Appendix Table A9 show estimates from equation (10). 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 7 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.¹⁰⁹ 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 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.

7 Distinctive Large-Employer Attributes and Skill Accumulation

This section 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. Whenever I can, I provide descriptive evidence between firm size and these attributes in the context of Spain.

7.1 Formal Training and Education

The literature has provided evidence that 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 arguably general skills like computing and basic education.¹¹⁰

Appendix Table A11 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 employed in 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%).¹¹¹

¹⁰⁹In particular, it plots $365 \cdot (\hat{\lambda}_1 + \hat{\lambda}_2 Exp)$.

¹¹⁰The 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.

¹¹¹Formal 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.

7.2 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 processing" 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 ([Antonacopoulou 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.¹¹² 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 seem to be 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 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., 2018](#); [Nix, 2017](#); [Jarosch et al., 2018](#)). [Bloom and Van Reenen \(2006\)](#) show that larger firms tend to be better managed. The correlation between size and management quality is also present for Spanish employers (see Appendix Figure A8).

7.3 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](#)).¹¹³ 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

¹¹²This relates to the literature on learning about heterogeneous worker abilities as a source of wage growth. See [Rubinstein and Weiss \(2006\)](#) for an overview.

¹¹³The 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.

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 well-known 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. I show in Appendix Table A12 that during the 1990s and early 2000s, larger employers in Spain were more likely to invest on R&D and foreign technology transfers.

8 Conclusion

This paper sheds new light on ways in which firm heterogeneity can affect workers' labor market outcomes. For a given worker, the potential career trajectories that arise from starting out at one employer or another can be substantially different.

I have focused attention on firm size. Size is an employer attribute which is transparently measured and observable to labor market participants and policymakers. Further, size correlates with meaningful employer characteristics: productivity, workforce training, higher wages, or managerial quality among others. The findings from this paper imply that, indeed, size summarizes relevant firm attributes for young workers.

I document a causal link between starting the working life at a larger firm and enjoying better long-term labor market outcomes. I am able to identify a causal effect because, in spite of the importance of a first-job match, there is some randomness involved in the process of creating this match. My instrumental-variables approach leverages the part of this randomness that is driven by idiosyncratic hiring shocks of large firms at the time of young workers' labor market entry.

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 (LATE). I find that these workers are younger and with lower earnings potential. While not representative of the whole population, this group of people might be of special interest for researchers or policymakers. My findings suggest that these workers, either due to skill-development or worse outside options, seem to disproportionately benefit from starting out at a large first employer.

The second part of the paper shows evidence consistent with two mechanisms. The first channel is a "job ladder" one. I show that starting at a larger first employer leads to holding subsequent jobs also at larger firms. A simple job search model in which large-firm workers are more selective in accepting subsequent offers could give rise to this channel. The second channel is a human capital one. My results are consistent with young people

facing better on-the-job skill development opportunities when employed at large firms. It is worth noting that the two channels naturally complement and reinforce each other.

Human capital as a driving mechanism can have interesting implications. For instance, it could carry efficiency considerations. In the presence of worker mobility and imperfect wage adjustment, firms that are better at increasing young workers' productivity in persistent ways might not fully internalize this in their operational decisions. Further, researchers and policy-makers worry about the aggregate losses that can arise from an inefficient firm-size distribution induced by size-specific policies and regulations (IMF, 2015; Guner et al., 2008). These losses could be underestimated if large employers provide better skill-development opportunities to their workers.

Finally, a better understanding of what it is that "good human-capital" firms do well could be informative for worker training and active labor market policies. Policy could also be used to encourage these 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.

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., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica* 67(2), 251–333.
- 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.
- 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 (2018). A distributional framework for matched employer employee data. Mimeo, University of Chicago and New York University.
- 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., J. Robert E. Lucas, and E. Rossi-Hansberg (2018). Learning, career paths, and the distribution of wages. *forthcoming in American Economic Journal: Macroeconomics*.
- 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 (2000). Adapting to circumstances: The evolution of work, school, and living arrangements among North American youth. In D. G. Blanchflower and R. B. Freeman (Eds.), *Youth employment and joblessness in advanced countries*, pp. 171–214. 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., C. Lelarge, and J. Van Reenen (2016). Firm size distortions and the productivity distribution: Evidence from France. *American Economic Review* 106(11), 3439–79.
- 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.
- 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.
- 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.
- 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. Meyerson 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.
- Lise, J. (2013). On-the-job search and precautionary savings. *The Review of Economic Studies* 80(3), 1086–1113.
- 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.
- McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.
- 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.
- Nagypál, É. (2007). Learning by doing vs. learning about match quality: Can we tell them apart? *The Review of Economic Studies* 74(2), 537–566.
- 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.
- Nybom, M. (2017). The distribution of lifetime earnings returns to college. *Journal of Labor Economics* 35(4), 903–952.
- Oi, W. Y. and T. L. Idson (1999). Firm size and wages. *Handbook of Labor Economics* 3, 2165–2214.
- Oreopoulos, P. and U. Petronijevic (2013). Making college worth it: A review of the returns to higher education. *The Future of Children*, 41–65.
- 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.
- 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 (2018). Unlucky cohorts: Estimating the long-term effects of entering the labor market in a recession in large cross-sectional data sets. Working Paper 25141, National Bureau of Economic Research.
- 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 (2015). Firming up inequality. Working Paper 21199, National Bureau of Economic Research.

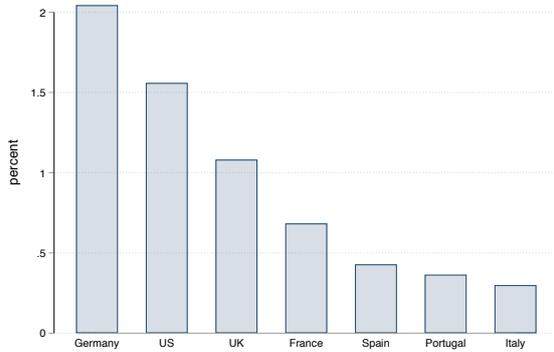
- Sorkin, I. (2018). Ranking firms using revealed preference. *The Quarterly Journal of Economics* 133(3), 1331–1393.
- Syverson, C. (2011). What determines productivity? *Journal of Economic Literature* 49(2), 326–65.
- Topel, R. (1991). Specific capital, mobility, and wages: Wages rise with job seniority. *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 -

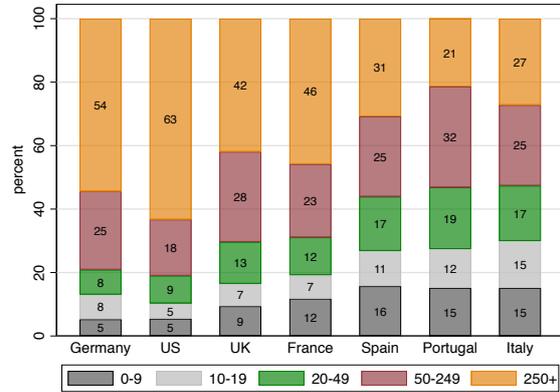
A Additional Figures and Tables

Figure A1: Relatively few large firms in Spain: Firm size across countries

(a) Percentage of enterprises that are large (250+)



(b) Employees by business size (manufacturing)



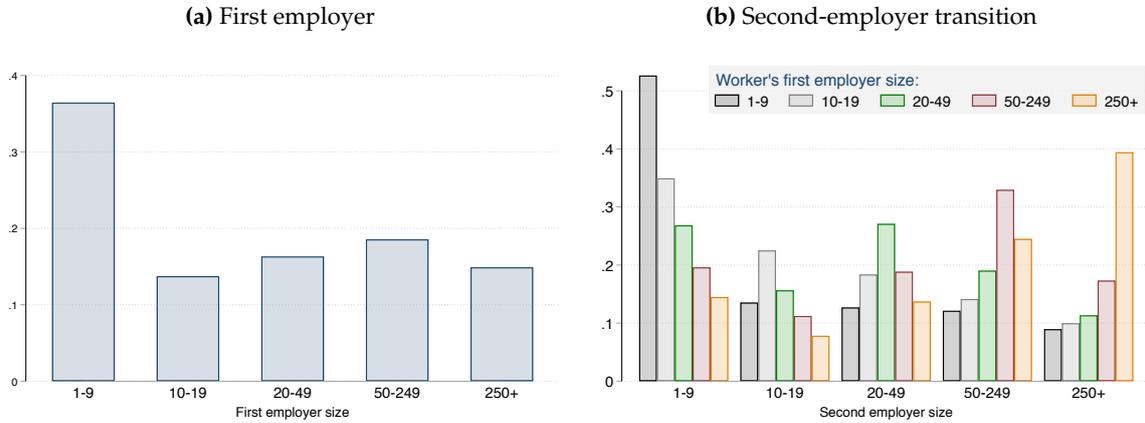
Notes: Source is OECD. Data refer to year 2013. Panel (a): Percentage of total number of enterprises (excludes self-employed) in each country that have 250 employees or more. Panel (b): Percentage of manufacturing workers working in each employer size category. Categories might not add up to 100 due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

Figure A2: Spanish regions



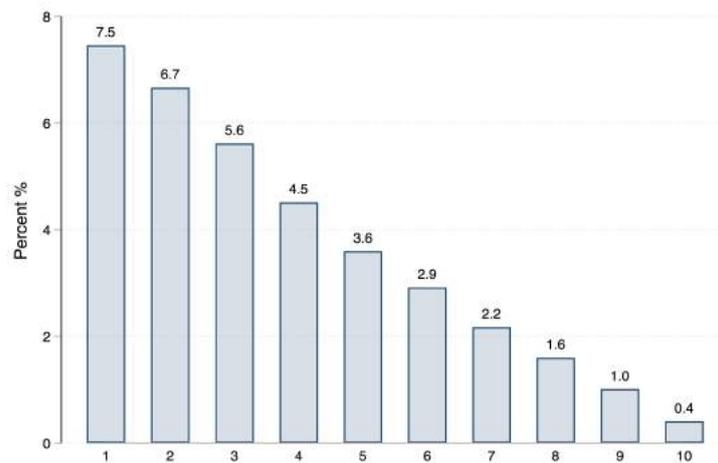
Notes: Regional division of Spain into 17 regions plus two North Africa enclaves. Each region is further divided into provinces. There are a total of 50 provinces.

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



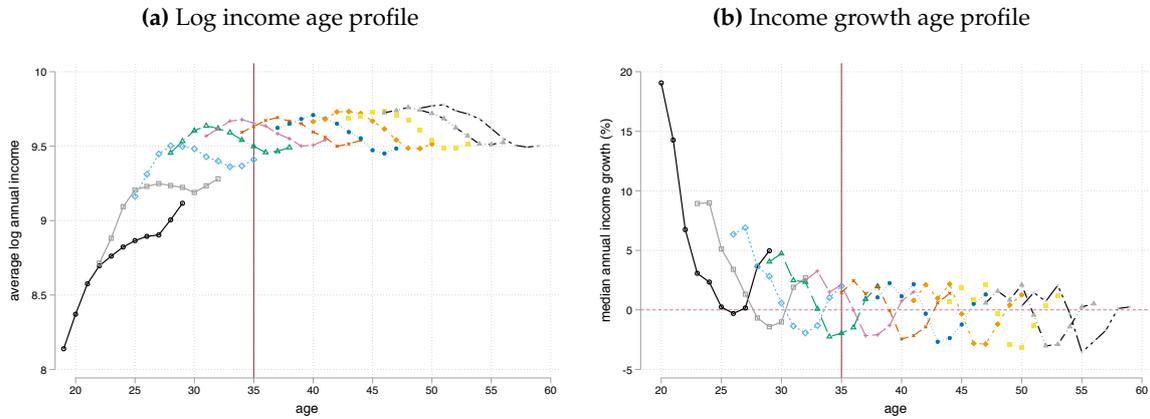
Notes: Panel (a): Distribution of first-employer size. Panel (b): Distribution of second-employer size, separately by first-employer size groups. Sample of male workers of all education levels, born in Spain between 1968–1980.

Figure A4: % of lifetime income coming from unemployment benefits: by decile of lifetime income



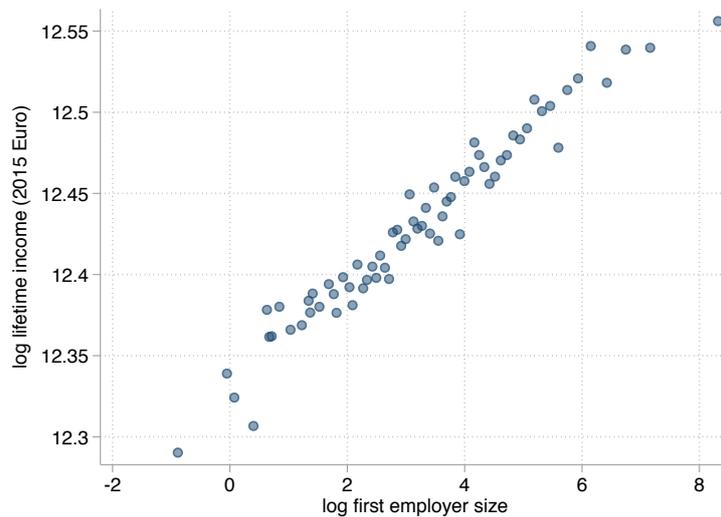
Notes: Percent of lifetime income that comes from unemployment benefits as opposed to wages. Separately by decile of lifetime income. Lifetime income in 2015 Euro as defined in equation (1) in the text. Sample of male workers of all education levels, born in Spain between 1968–1980.

Figure A5: Income stabilizes by age 35: Annual income and growth age profiles (2006–2015)



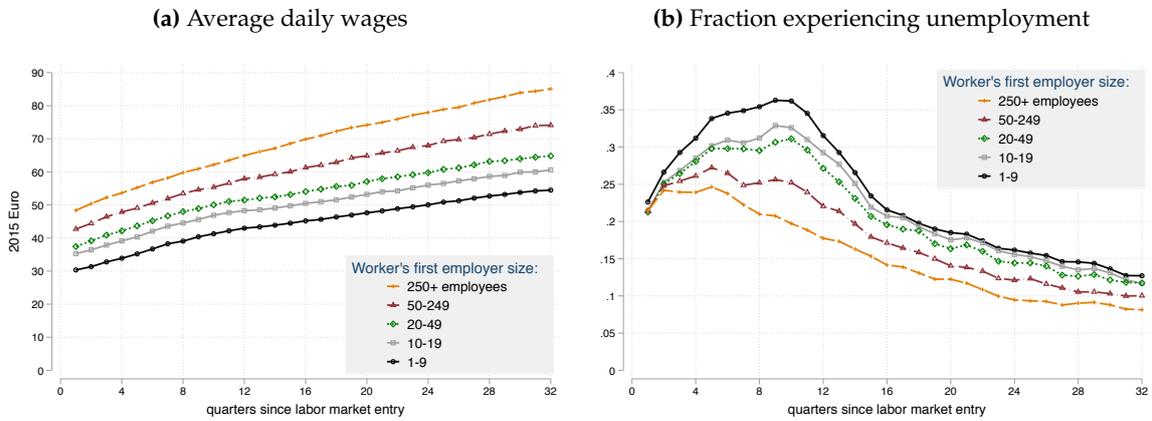
Notes: Age profiles for different cohorts in log annual income and annual income growth rates. Top panel: average log annual income by cohort and year. Bottom panel: median annual income growth rate by cohort and year. Growth rate g_t between annual income Y_{t-1} and Y_t computed as $100 \times \frac{Y_t - Y_{t-1}}{\frac{1}{2}(Y_t + Y_{t-1})}$. 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 A6: Lifetime income and first-employer size correlation: controlling for sector



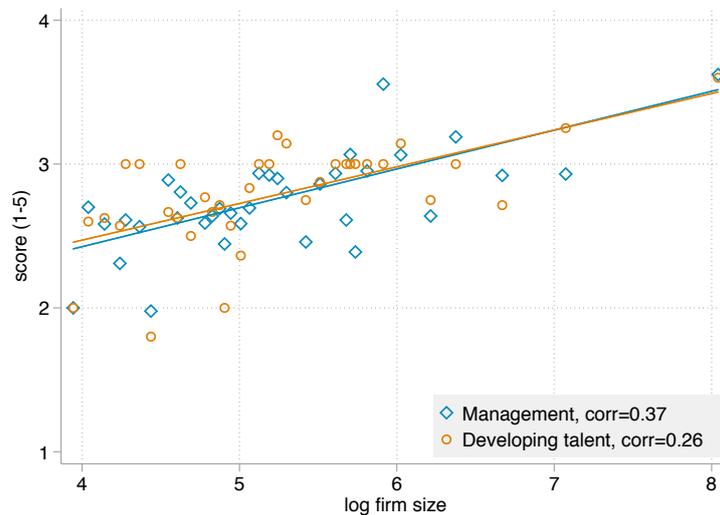
Notes: Binned scatterplot. Log lifetime income (as defined in the text) on the vertical axis. Log size of worker's first employer on the horizontal axis. Both variables net out of 58 first-employer sector fixed effects. Sample of male workers of all education levels, born in Spain between 1968–1980.

Figure A7: Daily wages and unemployment trajectories by first-employer size category



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.

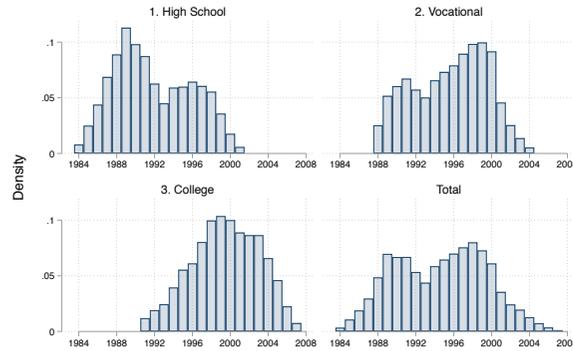
Figure A8: Firm size and managerial quality in Spain



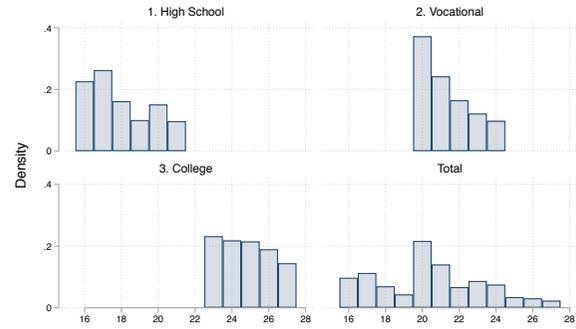
Notes: Source is World Management Survey, 2013 wave. Data on 214 manufacturing plants in Spain. Size refers to firm (not plant) size. Management is the average score of all survey management questions. Developing talent is the score of a single question (which is also included in the overall Management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

Figure A9: Additional first labor market semester summary statistics, by educational attainment

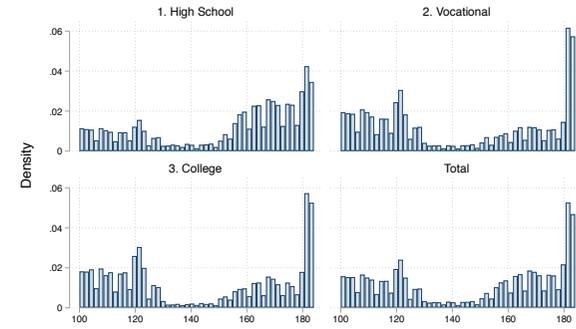
(a) Year of first labor market experience



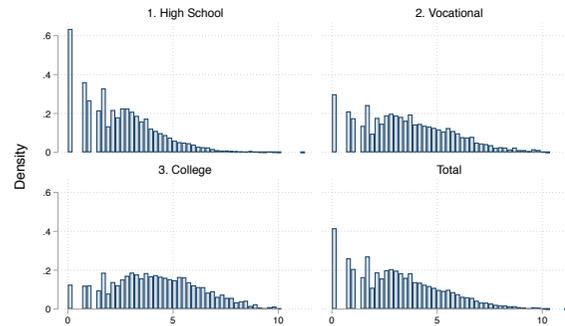
(b) Age at first labor market experience



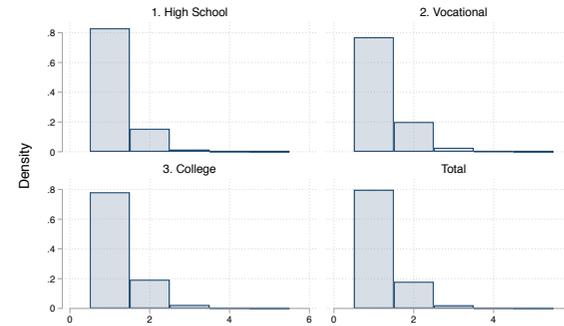
(c) Days worked during first labor market experience



(d) First employer log size

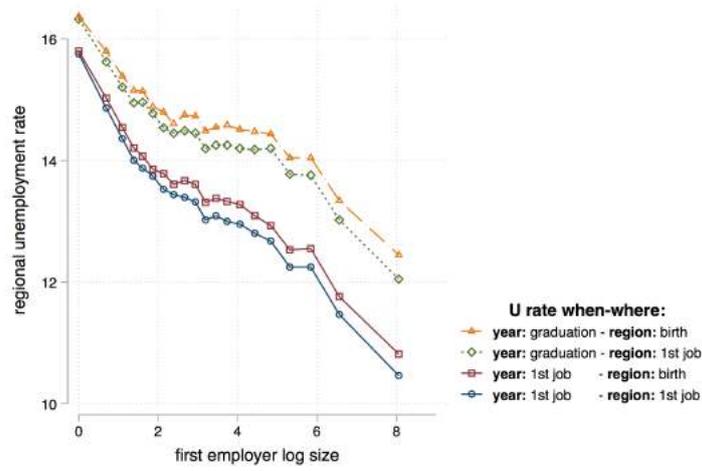


(e) Number of employers during first labor market experience



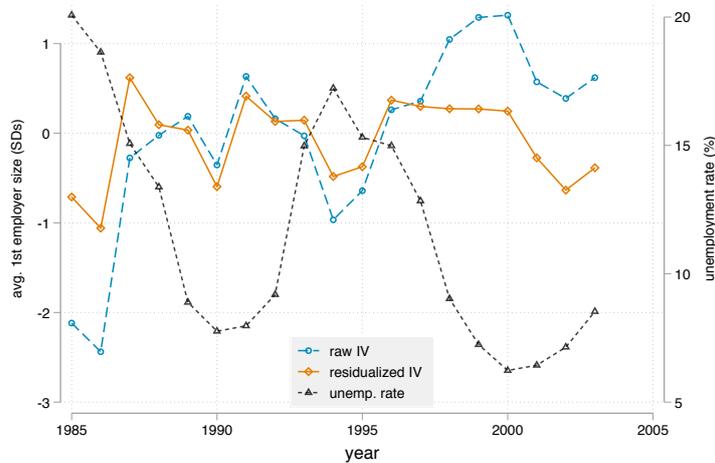
Notes: First labor market experience defined as the first six continuous months after predicted graduation (defined in text) that a person works for 100 days or more. All figures plotted separately by education level and overall. Panel (a): Calendar year of first labor market experience. Panel (b): Age 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 A10: First-employer size and business cycle variation



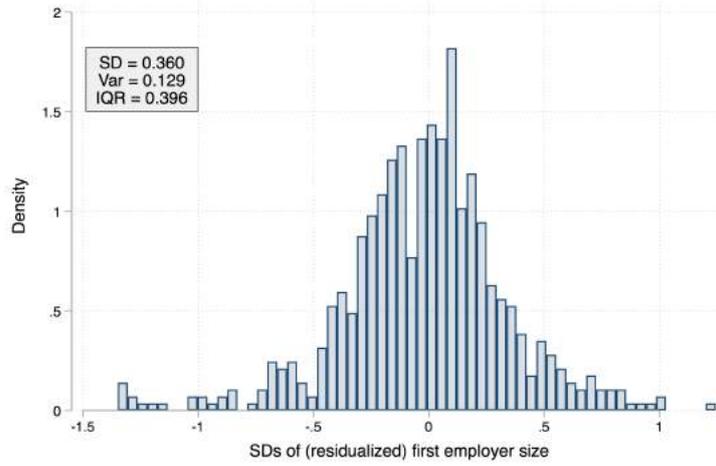
Notes: Binned scatterplot at the worker level. X-axis is log first-employer size of a given worker. Y-axis is the regional unemployment rate in different years/regions associated with the worker. Different colors represent the unemployment rate at different combinations of region (of birth or of first employment) and time (of predicted graduation or of first employment).

Figure A11: IV residual variation and business cycle variation in Catalunya region



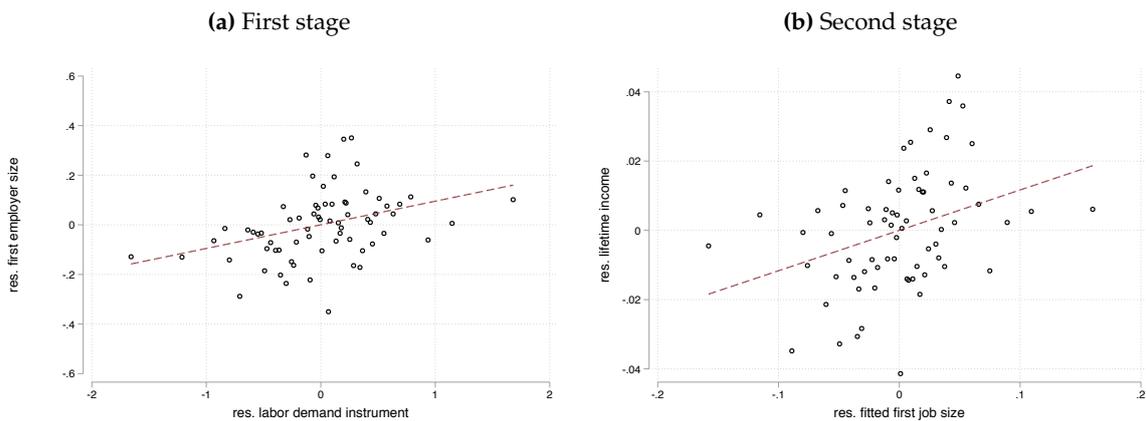
Notes: Time series evolution of the unemployment rate in Catalunya (black triangles), the instrument \bar{s}_{-i}^{rec} described in the text (blue dots), and residuals from a regression of \bar{s}_{-i}^{rec} on region of birth, education, and cohort fixed effects and a flexible of the regional unemployment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

Figure A12: 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 A13: IV-TSLS elasticity of lifetime income w.r.t. first-employer size



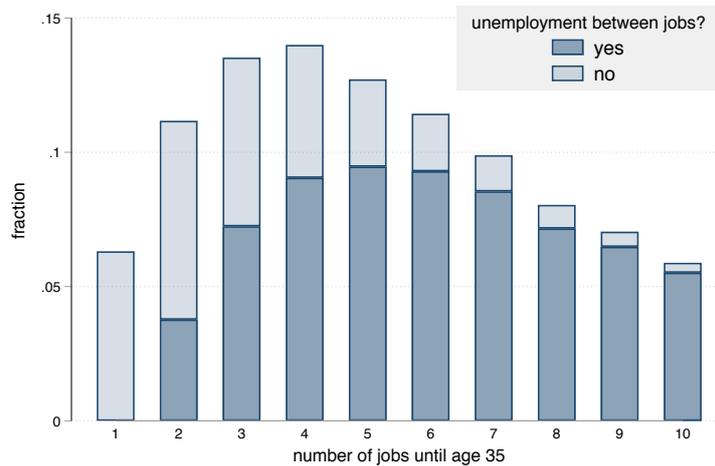
Notes: Binned scatterplots of first stage and second stage residual variation from equation (3) in the text, instrumenting for log first job size $s_{J(i)}$ using the instrument \bar{s}_{-i}^{ec} described in the text. The outcome variable is log total income after first job semester (described in text) up until age 35. Sample of male workers of all education levels, born in Spain between 1968–1980.

Figure A14: Time spent and income earned at the first job: Subsample of “likely compliers”



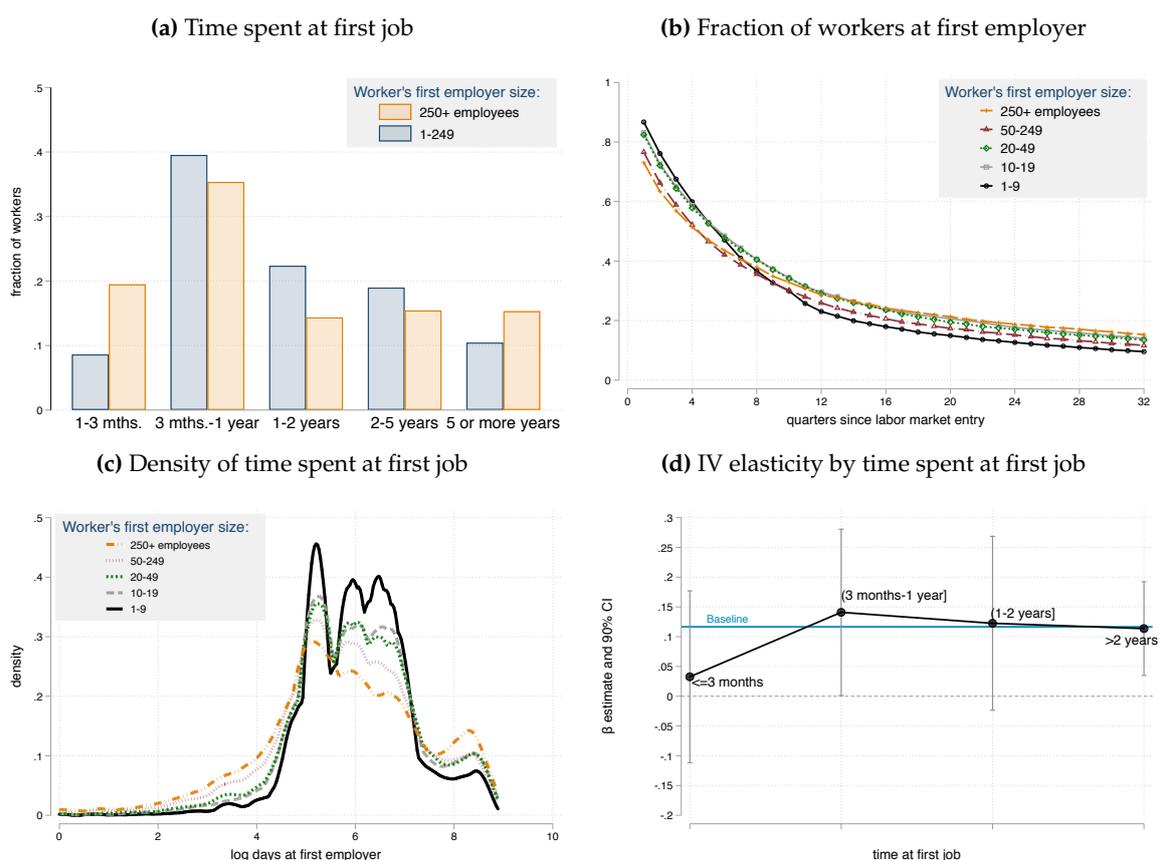
Notes: Panel (a): Distribution of time spent at first job. Panel (b): Distribution of the fraction of lifetime income earned at the first job. Lifetime income defined in text. First job is that held during the first continuous six months after predicted graduation in which a person works for 100 days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

Figure A15: Unemployment between jobs: By number of jobs until age 35



Notes: Total bar height indicates the fraction of workers who hold a given number of jobs between their first job and age 35. Within each bar, the shaded region indicates the proportion of workers who have experienced unemployment between at least one pair of jobs. I count as unemployment gaps those that are at least 2 months long. Restricted to workers in the lifetime analysis sample that hold 10 jobs or less between their first job and age 35 (75% of original sample). By construction, those who only hold one job experience no unemployment gap.

Figure A16: Time spent at first job by employer size, and IV results 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. Sample of male workers of all education levels, born in Spain between 1968–1980. 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 (3)), estimated separately for four groups of workers based on the time spent at the first employer. Group ≤ 3 months: $N=7,455$. Group (3 months–1 year]: $N=29,405$. Group (1–2 years]: $N=18,138$. Group > 2 years: $N=24,943$.

Table A1: IV residual variation uncorrelated with the business cycle

	Dependent variable = \bar{s}^{rec}				
	(1)	(2)	(3)	(4)	(5)
unemployment rate	-0.0828*** (0.0076)	-0.0523*** (0.0081)			
GDP growth			0.0707** (0.0300)	0.0263* (0.0151)	0.0055 (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}_i^{rec} (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 \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

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

	OLS				First Stage	Second Stage			
	lifetime inc. 0% (1)	lifetime inc. 1% (2)	lifetime inc. 2% (3)	lifetime inc. 3% (4)	first job size (5)	lifetime inc. 0% (6)	lifetime inc. 1% (7)	lifetime inc. 2% (8)	lifetime inc. 3% (9)
first employer size	0.0276*** (0.0011)	0.0274*** (0.0011)	0.0271*** (0.0011)	0.0269*** (0.0011)		0.1166** (0.0481)	0.1195** (0.0486)	0.1225** (0.0492)	0.1255** (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 TSLS estimates, by varying annual discount rate. Standard errors clustered at the level of region of birth \times education \times 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		Second Stage	
	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 \times education \times 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 (3) in the text. Second row refers to the estimation of an equation similar to (3) 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: Quarterly income and time-varying first-employer size elasticity

	OLS		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 \times 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 (6) in the text. Regressions at the worker \times 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 \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Table A6: 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 \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Table A7: Differential returns to experience at large employers: Monthly wages

	(1)	(2)	(3)	(4)
<i>bigExp</i>	115.2666*** (4.1355)	81.7438*** (4.0561)	115.7765*** (4.1367)	80.7595*** (4.0755)
<i>bigExp</i> · <i>Exp</i>	-0.0117*** (0.0008)	-0.0057*** (0.0008)	-0.0070*** (0.0010)	0.0005 (0.0010)
<i>bigExp</i> · <i>Tenure</i>			-0.0091*** (0.0008)	-0.0115*** (0.0007)
<i>Exp</i>	193.8450*** (3.5338)	200.7712*** (3.4959)	192.5913*** (3.5330)	199.3433*** (3.4948)
<i>Exp</i> ²	-0.0283*** (0.0004)	-0.0297*** (0.0004)	-0.0285*** (0.0004)	-0.0300*** (0.0004)
<i>Tenure</i>	130.0217*** (1.7206)	117.7391*** (1.7191)	131.7443*** (1.7173)	119.9342*** (1.7124)
<i>Tenure</i> ²	-0.0219*** (0.0004)	-0.0196*** (0.0004)	-0.0210*** (0.0004)	-0.0184*** (0.0004)
Current employer size category FE	no	yes	no	yes
Clusters (workers)	125232	125232	125232	125232
N (worker \times month)	16198288	16198288	16198288	16198288

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-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.

Table A8: Differential wage returns to experience at large employers, by current employer size

	(1)	
<i>bigExp</i>		
missing	68.5654***	(26.1499)
1-5	69.8272***	(14.1053)
6-19	99.1858***	(10.8146)
20-49	92.5799***	(11.0074)
50-249	100.8000***	(8.9762)
250+	45.0086***	(7.8794)
<i>bigExp</i> · <i>Exp</i>		
missing	0.0038	(0.0089)
1-5	-0.0071**	(0.0034)
6-19	-0.0101***	(0.0025)
20-49	-0.0082***	(0.0025)
50-249	-0.0086***	(0.0019)
250+	-0.0003	(0.0016)
Current employer size category FE	yes	
Clusters (workers)	125232	
N (worker × month)	16198288	

Notes: Dependent variable is log monthly wage. Experience measured in days. Estimates of γ_1 and γ_2 from equation (8) interacted with current employer size category fixed effects $\psi_{s(i,t)}$. *bigExp* is experience acquired in employers with 250+ employees. *Exp* is overall experience (including *bigExp*). Regression includes worker fixed effects, experience (quadratic), tenure (quadratic), age (quadratic), unemployment rate (quadratic), 21 sector fixed effects, fixed-term contract fixed effects, and month fixed effects. Standard errors clustered at the worker level in parentheses. * 0.10 ** 0.05 *** 0.01.

Table A9: Differential returns to experience at large employers: Promotion arrival rate

	(1)	(2)
<i>bigExp</i>	1.4458*** (0.2675)	1.6964*** (0.2847)
<i>bigExp</i> · <i>Exp</i>	-0.0001** (0.0001)	-0.0002*** (0.0001)
<i>Exp</i>	-6.6838*** (0.2043)	-6.7633*** (0.2046)
<i>Exp</i> ²	0.0013*** (0.0000)	0.0013*** (0.0000)
<i>Prof.cat.</i> – <i>duration</i>	16.7451*** (0.1343)	16.7498*** (0.1343)
<i>Prof.cat.</i> – <i>duration</i> ²	-0.0028*** (0.0000)	-0.0028*** (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.

Table A10: Career outcomes: Reduced-form estimates

	Lifetime income				Other outcomes		
	lifetime inc. 0% (1)	lifetime inc. 1% (2)	lifetime inc. 2% (3)	lifetime inc. 3% (4)	average daily wage (5)	days worked (6)	lifetime earn. 0% (7)
labor demand instrument	0.0111*** (0.0040)	0.0114*** (0.0040)	0.0117*** (0.0040)	0.0120*** (0.0041)	0.0078** (0.0036)	0.0027 (0.0016)	0.0105** (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 \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Table A11: 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.

Table A12: 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*** (0.0044)	0.0537*** (0.0045)	0.0312*** (0.0034)	0.0331*** (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 (389 average number of employees) 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.

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). While access to these data is restricted, 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.

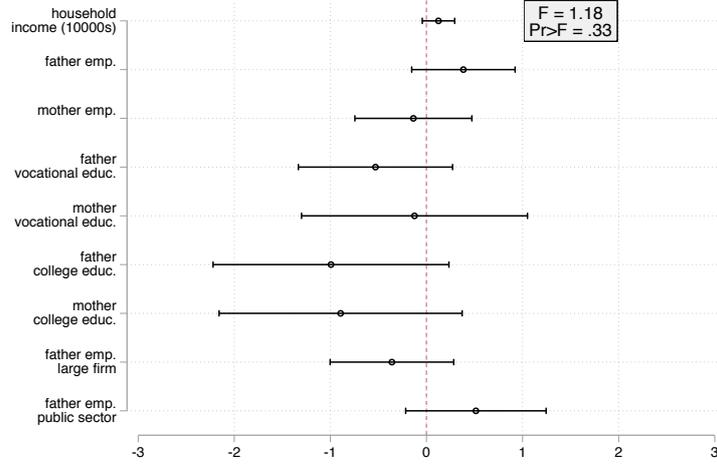
Using EU Households Panel (*Panel de Hogares de la UE*, or PHOGUE) allows me to observe the relevant information for four birth cohorts from my sample (1977–1980). Collapsing the data to the $\{rec\}$ cell level I run the following regression:

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

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, reassuringly, I fail to reject the joint test $\psi = 0$.

¹I 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 the potential endogenous response of educational investment decisions to the large-firm demand shocks that the IV leverages. I check for this possibility studying whether, after controlling for the unemployment rate, 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:

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

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

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 ν s and κ s 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 composition at 17	-0.0053 (0.0043)	-0.0028 (0.0052)			-0.0023 (0.0031)	-0.0044 (0.0055)
labor demand composition at 20			-0.0036 (0.0028)	-0.0071 (0.0051)	-0.0033 (0.0027)	-0.0065 (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 4.5, 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. The first robustness check changes 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 in this case 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, and they are almost identical to baseline.

Past business cycle conditions. As previously discussed, one might 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

²I 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.

specification which controls for past unemployment at a workers' region of birth. In particular, it additionally controls 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 again 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 (3) 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 (3) 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 (3) 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.

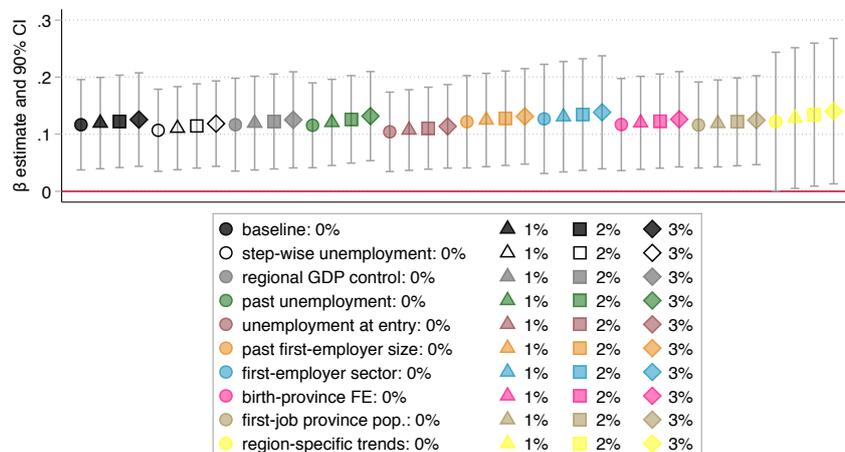
Provincial size of first job. Larger employers are typically located in more populated areas. One could argue that this is part of the set of attributes defining large firms. However, we would like to know if the first-employer size premium is simply driven by geographical effects of more populated areas. I find that this is not the case by estimating equation (3) with an additional control: (log) population of the province where the worker held his first

³In 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.

⁴Appendix 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 (3).

job. The results from this specification are represented by the brown markers in Appendix Figure B2 and are essentially identical to baseline.

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 (3) 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.

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. 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 these 11 years:

$$Y_i^{05-15} = \sum_{t=2005}^{2015} y_{it}. \quad (\text{B4})$$

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 (3), using $\ln(Y_i^{05-15})$ as dependent variable.⁸ Appendix Table B2 shows the OLS, first stage, and 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, TSLS is equal to .1408 compared to .1166–.1255.

The second robustness 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 Appendix Table 7). 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 7. Reassuringly, the TSLS estimates are practically the same, .085 compared to .089.

⁵This 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.

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

⁷This 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.

⁸I 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.

⁹Again, 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	Second Stage
	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	Second Stage
	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 (3), 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:

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

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

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

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

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

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

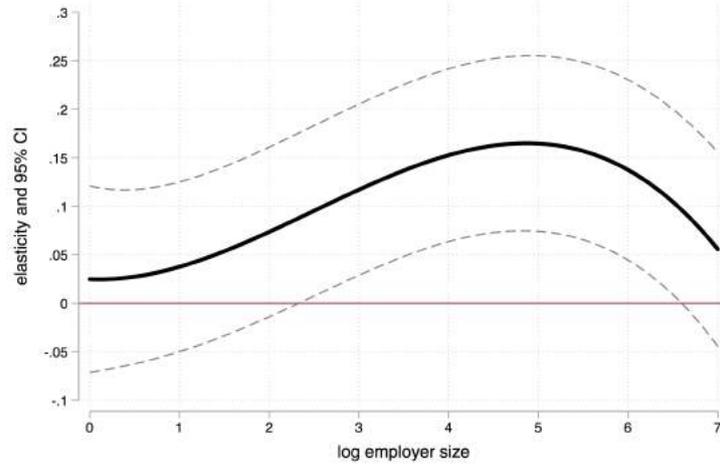
$$y_i = \beta_1 s_{J(i)} + \beta_2 s_{J(i)}^2 + \beta_3 s_{J(i)}^3 + \beta_4 s_{J(i)}^4 + \delta' X_i + \sum_{l=0}^4 \kappa_l \hat{\nu}_i s_{J(i)}^l + \varepsilon_i. \quad (\text{B8})$$

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 when 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.

¹⁰The 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 $\hat{\epsilon}(s) = \hat{\beta}_1 + 2\hat{\beta}_2s + 3\hat{\beta}_3s^2 + 4\hat{\beta}_4s^3$. 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*.

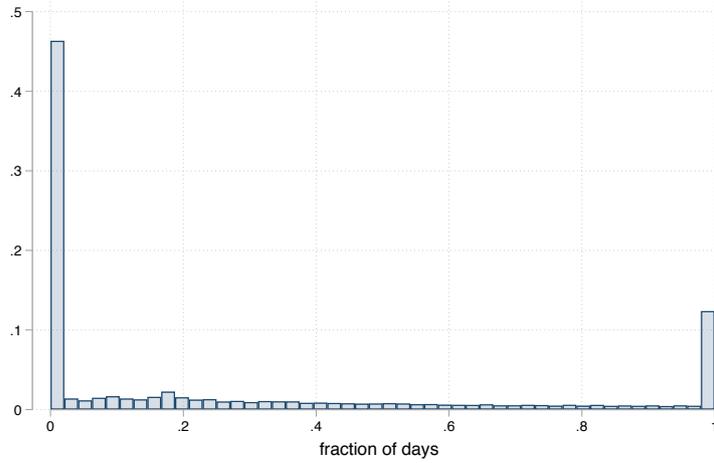
To test for the link between first-employer size and these two indices, I use them as outcome variables in OLS and IV estimations of equation (3). Table B5 shows the results from this exercise. Columns (1) and (2) show that in OLS first-employer size does a good job at predicting job security experienced 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.

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

¹²Given 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 (see Section 3).

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	Second Stage	
	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*** (0.0011)	0.0098*** (0.0009)		-0.0640* (0.0372)	0.0543 (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 (3), 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 \times education \times birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

B.8 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 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

$$V_t(u_t) = u_t + \beta \left[\lambda \mathbf{E}[\max\{V_{t+1}(u_t), V_{t+1}(u)\}] + (1 - \lambda)V_{t+1}(u_t) \right], \quad (\text{B9})$$

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:

$$\mathbf{E}[\max\{V_{t+1}(u_t), V_{t+1}(u)\}] = F(u_t)V_{t+1}(u_t) + \int_{u_t}^{\bar{u}} V_{t+1}(u)f(u)du. \quad (\text{B10})$$

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:

$$\mathbf{E}(u_{t+1}|u_t) = [(1 - \lambda) + \lambda F(u_t)] \cdot u_t + \lambda(1 - F(u_t)) \cdot \mathbf{E}(u|u > u_t). \quad (\text{B11})$$

It follows that:¹³

$$\frac{\partial}{\partial u_t} \mathbf{E}(u_{t+1}|u_t) = (1 - \lambda) + \lambda F(u_t) > 0. \quad (\text{B12})$$

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

$$V_t(u_t) = u_t + \beta \left[(1 - \delta)\lambda \mathbf{E}[\max\{V_{t+1}(u_t), V_{t+1}(u)\}] + (1 - \delta)(1 - \lambda)V_{t+1}(u_t) + \delta D_{t+1} \right]. \quad (\text{B13})$$

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

$$D_t = \beta \left[\lambda \mathbf{E}[V_{t+1}(u)] + (1 - \lambda)D_{t+1} \right]. \quad (\text{B14})$$

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

$$K_{t+1} = K_t + A_t K_t, \quad (\text{B15})$$

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 develop-

¹³Using the fact that $\frac{\partial}{\partial u_t} \mathbf{E}(u|u > u_t) = \frac{f(u_t)}{1 - F(u_t)} [\mathbf{E}(u|u > u_t) - u_t]$.

ment they offer.¹⁴ Under this setup, the value of employment in period t at a firm with human capital productivity A_t is given by

$$V_t(K_t, A_t) = RK_t + \beta \left[(1 - \delta) \lambda \mathbf{E} [\max\{V_{t+1}(K_{t+1}, A_t), V_{t+1}(K_{t+1}, A)\}] + (1 - \delta)(1 - \lambda)V_{t+1}(K_{t+1}, A_t) + \delta D_{t+1}(K_{t+1}) \right]. \quad (\text{B16})$$

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

$$D_t(K_t) = \beta \left[\lambda \mathbf{E} [V_{t+1}(K_t, A)] + (1 - \lambda)D_{t+1}(K_t) \right]. \quad (\text{B17})$$

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,

$$K_t = g(K_0, \{A_\tau\}_{\tau=0}^{t-1}). \quad (\text{B18})$$

Finally, note that the human capital accumulation function (B15) 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

$$K_{t+1} = K_t + A_t f(a_t) K_t, \quad (\text{B19})$$

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

B.9 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.

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.¹⁵ Then, workers with big-firm experience are paid more not because of what they have learnt at these jobs, but because employers believe

¹⁴In 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.

¹⁵Under 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.

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.¹⁶

I estimate specifications of equation (8) that allow for the differential value of large-employer experience to vary by current employer tenure. In particular I augment equation (8) by estimating

$$\ln w_{it} = \alpha_i + \psi_{s(i,t)} + \gamma_1 bigExp_{it} + \gamma_2 (bigExp_{it} \cdot Exp_{it}) + \gamma_3 (bigExp_{it} \cdot Tenure_{it}) + X'_{it} \delta + \varepsilon_{it} \quad (\text{B20})$$

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 (B20) regressors,

$$\frac{\partial \mathbf{E}(\ln w_{it} | Z_{it})}{\partial bigExp_{it}} - \frac{\partial \mathbf{E}(\ln w_{it} | Z_{it})}{\partial smallExp_{it}} = \gamma_1 + \gamma_2 Exp_{it} + \gamma_3 Tenure_{it} \quad (\text{B21})$$

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 A7 show estimates of equation (B20) 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.¹⁷

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.

[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 (8) 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

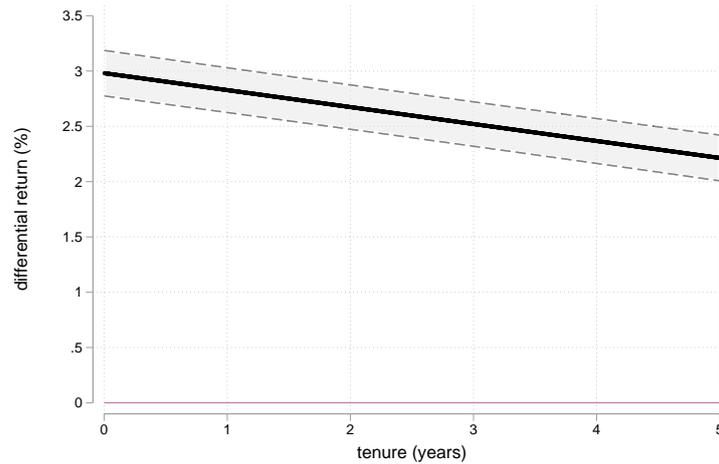
¹⁶One caveat of my approach is that the original test of [Altonji and Pierret \(2001\)](#) requires that the wage return to unobserved ability also varies over time in order to load the effect of learning about employer ability. In my case, I rely on the worker fixed effect as capturing unobserved ability. Since this effect is fixed over time, I might be underestimating the rate of decay of the return of large-employer experience.

¹⁷To 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.

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.¹⁸

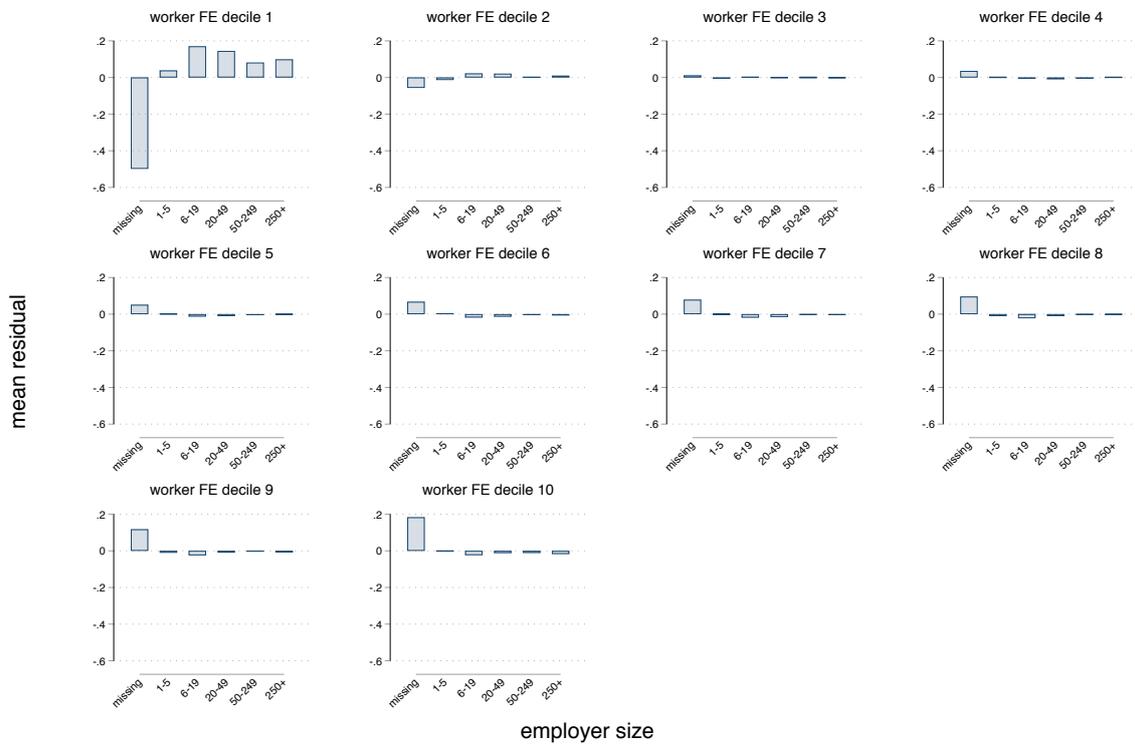
Figure B5: Differential wage return to one year of large employer experience, by current employer tenure



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 (B20) (in Table A7, 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.

¹⁸Mean 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 (8) with cells defined by decile of estimated worker effects (α_i) interacted with employer size category.

B.10 Details on the promotion variable

In Section 6.2 I study the differential returns to large-firm experience in terms of career progression through promotions. I construct a promotion proxy using information in social security data on professional categories.

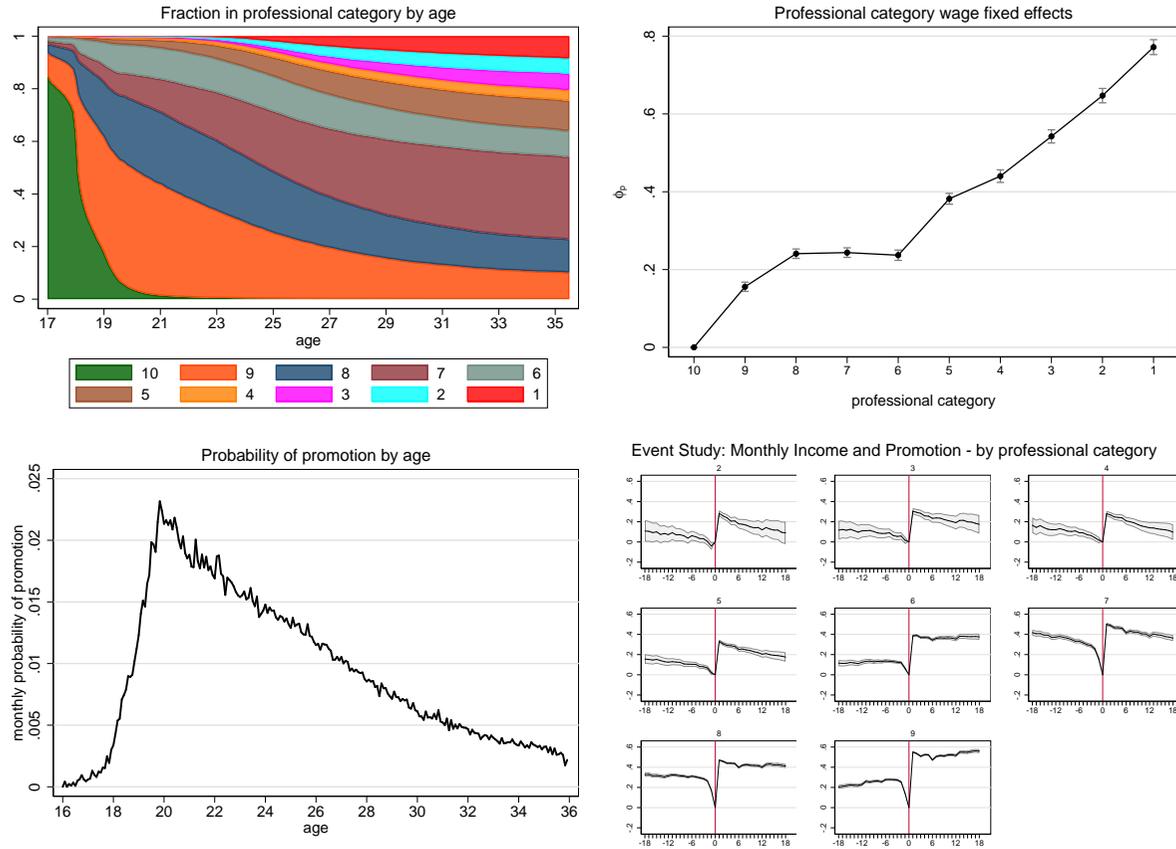
In particular, the data include a professional category variable (“*grupo de cotización*”) that allows the creation of a promotion proxy. This variable is determined by the type of job a worker performs and not by her education level. There are originally 13 categories which I group into 10. I group 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.

Appendix Figure B7 provides support for the promotion definition I use. The top-left panel shows that as workers age they rise through the professional category ranks. The top-right panel shows that the wage premium associated with each professional category is increasing in rank.¹⁹ The bottom-left panel shows that the monthly probability of promotion as a function of worker age follows an intuitive pattern: highest in the early twenties and declining but non-zero from then onwards. Finally, the bottom-right panel shows in an event-study framework that workers experience a positive jump in their monthly earnings whenever a promotion arrives.

¹⁹The panel plots professional category fixed effects ϕ_p from the regression $\ln w_{it} = \alpha_i + \psi_{s(i,t)} + \phi_{p(i,t)} + X'_{it}\delta + \varepsilon_{it}$, where w_{it} is the monthly wage, α_i are worker fixed effects, $\psi_{s(i,t)}$ are current employer size category fixed effects, and X_{it} includes time-varying controls: a quadratic term for total experience, tenure at current employer (quadratic), current unemployment level (quadratic), type of labor contract (permanent or fixed-term), sector fixed effects, and time (year-month) fixed effects.

Figure B7: Professional categories and promotions



Notes: Top-left panel plots the fraction of workers in each of ten professional categories by age. Top-right panel plots the professional category fixed effects and 99% confidence intervals from a regression of log monthly wages on worker fixed effects, current employer size fixed effects, experience (quadratic), tenure (quadratic), current unemployment rate (quadratic), sector fixed effects, type of contract, and month fixed effects. Bottom-left panel plots the monthly probability of promotion by age. Bottom-right panel plots monthly (log) income event study estimates and 95% confidence intervals of time-from-promotion dummies, separately for each initial (before promotion) professional category (movements away from professional category 10 are not counted as promotions since they are based on worker age). Regressions in event study control for current employer size category, employment status, experience (quadratic), tenure (quadratic), unemployment rate (quadratic), sector fixed effects, type of contract, and year fixed effects.

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 4.6) 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 more rural areas 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 $s = 0, 1, 2, \dots, J$ are denoted by Y_{si} .²⁰ The instrument (labor demand environment) is represented by Z_i . It could be binary $Z_i \in \{0, 1\}$, or multivalued $Z_i \in \{0, 1, 2, \dots, K\}$. 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_{Z_i} 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: $S_{1i}, S_{0i}, Y_{0i}, Y_{1i}, \dots, Y_{Ji}$ 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

²⁰Positive 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

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

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. These 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

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

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 is interested in knowing whether 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:

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

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 $k = 0, 1, \dots, K$ 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

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

where

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

Their Theorem 2 concludes that in the multivalued instrument case

$$\begin{aligned} \beta^{TSLS} &= \sum_{k=1}^K \mu_k \beta_{k,k-1} \\ &= \sum_{k=1}^K \sum_{s=1}^J \mu_k \omega_s^{k,k-1} \cdot E[Y_{si} - Y_{(s-1)i} | S_{ki} \geq s > S_{k-1,i}] \end{aligned} \quad (\text{C6})$$

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):

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

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

$$F(s|Z_i, X_i) = \Lambda\left(g(Z_i, X_i; \theta(s))\right) \quad \text{for all } s \in \mathcal{S} \quad (C8)$$

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 $g(Z_i, X_i; \theta(s))$ to be the same linear function of the instrument and controls used in the TSLS estimation,

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

where the controls X_i are the same as in the main IV specification from equation (3): 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 $\theta'(s) = [\gamma_0(s), \gamma_1(s), \delta(s)]$ for each $s \in \mathcal{S}$ involves running the following $|\mathcal{S}|$ logit regressions:

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

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

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

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

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

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: (C11)

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

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:

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

where:

$$\hat{\phi}(k) = \left[\frac{1}{N} \sum_{i=1}^N \sum_{s=1}^J \left(-\hat{\gamma}_1(s) \cdot \Lambda(\hat{\gamma}_0(s) + \hat{\gamma}_1(s)k + X_i' \hat{\delta}(s)) \cdot \left[1 - \Lambda(\hat{\gamma}_0(s) + \hat{\gamma}_1(s)k + X_i' \hat{\delta}(s)) \right] \right) \right]^{-1}$$
(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. Appendix Figure C2 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.²¹ 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).²²

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

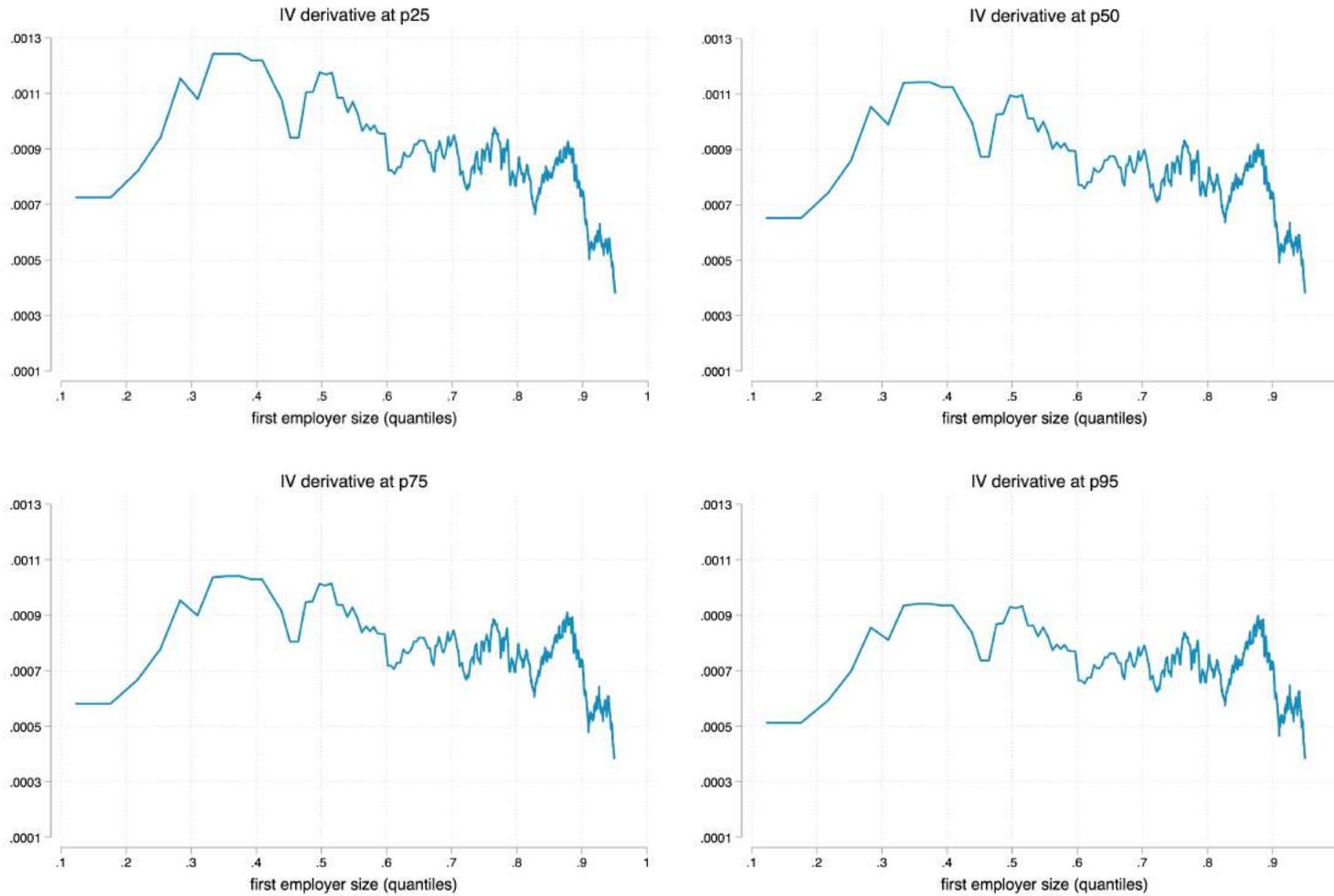
²¹Results are similar for other IV values. A high IV value represents a substantial large-firm hiring shock like the DuPont example I discuss in Section 4.

²²I classify workers as rural- or urban-born based on their province of birth. I use data from Goerlich Gisbert and Cantarino Martí (2015) who estimate the fraction of urban (living in urban clusters, based on population and population density) and rural population (living outside these clusters) at the municipality level using data from 2006. I use their province aggregates and I classify as (more) rural provinces with over 15% of its population being rural. This number is around the population weighted median across provinces in the original data, and it is close to the median in my sample.

employer sizes: The instrument seems to impact more the movements across this part of the distribution for urban-born and college workers.

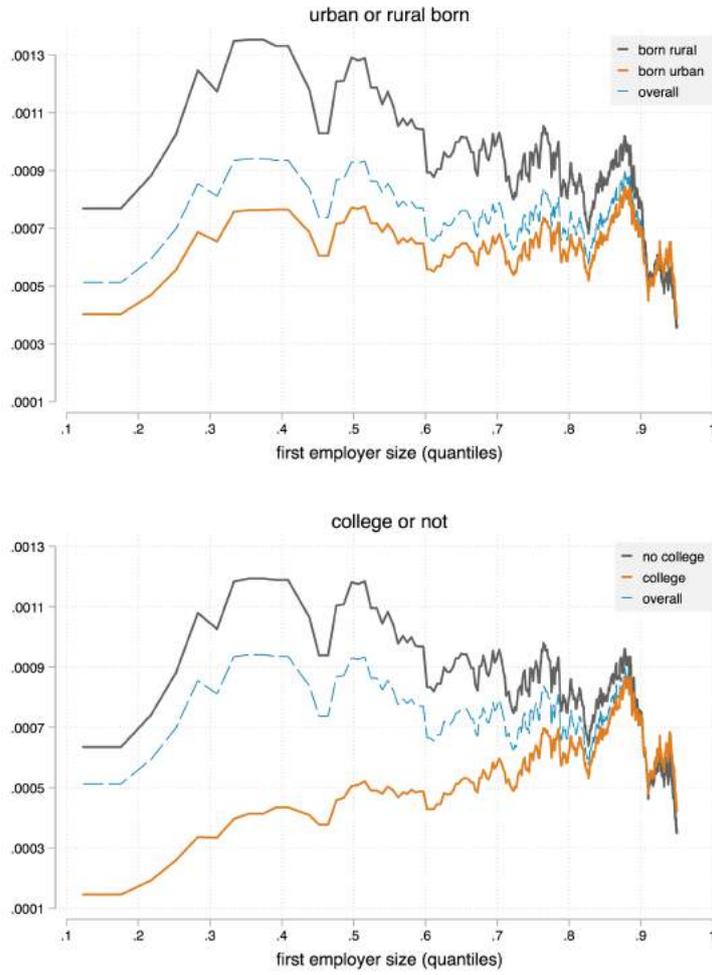
Overall, the analysis carried out in this section supports the intuition laid out in the paper with respect to the comparison between OLS and TSLS elasticities: The “compliers” who are mostly driving the TSLS estimates are less educated and come from less urban places of Spain. The “large” TSLS estimates (in comparison to OLS) seem to imply that these are the workers who are more sensitive to their first employer characteristics. This is plausible. The fact that these workers are younger and have less education is consistent with them benefiting the most from the human capital opportunities large employers offer. In addition, these workers will more easily get “stuck” in bad jobs at the beginning of their careers, and face worse outside options if they do not match with a good first employer.

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 .

Figure C2: Compliers' characteristics: weight function heterogeneity from flexible 1st stage



Notes: Estimated weight function from equation (C12) for different subgroups. As a function of first employer size s , holding the instrument constant in the 95th percentile.

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 $j \in \{1, 2, \dots, J\}$ 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

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

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.²³

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.²⁴ Workers choose the firm j which provides the highest utility. Given the distributional assumption, the probability of choosing firm j is given by (McFadden, 1974):

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

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

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

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

²³A 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 $b = 0$.

²⁴It 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 .²⁵

$$Y_j = A_j N_j.$$

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.²⁶

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

²⁵The 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.

²⁶In 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, $\mathbb{1}\{\cdot\}$ 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.²⁷ 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 $\theta \in (0.5, 1)$:

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

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:

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

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

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

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

²⁷This 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 :

$$\begin{aligned}
\mathbf{E}(w_{ik_1}^1 | k_0 = j) &= \mathbf{E}\left[\mathbf{1}\{S_i = H\} \cdot \beta \ln w_{Hk_1}^* | k_0 = j\right] + \mathbf{E}\left[\mathbf{1}\{S_i = L\} \cdot \beta \ln w_{Lk_1}^* | k_0 = j\right] \\
&= \beta \left[q_j \cdot \underbrace{\mathbf{E}(\ln w_{Hk_1}^*)}_{\equiv \tilde{w}_H} + (1 - q_j) \cdot \underbrace{\mathbf{E}(\ln w_{Lk_1}^*)}_{\equiv \tilde{w}_L} \right] \\
&= \beta \left[\underbrace{\tilde{w}_L + q_j(\tilde{w}_H - \tilde{w}_L)}_{\equiv \psi_j} \right]
\end{aligned}$$

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

$$u_{ij}^0 = \beta \ln w_j + \psi_j + \varepsilon_{ij}^0, \quad \text{where } \psi_j \equiv \beta \left[\tilde{w}_L + q_j(\tilde{w}_H - \tilde{w}_L) \right].$$

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.