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

Jaime Arellano-Bover

October 29, 2022

Abstract

I study the long-term effects of landing a first job at a large firm versus a small one using Spanish administrative data. Size could be a relevant employer attribute for inexperienced workers since large firms are associated with greater productivity, wages, and training. The key empirical challenge is selection into first jobs based on unobserved worker characteristics. I develop an instrumental-variables approach that, keeping business-cycle conditions fixed, leverages variation in the *composition* of labor demand that labor-market entrants face. Initially matching with a larger firm persistently improves long-term outcomes, even through subsequent jobs. Mechanisms suggest better skill-development at large firms.

An earlier version of this paper was the first chapter of my dissertation at Stanford University; I thank Caroline Hoxby, Ran Abramitzky, Luigi Pistaferri, and Isaac Sorkin for their advice and encouragement. I have benefited from useful comments from the editor Fabian Lange, Manuel Arellano, Barbara Biasi, Nick Bloom, Mark Duggan, Thomas Ginn, Pete Klenow, Guido Martirena, Francesc Ortega, 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 other venues. 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.

Affiliations: Tor Vergata University of Rome, Einaudi Institute for Economics and Finance (EIEF), and IZA. E-mail: ***j.arellano.bover@gmail.com***.

1 Introduction

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 Alternatively, imagine she starts at an unproductive firm with no training schemes, outdated technologies, unsophisticated management, and few coworkers.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

On the one hand, young workers are mobile (Topel and Ward, 1992), so initial matches might not be relevant in the long run; there will be time to find a good job later on. On the other hand, first employers could affect career paths: search for ensuing jobs could vary based on first-employer quality, and opportunities to learn useful skills might differ across firms. For a young adult in her formative years, these distinctions could persistently impact her working life. There is abundant evidence on firm-driven wage inequality focusing on contemporaneous worker-firm matches (e.g., Abowd et al., 1999; Card et al., 2018). However, much less is known of how workers are impacted by past employment at heterogeneous firms.

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

The IV strategy uses variation in regional labor demand *composition*. The logic underlying the IV is that idiosyncratic shocks in the hiring decisions of large firms can generate

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

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

[3] The relevant thought experiment is random assignment of entrants to be hired by a larger or smaller firm. Firms that differ in size are typically different in other attributes, all of which form part of the thought experiment of being hired by potential first employers of different sizes and likely drive first-employer size effects. In contrast, exogenously increasing the size of a given firm is not the relevant thought experiment.

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

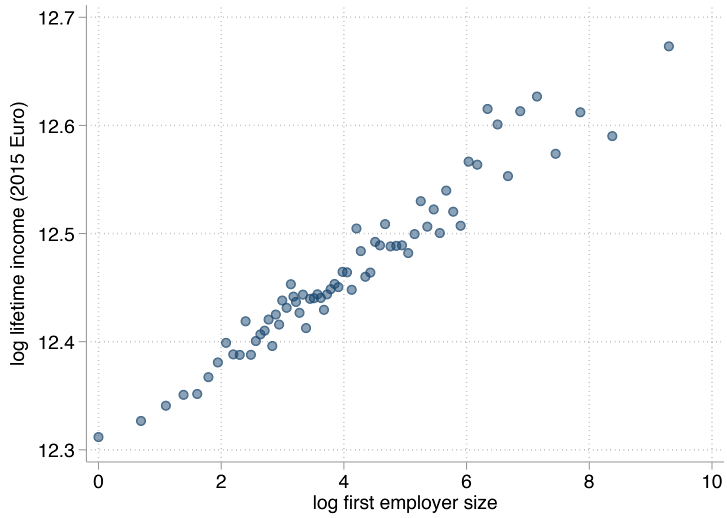
The Spanish context provides rich variation in large-firm hiring shocks. During 1985–2003, the years of labor market entry I study, Spain underwent an economic transformation following adhesion to the EU in 1986 (Chislett, 2002). This period was characterized by an opening to trade, growth in foreign firms' investments, market reforms, and expansion of regional infrastructures. These factors led to great dynamism in large firms opening and expanding operations across different parts of the country. This variation allows keeping constant the effects that cyclical conditions at entry might have on long-term outcomes, which has been the focus of previous work (see von Wachter, 2020). I keep cyclical conditions constant by controlling for regional unemployment rates, thus only using variation in large-firm hiring that is uncorrelated with business-cycle trends.

My results show that matching with a good first employer can shape workers' long-term career prospects. The raw data display a positive correlation between lifetime income (a cumulative measure of many years of monthly labor income) and first-employer size. Figure 1 illustrates this unconditional correlation.[4] The main result is that adding controls and using the IV approach to account for workers' unobservable characteristics confirms the patterns in the raw data: I estimate a positive IV elasticity of lifetime income with respect to first-employer size, equal to 0.12. This magnitude is meaningful: a one standard-deviation increase in log first-employer size is associated with a 27.7% increase in lifetime income. The first stage, which does a good job at predicting first-employer size, implies that, at least for some, luck plays a role in the key process of matching with heterogeneous first employers. The effect on lifetime income can be attributed both to an increase in average daily wages, explaining 74% of the effect on income, and an increase in total days worked, explaining the remaining 26%. Accounting for first-employer sector fixed effects does not change results.

The IV estimate of the elasticity between lifetime income and first-employer size is about four times larger than the OLS, similarly to related IV-OLS comparisons in the literature (Kahn, 2010). This comparison might run counter to what simple intuition would predict about unobserved ability being positively correlated with first-employer size. However, in a context of heterogeneous effects, this is consistent with "compliers," those whose first-employer match is more susceptible to the IV, benefiting the most from a first job at larger firms. Building on Angrist and Imbens (1995), I estimate "complier weights" that shed light

[4]US panel survey data show a similar correlation (see online Appendix Figure A1).

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



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

on who are the people whose first jobs are most affected by the variation the IV captures. I find that compliers tend to be less educated and from less urban areas. This LATE result indicates I capture the causal effect for younger entrants with lower earnings potential, who might be of special interest.

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

On the mechanisms behind the first-employer size effect, I first confirm that the lifetime effect is truly persistent, not solely stemming from time spent at the first job. Evidence of persistence includes the low fraction of lifetime income that is earned at the first job (due to job mobility and wage growth), and first-employer effects that are still present at age 35 (an age at which income trajectories have stabilized and 93% of workers have left their first job). Based on this persistence, likely mechanisms are related to what that the literature identifies as main sources of life-cycle wage growth: human capital accumulation and job search (Rubinstein and Weiss, 2006).

I find that first-employer size has a positive causal effect on the size of ensuing employers (i.e., second employer, and employer at age 35). That is, a larger first employer leads to larger subsequent ones, implying that first-job matches affect future firm-to-firm mobility.

Persistence in employer characteristics could be related to human capital if skills acquired at large firms are more valuable in other large firms. Additionally, such persistence could arise from job ladder effects driven by search frictions.

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

Contribution to the literature. This paper forms part of a broad literature showing how the identity of the firm hiring an individual matters for labor-market outcomes and, as a result, workers' luck matching with employers affects careers. This group includes work related to the two-way fixed effects model of Abowd et al. (1999) (AKM), the firm-size wage premium (Brown and Medoff, 1989; Oi and Idson, 1999), long-term losses arising from job displacement (Jacobson et al., 1993), or the effects of entering the labor market during downturns (von Wachter, 2020). The big-picture contribution of this paper, relative to all these works, is to document the dynamic effects of first-job employer-employee matches.

Works building upon AKM that document firms' contribution to inequality (e.g., Card et al., 2013, 2018; Song et al., 2019) focus on contemporaneous worker-firm matches, while evidence on dynamic effects of employment at heterogeneous firms is more limited.[5] I contribute to this literature by establishing a causal link between first-employer attributes and career outcomes, with implications for long-term inequality.[6] Moreover, my cross-sectional IV identification method innovates relative to approaches relying on workers' moves across firms. Lastly, my findings could have implications for AKM firm premia: large firms' AKM pay premia would be overestimated in contexts where past large-firm employment is associated with greater human capital growth *and* large-firm employment today.[7]

While the firm-size premium literature goes back all the way to Moore (1911), there is no consensus on why the premium exists or whether it has a causal component.[8] This paper documents, with a causal interpretation, that first jobs at large firms lead to persistently better career outcomes.[9] Relative to the job displacement literature, my findings help explain larger losses among workers displaced from large firms (Fackler et al., 2021), and especially

[5]Abowd et al. (2018) and Bonhomme et al. (2019) provide some evidence on dynamic implications of employment at heterogeneous firms. See also Di Addario et al. (2021). Gregory (2019) and Arellano-Bover and Saltiel (2021) study heterogeneous human capital acquisition across firms.

[6]Other papers on first jobs study specialized workers such as Ph.D. economists, MBAs (Oyer, 2006, 2008), CEOs (Schoar and Zuo, 2017), or medical doctors (Arora et al., 2021).

[7]Kline et al. (2020) show a positive correlation between firm size and AKM pay premia.

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

[9]Without focusing on entrants, Sorenson et al. (2021) document persistent penalties associated with employment at startups (i.e., young firms).

so evidence among young workers (von Wachter and Bender, 2006).

Lastly, the "graduating-in-a-recession" effect (Kahn, 2010; Oreopoulos et al., 2012; Altonji et al., 2016; Schwandt and von Wachter, 2019) is related to the first-employer size effect since, as this literature shows, inexperienced workers are more likely to be hired by large firms during booms. In spite of this body of work, evidence on mechanisms does not abound.[10] This paper improves our understanding of the mechanisms behind this literature. By studying first-employer heterogeneity—one of the suggested mechanisms—but doing so while keeping constant business-cycle fluctuations, I quantify how much of the "graduating-in-a-recession" effect can be explained by the first-employer size effect.[11]

The rest of the paper is organized as follows. Section 2 describes the data, measurement, and context. Section 3 presents causal effects of first-employer size on long-term outcomes. Section 4 studies persistence and mechanisms. Section 5 concludes. Several online appendices contain supplementary results and robustness checks.

2 Data, Measurement, and Context

2.1 Spanish Social Security Administrative Records

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

The data include 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. Education is is observed thanks to a merge with municipal registries. While education is a key variable when studying youth labor market entry, many times it is not recorded in administrative labor market data, making MCVL well-suited for this topic. I group educational attainment 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, contract type (permanent/temporary), professional category (*grupo de cotización*), and each month's pay-

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

[11] Other work on mechanisms behind graduating-in-a-recession effects are Kwon et al. (2010); Liu et al. (2016); Wee (2016); Arellano-Bover (2022).

roll 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. I follow a procedure similar to Bonhomme and Hospido (2017) to impute monthly earnings for censored observations.[12] Censored observations are few: 8.7% and 3% of month-person observations in my sample are top- and bottom-coded, respectively. Since the taxable base of the self-employed is not a function of their monthly income, I do not observe earnings for them.

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

Social security records are matched with uncensored annual earnings tax data for the years 2005–2015. The downside from using tax records to study long-term effects is that the time series is significantly shorter and residents of two Spanish regions, the Basque Country and Navarre, are not in the data. I use tax data to show that the main lifetime results are robust when using measures of cumulative earnings derived from uncensored tax earnings. I also use tax data to document the cross-sectional firm-size premium.

Employers are represented in the data through their anonymized social security account numbers. For workers in the general regime of social security,[13] each firm has one account for each province in which it employs workers. There are 50 provinces in Spain which are further grouped into 17 autonomous regions. An employer identifier in the data thus represents a firm-province combination. This notion of employer is equivalent to a firm for single-establishment firms, and smaller than a firm—closer to an establishment—for firms operating in several provinces. Firm-province is the employer definition I use throughout the paper.[14] Since this paper focuses on size, and to the extent that large firms are large employers relative to other employers in the provinces in which they operate, using employer or firm size should not make much of a difference, other than compressing the size distribution. A drawback of this employer definition might arise from rare cases in which I assign a small first employer to workers who are in fact matched to a large firm in a province in which it has a small presence. Unfortunately, I do not observe firm size whenever it differs from firm-province.

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

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

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

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

employer unit because employer size is observed at this level. Since I observe a sample of workers and not the population, I cannot "aggregate up" from employer size to firm size.

In the original MCVL data typically made available to researchers, employer size is only observed starting on 2004. However, I obtained a new data extract recording the evolution of size for the employers in my sample, going back to 1980. This extract allows measuring employer size at any point in time during the sample years of labor market entry, which in this study is key in order to avoid reclassification bias (i.e., assigning a large first-employer to a worker who had a small first-employer that grew).[15]

Supplementary data sources are described in online Appendix B, Section B.1.

2.2 Sample Selection

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

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

The data requirements for the cross-sectional long-term analysis are stringent since each observation aims to capture information about the full labor market history of a given worker. For each person, I require information on his first labor market experience, and enough lifetime information on employment and earnings. Thus, I impose additional restrictions for this analysis that reduce the number of workers in the sample. I include those who, between 16 and 35 years, have sufficient attachment to the formal labor market: those who are employed for half or more of the months since labor market entry up until the year they turn 35. This type of sample selection criteria is present in other studies analyzing lifetime income (Guvenen et al., 2022). I further 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,

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

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

25 for vocational, and 28 for college),[17] or in a Social Security regime different than the general regime. All these restrictions result in a sample of around 80,000 people, 50% of those originally in the raw data. Table 1 shows 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 3. 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.

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

2.3 Definitions and Measurement

First labor market experience. I define a worker's first labor market experience as the first six continuous months after predicted graduation that a person works for 100 days or more. This definition aims to capture the first relevant job after finishing formal education, while avoiding summer work or very temporary employment. Workers in my sample entered the workforce during the late 1980s, 1990s, and early 2000s.

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

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

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

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

There is a tradeoff between how many cohorts are studied 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), 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 when incomes stabilize and trajectories are more easily predictable.[18] Table 1 provides summary statistics for this measure. The median is 254,142 Euro (2015).[19]

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

[18] 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 online Appendix Figure A2 for evidence for Spain on income profiles stabilizing during the mid 30s.

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

zeroes or unemployment benefits) deals with this issue.

2.4 Large and small firms in Spain

Spain has relatively few large firms. According to 2013 OECD data, 0.4% of Spanish enterprises have 250 employees or more. This percentage is comparable to that from Portugal or Italy but far below Germany (around 2%) or the US (around 1.5%; see online Appendix Figure A3). 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).

As a result, compared to other contexts, few young workers are employed at large firms, which the literature suggests tend to offer more desirable jobs (Sorkin, 2018). Firm attributes associated with a large size are likely similar in Spain and other countries (see online Appendix D). However, compared with Germany or the US, the outside option of a young Spaniard who does not match with a large employer might disproportionately be a very small and possibly unproductive firm.[20] 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.

The size premium in the cross-section

Before focusing on first jobs and long-term outcomes, Table 2 shows the cross-sectional relationship between contemporaneous employer size and contemporaneous job outcomes. Using the annual panel tax records 2005–15, I estimate regressions where (log) daily wages or contract type are outcome variables, the explanatory variable is current (log) employer size, and controls include workers' age and education, firm sector, and year fixed effects.[21] Column (1) shows that the unconditional elasticity between employer size and wages is equal to 0.083. This magnitude lies between the range of comparable estimates by Colonnelli et al. (2018) in Sweden and UK (0.01–0.02), Brazil (0.09), and Germany (0.12–0.14).[22] Column (2) adds controls making the estimated elasticity drop to 0.059. This magnitude aligns with the elasticities for Spain reported by Lallemand et al. (2007) using survey data and similar controls.

Column (3) in Table 2 allows the elasticity to differ according to workers' education. The elasticity is 33% larger for non-college workers (0.061) than for college-educated workers (0.045). Columns (4) and (5) show that large firms are slightly less likely to use fixed-term contracts, while column (6) shows that this overall negative correlation remains for non-college workers but it is even closer to zero and positive among college-educated workers. Overall, larger firms in Spain are associated with better job outcomes, especially among less educated workers.

[20] 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 US (online Appendix Figure A3).

[21] These controls mimic the ones I later use in the main analysis.

[22] Bloom et al. (2018) study the US size premium but consider discrete size categories instead of elasticities.

Table 2: Employer-size premium in the cross-section: wages and contract type

	(1) Daily wage	(2) Daily wage	(3) Daily wage	(4) =1 Fixed-term	(5) =1 Fixed-term	(6) =1 Fixed-term
Employer size	0.0832***	0.0585***		-0.0199***	-0.0110***	
	(0.0004)	(0.0005)		(0.0003)	(0.0003)	
Employer size × no college			0.0605***			-0.0143***
			(0.0005)			(0.0003)
Employer size × college			0.0454***			0.0096***
			(0.0012)			(0.0007)
Age	no	yes	yes	no	yes	yes
Education	no	yes	yes	no	yes	yes
Birth region FE	no	yes	yes	no	yes	yes
Sector FE	no	yes	yes	no	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
SE clusters (persons)	402,684	375,882	375,882	388,339	362,227	362,227
Observations	2,826,352	2,727,367	2,727,367	2,677,582	2,581,788	2,581,788

Notes: Annual panel 2005–2015, tax earnings, male workers ages 18–59. Employer size enters regressions in logs. Outcome variable in columns (1)–(3) is log daily wage; outcome variable in columns (4)–(6) is a dummy equal to one if contract is fixed-term (i.e., temporary). The sample mean of the fixed-term dummy is equal to 0.354. Regressions corresponding to columns (3) and (6) allow for education-specific age profiles. Sector fixed effects correspond to 3-digit sectors. Standard errors clustered at the worker level. * 0.10 ** 0.05 *** 0.01.

3 Size of First Employer and Career Outcomes

This section lays out the relationship between the size of a worker's first employer and long-term career outcomes. I document descriptive facts and discuss the IV approach that accounts for endogenous sorting of workers and firms. The thought experiment I wish to capture is random assignment of young workers to be hired by firms of different sizes, with other firm attributes associated with size *forming part* of this thought experiment. I do not capture a hypothetical exogenous increase in the size of a given firm. Larger firms are characterized by attributes that could impact young workers and likely drive any first-employer size effect.[23]

3.1 Descriptive Facts

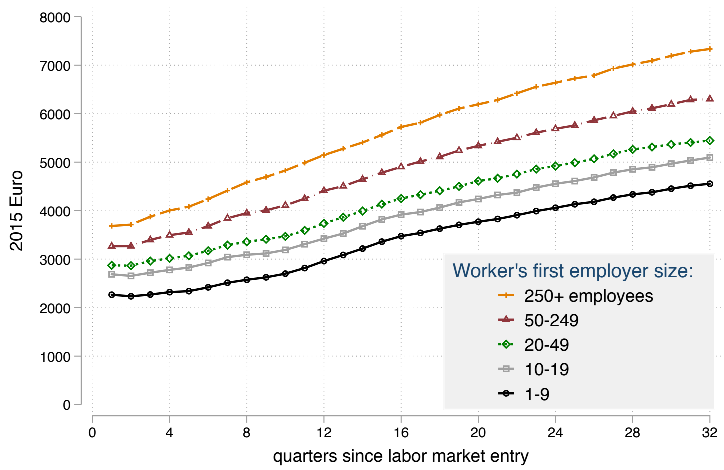
There is an unconditional positive relationship between the size of a worker's first employer and long-term career outcomes. Figure 1 (shown in Section 1) plots the unconditional relationship between the lifetime income measure and first-employer size. There is a strong positive relationship between the two variables which is linear in logs. The correlation coefficient is equal to 0.21. This relationship is not explained away by firms' industry (see online Appendix Figure A4).

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

[23] Online Appendix D discusses what these attributes 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 online Appendix D lays out, there are reasons to believe size could be a good proxy for such a scalar measure (e.g. training, productivity, new technologies).

patterns arise when examining employment and daily wages (online Appendix Figure A5).

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.

3.2 Empirical Approach: Estimating Equation and IV

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

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

where y_i is (log) lifetime income for worker i and $s_{J(i)}$ is the (log) number of employees of firm J where i held his first job.[24] c indexes birth cohorts, e refers to three educational attainment levels—high school, vocational, college—and r indexes regions of birth. $t_0(e,c)$ indexes a worker's *predicted* graduation year, which is a function of birth year c, and educational attainment e. Based on standard Spanish completion times, I assign year of predicted graduation as the year in which people with high school degrees turn 17, 20 for vocational education, and 23 for college education. The δs represent region of birth, education, and cohort fixed effects, while $f(u_{r,t_0(e,c)})$ is a flexible function of the unemployment rate in region r in year $t_0(e,c)$, capturing business cycle variation. At baseline, $f()$ is a quartic function which is allowed to differ across education groups. All variation in equation (2) is cross-sectional since each worker only has one first job and one measure of lifetime income.

OLS estimates of β are likely biased due to unobserved determinants of lifetime income that are plausibly correlated with first-employer size. For instance, large firms might be able to hire young workers who are more productive and would earn higher wages throughout their career regardless (see Arellano-Bover, 2021).[25] Similarly, young people who are able

[24] Size is measured at the year the worker joined the firm. Later, I show that results are robust to alternative size measures.

[25] Arellano-Bover (2021) provides evidence on first-job selection into large firms, based on education and cognitive skills, using PIAAC survey microdata from 31 countries.

to match with a large firm might be more proactive in their job search strategies, a skill that can lend returns throughout the working life. These and related reasons are the motivation for an IV strategy.

IV motivation

My IV approach uses variation in the composition of regional labor demand for inexperienced workers across time and space. The origin of such variation is tightly linked to firm size and the firm-size distribution. A well-known fact, which holds internationally, is that the firm size distribution is fat-tailed, being well approximated by a power law (Gabaix, 2009). This feature implies that large firms' *idiosyncratic* shocks, instead of disappearing on average, can generate quantitatively meaningful aggregate variation (see Gabaix, 2011). In other words, large-firm expansions and openings of new operations will make large firms hire batches of inexperienced workers differentially across years, and this variation will not necessarily average out, even if driven by purely idiosyncratic firm shocks. The goal of my IV is to use such resulting variation, as manifested in large-vs-small firm labor-demand *composition*, while controlling for the overall variation in demand *level*.

Variation in labor demand composition across years and regions implies that different young workers, depending on when and where they enter the labor market, will be exposed to different propensities to join larger or smaller firms. While the timing of large firms' idiosyncratic hiring shocks is plausibly exogenous from the point of view of an individual worker, young entrants' choice of where and when to search for a first job is clearly not, and could be influenced by said shocks. As such, to construct the IV, it is not advisable to use information on the actual region and year of entry. Instead, I will link variation in demand composition to individual workers based on their region of *birth* and their *predicted* graduation year (i.e., the expected graduation year based on year of birth and typical length of education qualifications). Using birth region and predicted graduation year addresses endogeneity issues but has consequences for the interpretation of IV results based on who the resulting compliers are, a matter I discuss below in Section 3.5.

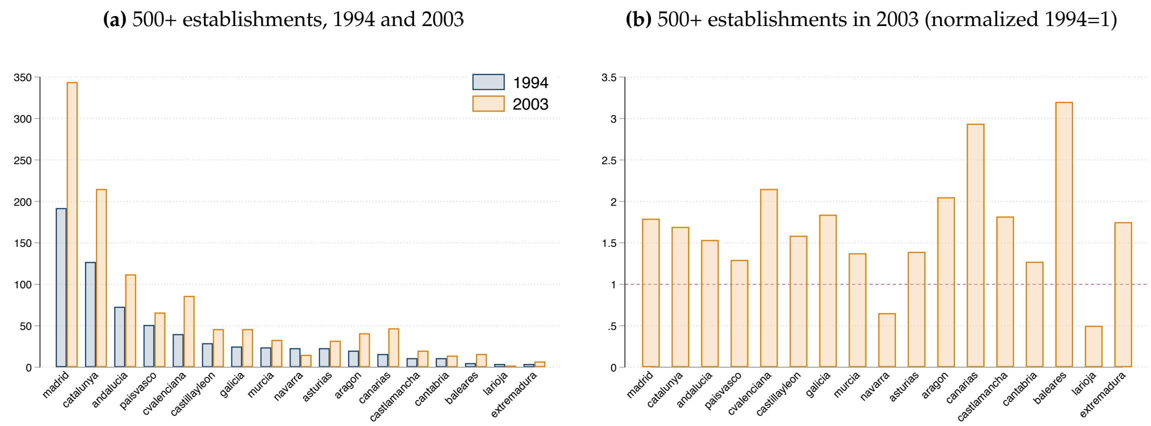
Simplified example. A schematic example based on a true event illustrates some IV intuitions. Consider two high school graduates who were both born in the Spanish region of Asturias, one year apart from each other. The graduation year of the younger person is 1993 and coincides with the opening in the region of a large and modern plant of the US multinational DuPont, which hires around 1,000 workers. The older worker's high school graduation was in 1992, one year earlier. This timeline suggests that the worker from '93 will be more likely to have his first job at DuPont than the worker from '92.[26] Similarly,

[26]In principle, the '92 worker could switch jobs after one year employed elsewhere and have his second job be at DuPont, starting in 1993. In practice, different types of frictions might limit such moves. The causal effect of interest consists of finding a first job at a larger firm, *leaving what happens later unrestricted*. Whether first jobs matter in this context is then an empirical question. If labor markets are very fluid, we might expect the '92 worker to quickly find a second job at DuPont. In such case, the data should not reflect any sizeable first-job effects on long-term outcomes. In practice, I find meaningful first-employer size effects on lifetime income. Furthermore, I later show that a larger first employer leads to a larger *second* employer, and a larger employer at age 35. This result speaks to mobility frictions and how subsequent moves are affected by initial matches.

given low regional mobility, 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.

Historical context. The goal of the IV is to aggregate and summarize variation resulting from large-firm hiring shocks across years and regions. 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 my sample years of labor market entry (1985–2003), Spain was undergoing a period of economic transformation following adhesion to the European Union in 1986 (Chislett, 2002). This period was characterized by an internationalization of the economy: an increased openness to trade, and growth of foreign firms' investments in the country. It also featured reforms towards market liberalization, and large investments in regional infrastructures. This context led to great dynamism in large firms opening and expanding across the country, contributing to the variation that the IV approach leverages. Figure 3 illustrates this trend using register data on the population of establishments. For each region, the figure shows the number of establishments with 500+ employees in 1994 (the first year for which these data exist) and in 2003 (the last year of labor market entry in my sample). In 15 out of 17 regions the number of large employers increased, and in most of them substantially so. This pattern holds even for regions that initially had 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, 1994–2003. Panel (a): number of 500+ establishments in each period. Panel (b): ratio between 2003 and 1994. Sample period of labor market entry is 1985–2003, DIRCE data go as far back as 1994. 500+ is the largest size category for which publicly available information exists.

IV construction

The goal is to construct an index that captures the variation in labor-demand composition a worker is exposed to in the time and place he is predicted to enter the labor market. That is, the index should capture variation in the degree to which labor demand for inexperienced workers is coming from large employers instead of small ones. This index will work as an IV, being used to predict the size of a worker's first employer.

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

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

Equation (3) illustrates the fact that I follow a leave-one-out approach. That is, if individual i got his first job in his predicted graduation year and in his region of birth, I exclude him from the calculation of $\bar{s}^{rec}_{-i}$.

3.3 IV Discussion

IV variation and the business cycle

I explicitly aim to partial out the effects of cyclical conditions at the beginning of the working life (von Wachter, 2020) from the effect of starting out at a larger or smaller employer. That is, the business cycle is a potential confounder of the first-employer size effect since it could impact both the size of a worker's first employer (Moscarini and Postel-Vinay, 2012) and also lifetime income through other channels. The empirical approach summarized in equation (2) is aimed at shutting down any impacts that business cycle conditions at entry might have on long-run prospects: by including region and cohort fixed effects, and explicitly controlling for a flexible function of the unemployment rate a worker faces during labor market entry. By flexibly controlling for the unemployment rate at the time of labor market entry, I try to replicate the thought experiment of comparing workers who were randomly assigned to firms of different sizes but shared common business cycle conditions.[27] In other words, my goal is to control for any *level* effects on labor demand resulting from large-firm hiring shocks, while using as IV the resulting variation in demand *composition*.

Figure 4 illustrates how cyclical conditions are held constant and the residual variation the IV approach uses. Panel (a) plots the correlation between the unemployment rate at entry and i) the IV $\bar{s}^{rec}_{-i}$ (*raw IV*), and ii) residuals from a regression of $\bar{s}^{rec}_{-i}$ 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).[28] After

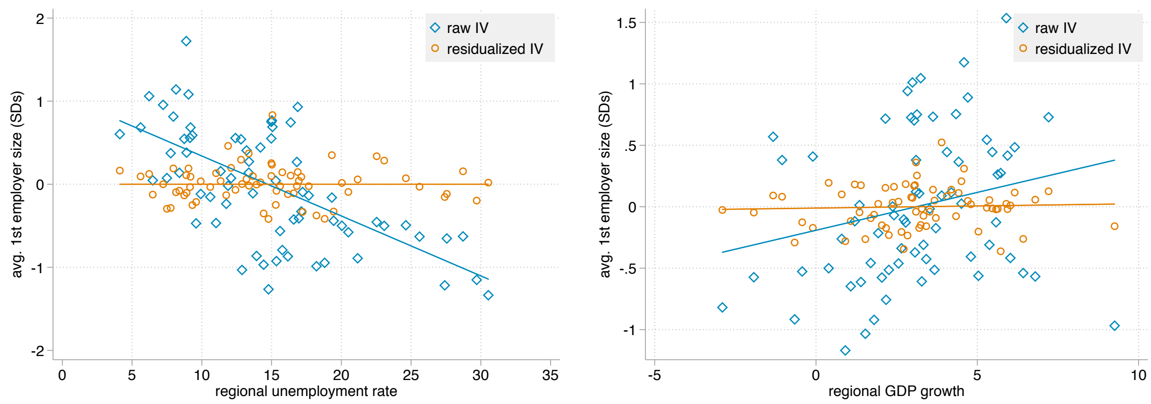
[27]The function $f()$ in equation (2) is allowed to differ across workers' education level since the *graduating-in-a-recession* literature finds heterogeneous impacts for workers of different skill levels. In robustness checks, I show that the main results are not sensitive to changing $f()$.

[28]The fact that young workers who enter the labor market during a recession start a smaller firms has been

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.[29] One could worry that using a single indicator might not perfectly capture cyclical variation. Panel (b) on Figure 4 allays these concerns using data on regional GDP growth rates. This additional cyclical indicator, since it is excluded from the specification in equation (2), is not mechanically unrelated to the IV residual variation. Reassuringly, a similar pattern emerges. There is a positive correlation between the instrument and regional GDP growth (blue diamonds). This again is consistent with large-firm hiring being procyclical. Controlling for fixed effects and the unemployment rate (orange circles) results in IV residual variation having a flat relationship with regional GDP growth. Overall, the residual IV variation identifying β seems orthogonal to business-cycle conditions.[30]

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

(a) Regional unemployment rate **(b)** Regional GDP growth



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

Instrument exclusion assumption

The instrument varies at the {region of birth $\times$ educational attainment $\times$ birth cohort}-level, except for the leave-one-out component, and follows the structure of the Bartik ap-

documented for Canadian college-educated workers in Oreopoulos et al. (2012) and for Austrian non-college workers in Brunner and Kuhn (2014).

[29]Online Appendix Figure A6 makes a similar point focusing on the time series variation of a large Spanish region.

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

proach discussed by Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2022). 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 hiring 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 get a DuPont job. Then, the 1993 Asturias high school cohort would be endogenously composed of more able young people (since in the absence of DuPont they would have attended college) *and* more likely to have a large first employer. This scenario would represent a violation of the exclusion assumption. Below I discuss in more detail the use of educational attainment in my empirical approach, and I find evidence lessening this type of concern.

IV and household characteristics at age 17

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

Educational attainment and potential endogenous responses

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

Certain features of my empirical approach somewhat relax these worries. As described above, business cycle conditions are kept constant. Given this approach, the educational response that would be worrying would come from responses to the large-firm hiring shocks captured by the instrumental variable $\bar{s}^{rec}_{-i}$, while holding business cycle conditions constant. Also note that an education enrollment response that is not followed by *completion* of the higher degree level would not be problematic for the exclusion restriction, it would simply reduce the relevance (predictive power) of the IV approach.

I test for endogenous education responses by checking whether, after controlling for unemployment rates, regional labor demand composition influences education attainment decisions. Section B.3 in online 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 education responses.

Autocorrelation of the instrument and persistent regional spillovers

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

These types of worries would be more pressing if the large-firm hiring shocks that the IV leverages were very persistent. The nature of the IV approach is to capture idiosyncratic hiring of large employers that are not sustained over time (such as plant openings, expansions, or hiring in batches). In line with this, a low autocorrelation of the IV residual variation would be desirable. Collapsing the data at the $\{rec\}$-level, the residual variation of the IV features an estimated autocorrelation equal to 0.15.[31] 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.

More on the variation underlying the IV and size as firm attribute of interest

The variation underlying my IV approach is tightly linked to firm *size* through the firm-size distribution; a similar IV approach would likely not be directly applicable for different firm attributes of interest. The key insight is that the power-law distribution of firm size implies that *individual* shocks to large firms do not average out (Gabaix, 2011). This is a statistical property of the firm-size distribution with roots that are well-grounded on economic mechanisms giving rise to power-law distributions (see Gabaix, 2009). What drives the IV variation in practice—based on the IV construction and on keeping macroeconomic conditions constant—is the aggregation of many idiosyncratic shocks to Spanish large firms which result in volatility across regions and years in the component of hiring that is driven by large firms. The DuPont plant opening case above is an example of one idiosyncratic hiring shock entering the aggregation. Other examples might include a firm experiencing a positive shock due to increased demand for their differentiated product, or a negative shock arising from a prolonged strike of a firm's workforce.

[31] 0.15 is the coefficient of a regression of the cell-level residualized IV on its one-year lag and a constant (N=610, robust standard error=0.049).

It would be illustrative to identify episodes of large-firm idiosyncratic shocks in the data, observe firms' identity, and investigate in local news or other sources the nature of the shock. This is in essence the "narrative" approach proposed by Gabaix (2011) to describe individual granular shocks. Unfortunately, I cannot follow this approach since I do not observe firm-level data, my worker-level data is a sample rather than the population, and its firm identifiers are anonymized. That is, while I argue that the large-firm hiring dynamics that construct my instrument are underpinned by granular shocks, my data is not well-suited to identify such shocks individually and "put a face" on them. However, this is conceptually similar to the standard treatment of the variation underlying the *graduating-in-a-recession* literature. In those studies, we take fluctuations in macroeconomic conditions as a given and choose some unemployment rate metric that summarizes them properly. This literature abstracts also from what are the root causes of the macroeconomic shocks that drive variation in unemployment rates.

Lastly, the nature of the IV implies that, even if aggregate conditions are kept constant, it could be that compliers find first jobs in firms that are large *and* disproportionately likely to be doing well. To the extent that size correlates with positive first-employer attributes, one could think of business conditions as being part of the bundle of employer qualities that the treatment captures.

3.4 Lifetime Income: Results

Table 3 shows OLS, first stage, and IV-TSLS results of estimating β in equation (2) using the proposed instrumental-variables approach (online Appendix Table A2 shows reduced-form estimates). 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 education level. I cluster standard errors at the {region of birth $\times$ educational attainment $\times$ birth cohort}-level since this is the level through which the IV operates (Abadie et al., 2017). Column (5) shows first-stage results. The instrument does a good job at predicting first job size, with an excluded instrument F-statistic of 24.3.[32] Columns (1) and (6) show, respectively, the OLS and IV elasticities of lifetime income with respect to first-employer size. The OLS elasticity estimate is .028. The IV-TSLS estimate is significantly larger and equal to .117. This elasticity implies that, at least for the relevant compliers, matching with a 10% larger first-employer leads to 1.17% higher lifetime income. Another way of interpreting the estimated magnitude is using the standard deviation of log first employer size, which is equal to 2.1. We can expect that matching with a first-employer that is larger by one standard deviation in log size to increase lifetime income by 27.7%.[33]

[32]Online Appendix Figure A7 displays the IV residual variation. Online Appendix Figure A8 provides graphical evidence of the IV result, showing first- and second-stage TSLS variation.

[33]Note that $100 \cdot [exp(2.1 \times .1166) - 1] = 27.74$

Table 3: Career outcomes and first-employer size: OLS and IV-TSLS estimates

	OLS				First Stage	IV-TSLS			
	lifetime income (1)	lifetime earnings (2)	average daily wage (3)	days worked (4)	first employer size (5)	lifetime income (6)	lifetime earnings (7)	average daily wage (8)	days worked (9)
first employer size	0.0276***	0.0280***	0.0311***	-0.0031***		0.1166**	0.1102**	0.0820**	0.0281
	(0.0011)	(0.0011)	(0.0009)	(0.0005)		(0.0481)	(0.0493)	(0.0402)	(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	79,941	79,941	79,941	79,941	79,941	79,941	79,941	79,941	79,941

Notes: All variables enter regressions in logs. 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. Regressions at the worker level. All regressions control for an education-specific quartic 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.

3.5 Comparison of OLS and IV results

The IV estimate is about four times larger than the OLS, which is in line with existing related evidence.[34] While this comparison might not align with what a simple unobserved ability bias would predict, the IV-OLS difference is consistent with the first-employer size effect being heterogeneous across workers, and it being larger for those whose first-employer match is more susceptible to the labor demand IV. That is, suppose that some people benefit more than others from having their first job at a larger firm. Suppose as well that those who benefit the most tend to get a first job at a large firm if there is idiosyncratically high large-firm hiring in their birth region, but not otherwise. Then, a LATE interpretation (Imbens and Angrist, 1994) would explain the relatively-high IV magnitude. I now provide evidence that is consistent with this scenario.

The first thing to ask is who, given the nature of the IV, might the likely compliers be. First, note that the geographic dimension of the instrument works through region of *birth*. The minority of people who migrate across regions for their first job will be less likely to be compliers.[35] More generally, highly motivated individuals will be more likely to do their best to match with their preferred type of employer under all scenarios of labor demand composition. Compliers, those who only match with large firms in years of differentially high large-firm hiring, might thus be of initially lower ability. This could arise as a supply-side effect if lower ability young adults take a more passive approach towards job search. It could also be a demand-side effect if large firms are able to screen job applicants and hire in order of perceived ability. In both cases, the marginal large-firm hires will be less able.

I explore more formally the possibility that the less able or less knowledgeable young workers comprise the group of IV compliers (online Appendix C provides details on the analytical and estimation framework I follow; here, I summarize the main takeaways). Building on results by Angrist and Imbens (1995), I estimate a flexible first stage using a distribution regression framework (Chernozhukov et al., 2013) that permits characterizing which parts of the firm-size distribution and which type of workers are driving the IV two-stage least square (TSLS) estimate. The intuition behind the Angrist and Imbens (1995) result is that, with a multivalued treatment and a multivalued instrument, the TSLS estimate can be written as a weighted average of causal responses to a unit change in treatment along the treatment and instrument distributions for the relevant compliers. I develop an approach to estimate such weights, allow them to vary across different groups of people. I focus on workers' education and place of birth as observable variables that might be related to complier status.

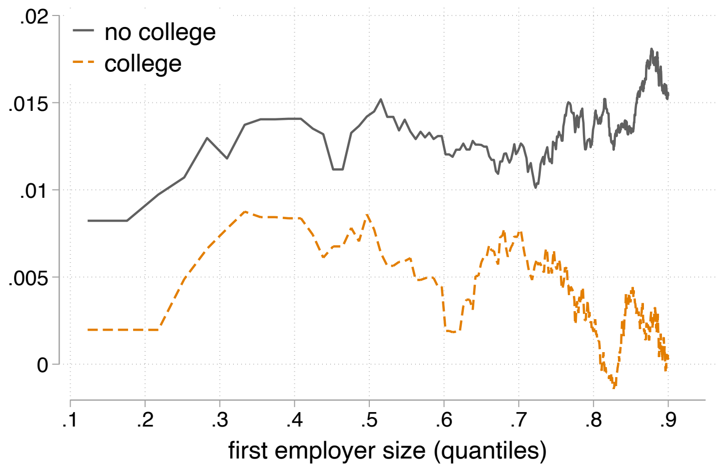
I find that non-college and rural-born workers likely play an important role in the complier group. Figure 5 illustrates this fact for non-college workers, plotting "complier weights" separately based on education. Weights for non-college workers are greater throughout the first-employer size distribution, implying that their initial job is always more sensi-

[34] Kahn (2010) finds IV effects of graduating in a recession that are 4.5 times larger than the OLS. She puts forward the explanation that compliers are the ones who do not re-optimize the place and time of entry in response to labor-market conditions.

[35] Table 1 shows that 87% of young workers in my sample held their first job in their region of birth.

tive to the variation of the IV.[36] This is consistent with Finamor (2022) who finds evidence that young people with lower earnings potential are less likely to re-optimize to labor market condition at the time of expected graduation and thus behave more as "compliers" with respect to this type of instrument based on region of birth and expected graduation year.

Figure 5: Compliers have lower earnings potential: weight function from flexible 1st stage



Notes: This figure plots the estimated differential impact of the instrumental variable on workers' first employer size, heterogeneously by workers' place of birth and education, and for different quantiles of the first employer size distribution. Estimates of the weight function from equation (C11) in online Appendix C, for different subgroups and overall, as a function of first employer size, and holding the instrument constant in the 95th percentile.

Overall, the evidence suggests that those with lower earnings potential are more influenced by the variation in demand composition that my IV uses. Additionally, there is evidence to indicate that these workers might benefit the most from large-firm employment (see results on the cross-sectional firm-size premium in Table 2 and the heterogeneous elasticity of lifetime income in Online Appendix Figure C3).[37] As such, a channel that is consistent with the evidence and explains the larger magnitude of the IV estimates is younger and less knowledgeable workers having the highest long-term benefits of a larger first employer. This could arise due to worse outside options, having a harder time moving away from a bad first job, or benefiting the most from large firms' on-the-job skill development opportunities.[38]

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

[36] Online Appendix Figure C2 additionally shows an equivalent figure and discussion for urban- vs. rural-born workers.

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

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

3.6 Lifetime Income: Robustness and Extensions

Robustness. The IV elasticity of lifetime income with respect to first-employer size is robust to a variety alternative specifications. Section B.4 in online Appendix B shows that the results are stable when accounting for first-employer industry, discounting the stream of income in lifetime income measures, controlling for regional unemployment differently, controlling for regional GDP growth, controlling for unemployment rates in previous years or during the year of labor market entry in addition to unemployment during predicted graduation, controlling for unemployment rates in years following predicted graduation up until age 35, measuring employer size as an average over years prior to labor market entry, using birth-province fixed effects, controlling for population size in the first-employer's province, and including region-specific time trends. Additionally, Section B.5 shows that the first-employer size effect is robust to using uncensored measures of income constructed from tax records.

3.7 Wages, Employment, Earnings, and Job Security

Wages, employment, and earnings. The lifetime-income effect of matching with a larger first employer could combine effects on different margins: quantity of work, average wages, and unemployment benefits. Here, I decompose the lifetime income effect into its different components. I estimate the elasticity β from equation (2) replacing lifetime income y_i with three different outcomes (in logs): average daily wages, total days worked, and lifetime earnings (which differ from lifetime income in that they do not include unemployment benefits). Table 3 shows OLS, first stage, and IV results from this exercise. Focusing on the IV estimates, column (7) shows that the elasticity of lifetime earnings is equal to .110, which is 94% of the elasticity of lifetime income equal to .117. 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%).

Fixed-term contracts. Spain features a "dual" labor market, with a stark difference between permanent and temporary (fixed-term) labor contracts (see Dolado et al., 2002). Table 2 shows a slight negative relationship between firm size and use of temporary contracts. If large first employers are more likely to offer permanent contracts, this should be interpreted as part of the bundle of positive features that are associated with greater first-employer size. Unfortunately, the data only starts recording contract type accurately in 1998, which is too late in time to analyze first-job contracts for most cohorts in my sample. However, I can study contracts at later ages. In online Appendix B, Section B.6, I study, as outcomes, measures of job security between ages 30–35 using contract type data. With the same IV approach as above, I find that, between ages 30–35, a larger first employer results in a lower probability of ever having a temporary contract, and a greater (although not statistically significant) probability of experiencing "total job security"—a composite index combining information on contracts and employment.

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

I quantify the relationship between the effect of starting one's career at a larger or smaller firm and the effect of entering the labor market during a recession, which has been the focus of previous work (e.g. Kahn, 2010; Oreopoulos et al., 2012). The persistent positive effects of starting at a large firm could partly explain the findings of this literature.[39]

I begin by estimating the following regression in my sample:

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

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

Table 4: First-employer size and unemployment rate at entry

	first employer size		
	(1)	(2)	(3)
unemployment rate at entry	-0.0099***	-0.0117***	-0.0166***
	(0.0035)	(0.0044)	(0.0050)
SE Clusters	661	442	364
Sample	All	HS & Voc.	Less urban HS & Voc.
Observations	79,941	66,998	29,724

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

Table 4 shows OLS estimates of γ for the whole sample and different subgroups. The negative estimates are consistent with previous literature and the evidence found on Figure 4. The estimate for the full sample in column (1), equal to -.0099, is very similar to that found in Oreopoulos et al. (2012). In column (2) I estimate γ for the subsample of workers without a college degree. The estimated coefficient is equal to -.0117, which is somewhat larger than for the whole sample. This suggests that, for this group of less educated workers, the size of their first employer is more sensitive to the cyclical conditions at the time of entry. Column (3) focuses on the subgroup of non-college workers who were born in less urban provinces of Spain. As discussed in Section 3.5, this group of workers are likely to be compliers in my IV approach and thus mostly driving the first-employer size causal effects. The estimate for this subgroup is even larger, equal to -.0166.

Next, I combine the estimates of γ with (i) the elasticity between lifetime income and first-employer size, and (ii) results from Fernández-Kranz and Rodríguez-Planas (2018),

[39] In an exercise similar in spirit but not focusing on entrants, Haltiwanger et al. (2018) decompose the cyclicality of job-to-job moves across the firm-wage ladder into i) the cyclicality of moves, and ii) the cyclicality of moving up conditional on moving. They then quantify the resulting implications for earnings growth.

who estimate the *graduating-in-a-recession* effect in Spain. Using these results, Table 5 shows that between 7% and 15% of the losses from entering the labor market during a recession could be explained by the fact that during bad economic times young people are more likely to match with a smaller first employer. For non-college workers from less urban provinces, this fraction is between 12% and 15%.

Table 5: Benchmark: First-employer size effect explaining entering-in-a-recession effect

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

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

4 Persistence and Mechanisms

Why does the first-employer size effect arise? I first show these effects are persistent—that is, not solely mediated by the time a worker spends at his first job. This persistence implies workers' trajectories in subsequent jobs are affected by initial matches. I show an example of such dynamics focusing on firm size of *subsequent* employers as a function of first-employer size. I then show evidence consistent with a human capital channel.[40]

4.1 Persistence

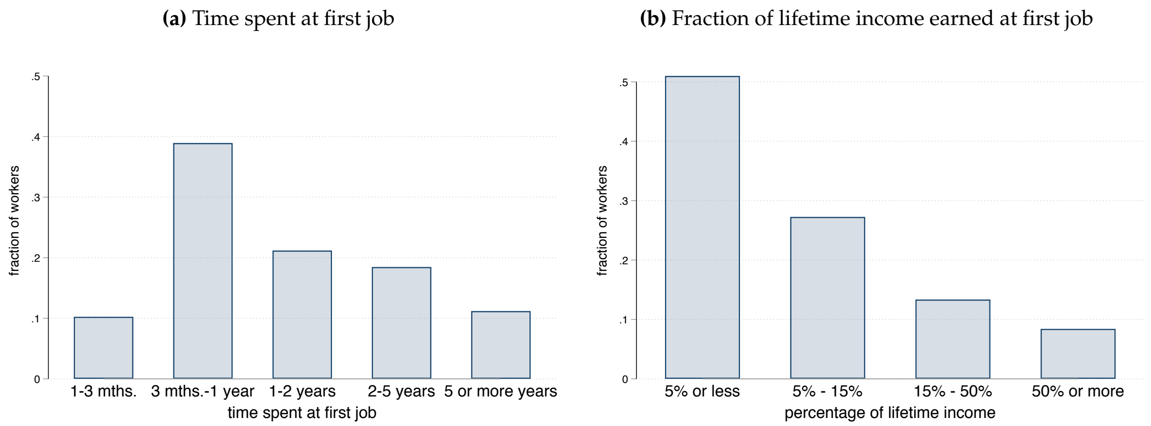
Descriptive evidence

Descriptive patterns in the data are consistent with persistent effects. First, young workers are very mobile. Figure 6 shows that most workers do not stay at their first employer for long. 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 career paths (see Figure 2).

The combination of high mobility and earnings growth implies that only a small fraction of lifetime income is directly earned at the first job. Figure 6 shows that this share is rather

[40] An inherent caveat in the persistence analysis is lack of exogenous variation in the employment dynamics following a first-employer match. Ham and LaLonde (1996) discuss the issues arising when researchers have at their disposal exogenous variation in some initial treatment, but no exogenous variation driving the subsequent employment dynamics. Analyses conditioning on employment patterns (e.g. time spent at first employer, unemployment spells) are more descriptive in nature than the lifetime analyses in Section 3.

Figure 6: 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.

low for most workers. Income earned at the first job represents 5% or less of lifetime income for half of the workers in my sample. This number is 5–15% for 28% of workers, 15–50% for 12% of workers, and 50% or more for less than 10% of the workers. These numbers are very similar for workers I previously identified as "likely compliers"—non-college, and born in less urban provinces (see online Appendix Figure A9).

Income at age 35

I directly test for persistence estimating a version of equation (2) in which the outcome variable is income earned during the calendar year a worker turns 35.[41] Table 6 shows the estimated elasticity, which is around 0.09. The estimated elasticity is unchanged when using uncensored tax earnings at age 35 (see online Appendix Table B3).

This result is evidence of persistence at subsequent jobs since at this age only 6.8% of people in my sample are working at their first employer. Moreover, this is the last year of income that enters the lifetime income measure. Previous evidence indicates that the majority of earnings growth occurs in the first ten years of the working life (Topel and Ward, 1992; Rubinstein and Weiss, 2006).[42] Thus, first-employer effects at age 35 (when the average person in the sample has been in the labor market for 15 years) suggest permanent effects past the actual years I consider in my lifetime income measure.

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

[42] I provide related evidence for Spain in online Appendix Figure A2.

Table 6: Income during age 35 and first-employer size: OLS and IV-TSLS estimates

	OLS	First Stage	IV-TSLS
	annual income age 35 (1)	first employer size (2)	annual income age 35 (3)
first employer size	0.0368*** (0.0015)		0.0894* (0.0538)
labor demand instrument		0.1010*** (0.0188)	
F-stat excl. instr.		28.89	
SE Clusters	661	661	661
Observations	70,588	70,588	70,588

Notes: All variables enter regressions in logs. OLS and IV-TSLS estimates of the elasticity of annual income during age 35 with respect to first-employer size. Dependent variable is 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. Regressions at the worker level. All regressions control for an education-specific quartic function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Firm size of subsequent employers

I find that starting at a larger employer leads to employment at larger subsequent employers, which could explain part of the persistent earnings effects. I estimate equation (2) using (log) size of a worker's second employer and (log) size of his employer at age 35 as dependent variables. Results are in Table 7. The IV elasticities between first employer size and that of subsequent employers are between 0.36 and 0.37.

Table 7: Subsequent employers and first-employer size: OLS and IV-TSLS estimates

	OLS	First Stage	IV-TSLS	OLS	First Stage	IV-TSLS
	second employer size (1)	first employer size (2)	second employer size (3)	employer size age 35 (4)	first employer size (5)	employer size age 35 (6)
first employer size	0.3232*** (0.0048)		0.3610** (0.1513)	0.2582*** (0.0063)		0.3745** (0.1557)
labor demand instrument		0.0999*** (0.0198)			0.0954*** (0.0198)	
F-stat excl. instr.		25.46			23.3	
SE Clusters	661	661	661	661	661	661
Observations	72,742	72,742	72,742	65,217	65,217	65,217

Notes: All variables enter regressions in logs. 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 an education-specific quartic function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Such persistence in ensuing employers' characteristics could be driven by a human capital accumulation (e.g. skills developed at large employers being more valuable at other large employers), but also by job ladder effects driven by search frictions.[43] Next, I show evidence that is consistent with a human capital channel, but harder to rationalize through

[43] A stylized job ladder framework in online Appendix B, Section B.7, illustrates this point. Section B.8 provides further evidence on the relationship between first-employer size and various measures of career dynamics.

search frictions.

4.2 Evidence of a Human Capital Mechanism

Better skill-development opportunities at larger firms could lead to persistent first-employer size effects, over and above any persistence potentially driven by search frictions.[44] How can we tell whether human capital acquisition contributes to the first-employer size effect? The key insight, present in models of on-the-job search, is that an involuntary unemployment spell cuts a job ladder progression. This is because an unemployed worker looking for a job does not have a current employer as an option to weigh against new offers. In this sense, this brings him to the "bottom" of the ladder.

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

I estimate the elasticity of different long-term outcomes with respect to first-employer size in the subsample of those experiencing unemployment between their first and second jobs. The IV results of equation (2) can be found in Table 8 (OLS estimates are in online Appendix Table A3). 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. Elasticities of comparable magnitudes to baseline also arise for average daily wages, lifetime earnings, subsequent employers' size, or income at age 35. The latter is noisily estimated but similar to baseline.

[44] See Rosen (1972), Gregory (2019), Arellano-Bover (2022), and Arellano-Bover and Saltiel (2021) for work on heterogeneous human capital accumulation across firms.

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

Table 8: Career outcomes and first-employer size: 1st–2nd job unemployment gap sample

	First Stage	IV-TSLS						
	first employer size (1)	lifetime income (2)	average daily wage (3)	lifetime earnings (4)	days worked (5)	second employer size (6)	annual income age 35 (7)	employer size age 35 (8)
first employer size		0.0900**	0.0794**	0.0832**	0.0038	0.5374***	0.0882	0.4298**
		(0.0403)	(0.0340)	(0.0422)	(0.0189)	(0.1928)	(0.0599)	(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	34,507	34,507	34,507	34,507	34,507	32,965	30,193	27,881

Notes: All variables enter regressions in logs. 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 employment gaps that are at least 2 months long. Regressions at the worker level. All regressions control for an education-specific quartic function of regional unemployment at predicted graduation year, region-of-birth fixed effects, three educational attainment levels fixed effects, and birth-cohort fixed effects. 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.

Further evidence on human capital.

Overall, the evidence is consistent with young workers acquiring more valuable skills at large firms. This aligns with evidence using panel data on cognitive skills (Arellano-Bover, 2022) and earnings growth (Arellano-Bover and Saltiel, 2021).[46] Here, I briefly describe additional pieces of evidence consistent with a human capital mechanism.

Section B.9 in online Appendix B includes additional evidence of persistent effects consistent with human capital. I estimate a time-varying version of the elasticity of lifetime income with respect to first-employer size and find an increasing first-employer size effect, meaning that a larger first employer results in higher earnings growth. I then show that workers with larger first employers experience greater wage growth when moving to their second job, holding constant first job tenure and second employer size. These results align with human capital but not search, since starting from a from a higher step in the job ladder would predict smaller wage growth at future job changes.

With a human capital mechanism, we might not expect persistent effects arising from human capital for those spending a very short amount of time at their first job. Online Appendix Figure A10 shows evidence consistent with this.

Online Appendix E provides further evidence on a human capital channel using the employment histories data in their panel form and measuring actual experience acquired in firms of different sizes. I find that, keeping constant contemporaneous employer characteristics, returns to past experience acquired in large firms are substantially greater than returns to experience acquired in small firms, and that this differential is more important the younger a worker is.

5 Conclusion

This paper sheds new light on how firm heterogeneity affects workers' prosperity in the long term. My findings imply that the identity of the firm where a young person lands their first job can have effects that last throughout one's career, and that firm size is a relevant measure that is correlated with meaningful employer characteristics. Compared to other firm attributes, size has the advantage that is easily observable to workers and policymakers and that, not being model-based, is transparently measured.

I am able to identify a causal link between first-employer size and long-term outcomes because, in spite of the importance of a first-job match, there is some randomness involved in the matching process. My IV approach leverages the part of this randomness that is driven by idiosyncratic hiring shocks of large firms. An inherent feature of the IV approach is that it estimates causal effects for workers whose first-employer match is most affected by idiosyncratic large-firm hiring shocks in their region of birth; that is, marginal large-firm hires. I find that these entrants are younger and with lower earnings potential. These "compliers" seem to disproportionately benefit from starting out at a large first employer. An

[46] Müller and Neubaeumer (2018) argue that training at a larger firm leads to lower unemployment later on. In a context of imperfect competition, if workers value learning, large firms could be large in part *as a result* of better learning opportunities (Card et al., 2018).

interesting question outside the scope of this paper is to understand the characteristics of an equilibrium in which the marginal large-firm worker has the largest long-term benefits from such a match. It could be that any human-capital benefits a worker derives are proportional to the costs she generates for the firm (through training, or the time it takes to learn the ropes of the job). Firms might not want to, or not be able to, discriminate wages based on these costs and benefits. Even if firms did lower wages to equalize long-term benefits, there are reasons why young workers could turn down such a deal (e.g. unawareness of long-term benefits, or liquidity and credit constraints).

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

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

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

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge (2017). When should you adjust standard errors for clustering? Working Paper 24003, National Bureau of Economic Research.

Abowd, John M, Francis Kramarz, and David N Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Abowd, John M, Kevin L McKinney, and Nellie 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.

Albæk, Karsten, Mahmood Arai, Rita Asplund, Erling Barth, and Erik Strøjer Madsen (1998). Measuring wage effects of plant size. *Labour Economics 5*(4), 425–448.

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

Angrist, Joshua D. and Guido 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.

Arellano-Bover, Jaime (2021). Who gets their first job at a large firm? The distinct roles of education and skills. *AEA Papers and Proceedings 111*, 465–69.

Arellano-Bover, Jaime (2022). The Effect of Labor Market Conditions at Entry on Workers' Long-Term Skills. *The Review of Economics and Statistics 104*(5), 1028–1045.

Arellano-Bover, Jaime and Fernando Saltiel (2021). Differences in on-the-job learning across firms. Discussion Paper 14473, IZA Institute of Labor Economics.

Arora, Ashna, Leonard Goff, and Jonas Hjort (2021). Pure-chance jobs versus a labor market: The impact on careers of a random serial dictatorship for first job seekers. *AEA Papers and Proceedings 111*, 470–75.

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

Bloom, Nicholas, Fatih Guvenen, Benjamin S. Smith, Jae Song, and Till von Wachter (2018). The disappearing large-firm wage premium. *AEA Papers and Proceedings 108*, 317–22.

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

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

Bonhomme, Stéphane and Laura Hospido (2017). The cycle of earnings inequality: Evidence from Spanish social security data. *The Economic Journal 127*(603), 1244–1278.

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

Borusyak, Kirill, Peter Hull, and Xavier Jaravel (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies 89*(1), 181–213.

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

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

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(S1), S13–S70.

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

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

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

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

Colonnelli, Emanuele, Joacim Tåg, Michael Webb, and Stefanie Wolter (2018). A cross-country comparison of dynamics in the large firm wage premium. *AEA Papers and Proceedings 108*, 323–27.

Di Addario, Sabrina, Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten (2021). It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages. Discussion Paper 14446, IZA Institute of Labor Economics.

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

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

Fackler, Daniel, Steffen Müller, and Jens Stegmaier (2021). Explaining wage losses after job displacement: Employer size and lost firm wage premiums. *Journal of the European Economic Association 19*(5), 2695–2736.

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

Finamor, Lucas (2022). Labor market conditions and college graduation. Mimeo, Yale University.

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

Gabaix, Xavier (2009). Power laws in economics and finance. *Annual Review of Economics 1*(1), 255–294.

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

García Pérez, José Ignacio (2008). La muestra continua de vidas laborales: Una guía de uso para el análisis de transiciones. *Revista de Economía Aplicada 16*(1), 5–28.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

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

Gregory, Victoria (2019). Firms as learning environments: Implications for earnings dynamics and job search. Mimeo, New York University.

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

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

Guvenen, Fatih, Greg Kaplan, Jae Song, and Justin Weidner (2022). Lifetime earnings in the United States over six decades. *American Economic Journal: Applied Economics 14*(4), 446–79.

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

Haltiwanger, John C, Henry R Hyatt, Lisa B Kahn, and Erika McEntarfer (2018). Cyclical job ladders by firm size and firm wage. *American Economic Journal: Macroeconomics 10*(2), 52–85.

Ham, John C and Robert 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, Todd L. and Daniel J. Feaster (1990). A selectivity model of employer-size wage differentials. *Journal of Labor Economics 8*(1, Part 1), 99–122.

Imbens, Guido W and Joshua 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.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan (1993). Earnings losses of displaced workers. *The American Economic Review 83*(4), 685–709.

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

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten (2020). Leave-out estimation of variance components. *Econometrica 88*(5), 1859–1898.

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

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

Liu, Kai, Kjell G. Salvanes, and Erik 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, Robert E (1978). On the size distribution of business firms. *The Bell Journal of Economics 9*(2), 508–523.

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

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

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

Moscarini, Giuseppe and Fabien 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, Steffen and Renate 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.

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

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

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

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

Petrongolo, Barbara and Maria 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, Sherwin (1972). Learning and experience in the labor market. *Journal of Human Resources 7*(3), 326–342.

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

Schoar, Antoinette and Luo 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, Hannes and Till von Wachter (2019). Unlucky cohorts: Estimating the long-term effects of entering the labor market in a recession in large cross-sectional data sets. *Journal of Labor Economics 37*(S1), S161–S198.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter (2019). Firming up inequality. *The Quarterly Journal of Economics 134*(1), 1–50.

Sorenson, Olav, Michael S. Dahl, Rodrigo Canales, and M. Diane Burton (2021). Do startup employees earn more in the long run? *Organization Science 32*(3), 587–604.

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

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

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

von Wachter, Till (2020). The persistent effects of initial labor market conditions for young adults and their sources. *Journal of Economic Perspectives 34*(4), 168–94.

von Wachter, Till and Stefan 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.

Wee, Shu Lin (2016). Delayed learning and human capital accumulation: The cost of entering the job market during a recession. Mimeo, Carnegie Mellon University.