Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

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

Jaime Arellano-Bover

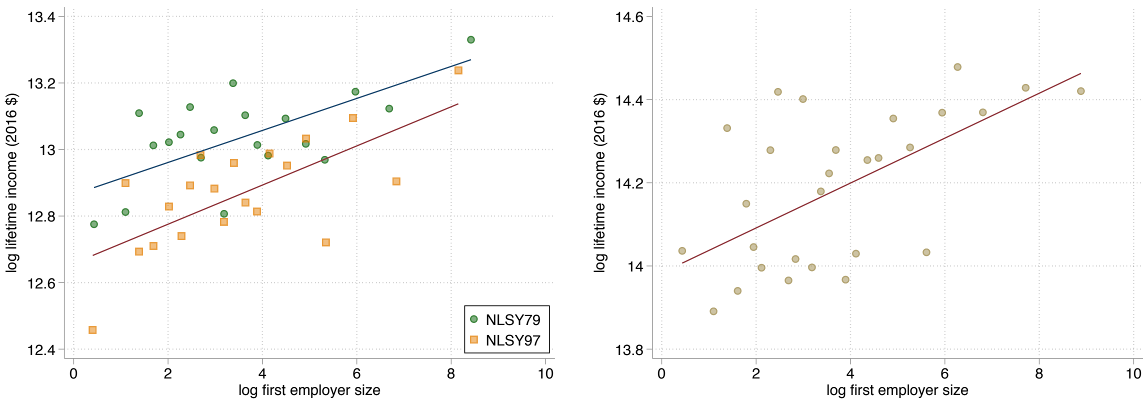
Online Appendices

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

A Additional Figures and Tables

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

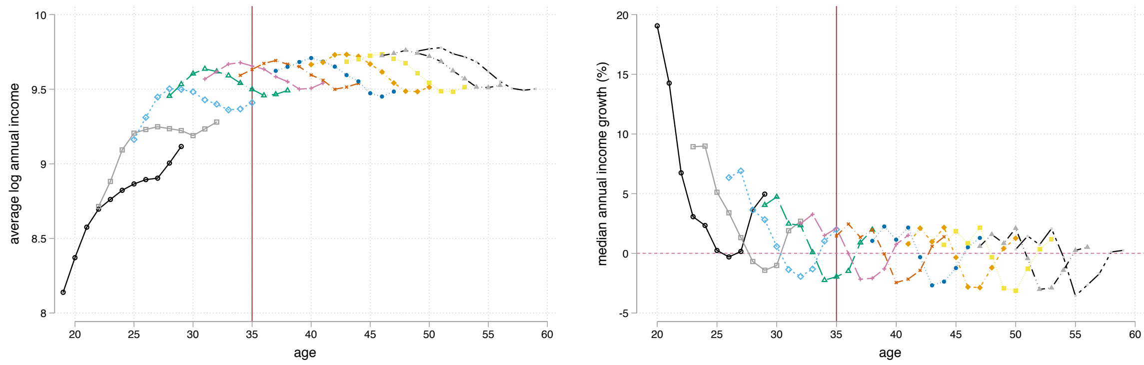
(a) NLSY79, NLSY97: Cumulative income until age 34 **(b)** NLSY79: Cumulative income until age 50



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

Figure A2: Income stabilizes by age 35: Annual income and growth age profiles (2006–2015)

(a) Log income age profile **(b)** Income growth age profile

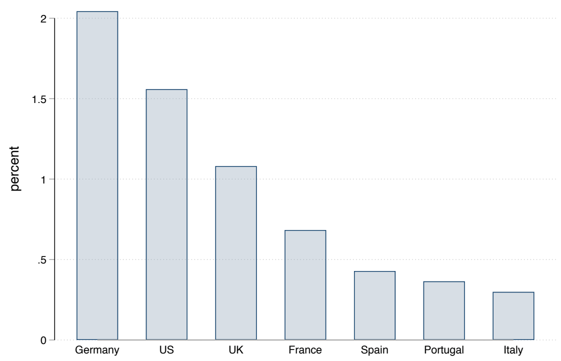


Notes: Age profiles for different cohorts in log annual income and annual income growth rates. Left panel: average log annual income by cohort and year. Right panel: median annual income growth rate by cohort and year. Growth rate g_t between annual income Y_{t-1} and Y_t computed as $100 \times \frac{Y_t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

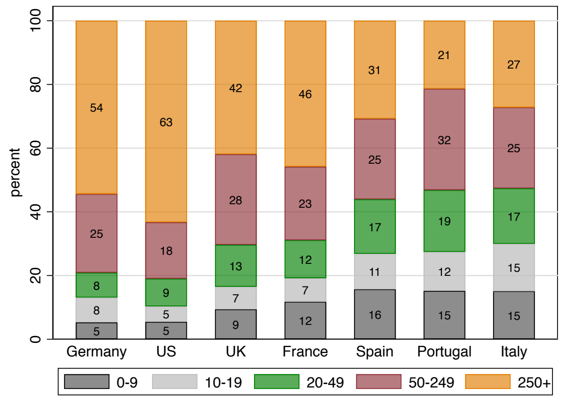
Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

Figure A3: Relatively few large firms in Spain: Firm size across countries

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

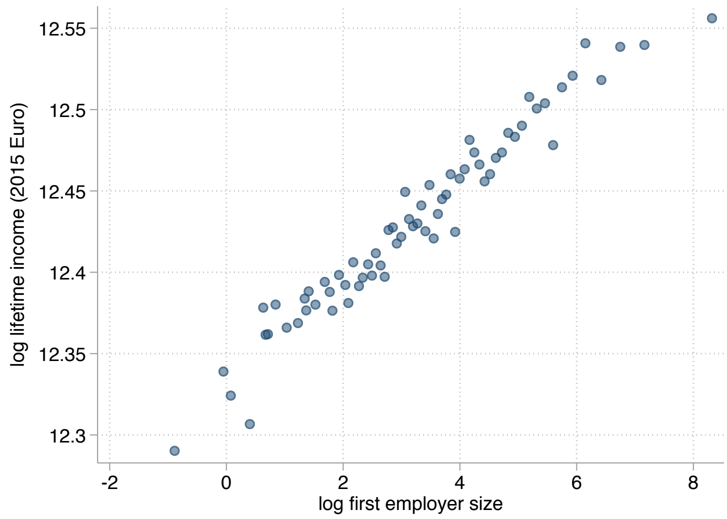


(b) Employees by business size (manufacturing)



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

Figure A4: Lifetime income and first-employer size correlation: controlling for sector

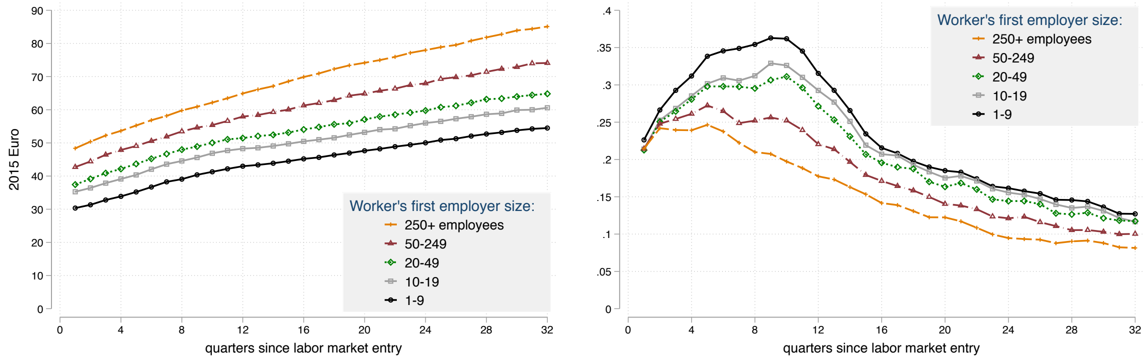


Notes: Conditional expectation of lifetime income as a function of first-employer size, adjusting for sector of first employer. Binned scatterplot. Log lifetime income (as defined in the text) on the vertical axis. Log size of worker's first employer on the horizontal axis. Both variables net out of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

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

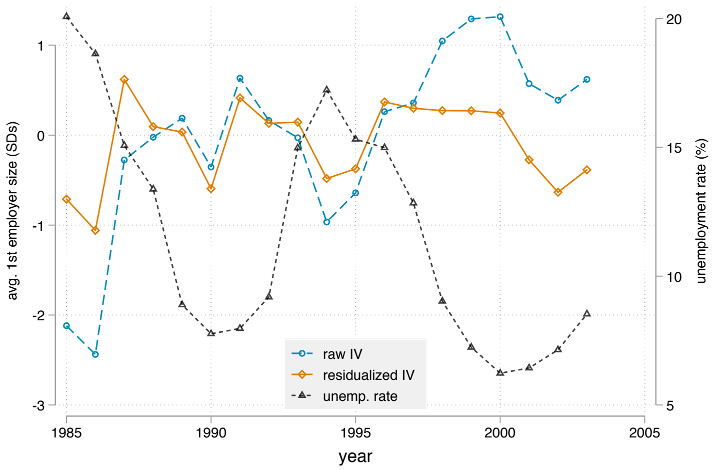
Figure A5: Daily wages and unemployment trajectories by first-employer size category

(a) Average daily wages **(b)** Fraction experiencing unemployment



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

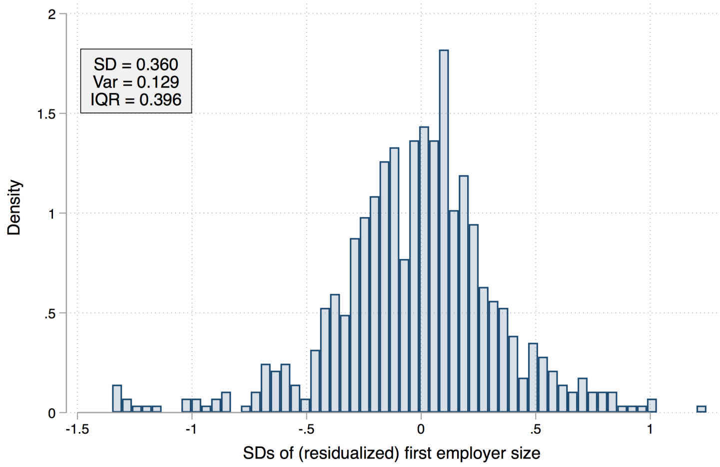
Figure A6: IV residual variation and business cycle variation in Catalunya region



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

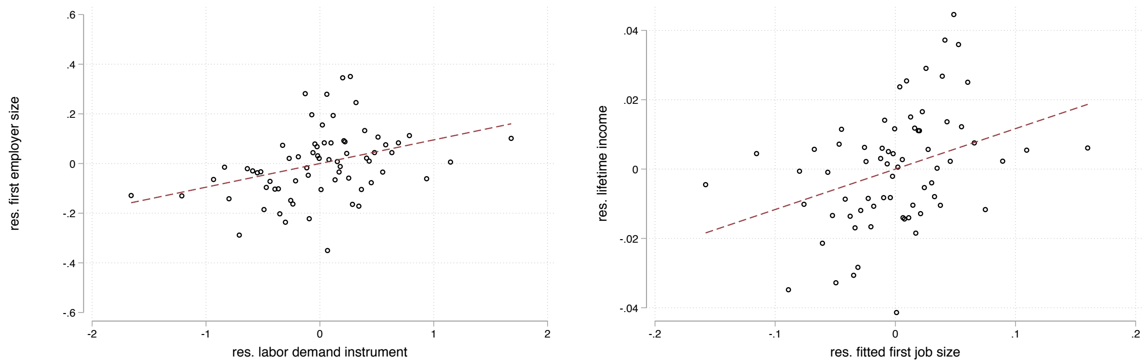
Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

Figure A7: Labor demand instrument: residual variation



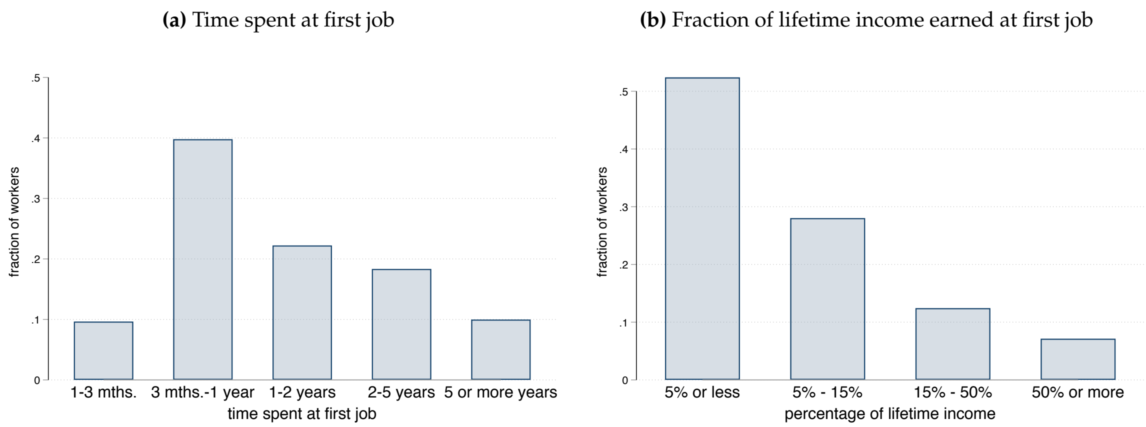
Notes: Histogram of (residualized) labor demand instrument across region of birth × education × year of birth bins. Expressed in units of standard deviations of (residualized) first employer size. In both cases residuals from a regression on a flexible function of unemployment rate at predicted graduation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

Figure A8: IV-TSLS elasticity of lifetime income w.r.t. first-employer size



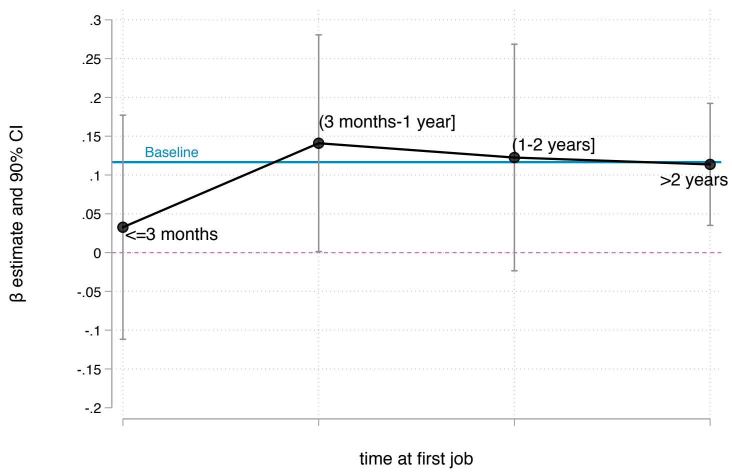
Notes: Binned scatterplots of first stage and second stage residual variation from equation (2) in the text, instrumenting for log first job size $s_{J(i)}$ using the instrument $\bar{s}^{rec}_{-i}$ described in the text. The outcome variable is log total income after first job semester (described in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

Figure A9: Time spent and income earned at the first job: Subsample of "likely compliers"



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

Figure A10: IV elasticity by time spent at first job



Notes: Elasticity of lifetime income with respect to first-employer size (TSLS estimates of equation (2)), estimated separately for four groups of workers based on the time spent at the first employer. Group ≤ 3 months: N=7,455. Group (3 months–1 year]: N=29,405. Group (1–2 years]: N=18,138. Group >2 years: N=24,943. Sample of male workers of all education levels, born in Spain between 1968–1980.

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

Table A1: IV residual variation uncorrelated with the business cycle

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

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

Table A2: Career outcomes: Reduced-form estimates

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

Notes: All variables enter regressions in logs. OLS estimates of the elasticity of lifetime income and other outcomes with respect to the labor-demand composition instrument. Columns (1)–(4): Lifetime income defined as sum of total income after first job semester (defined in text) until age 35, using 0, 1, 2, and 3 percent annual discounting. Column (5): Average daily wage defined as sum of total income over total days worked after first job semester (defined in text) until age 35. Column (6): Total days worked after first job semester (defined in text) until age 35. Column (7): Lifetime earnings defined as sum of total earnings after first job semester (defined in text) until age 35, using 0 percent annual discounting. 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.

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

Table A3: Career outcomes and first-employer size: 1st-2nd job unemployment gap sample, OLS estimates

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

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

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

B Additional Results, Extensions, and Robustness Tests

B.1 Additional data sources

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

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

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

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

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

Finally, I use the NLSY79 and NLSY97 surveys to document the correlation between first employer size and lifetime income in the US.

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

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

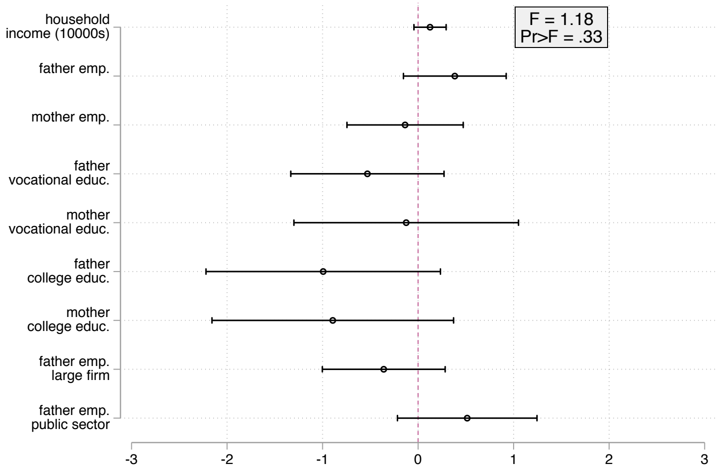
Using the EU Households Panel (*Panel de Hogares de la UE*, or PHOGUE) allows me to observe the relevant information for four birth cohorts from my sample (1977–1980). Collapsing the data to the $\{rec\}$ cell level I estimate the following regression:

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

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

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

Figure B1: IV specification check: Instrumental Variable and Cohort Household Characteristics



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

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

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

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

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

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

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

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

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

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

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

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

B.4 Lifetime income IV result: Additional robustness tests

In this section I show that the IV estimate of the elasticity of lifetime income with respect to first-employer size, discussed in Section 3.4, is robust. Figure B2 gathers the resulting IV elasticity estimates when using alternative specifications. Additionally, it shows results when discounting the measure of lifetime income (baseline estimates use the measure with no discounting). The black and round marker shows the baseline results from column (6) in Table 3.

Alternative flexible unemployment rate function. I change the way that I control for the regional unemployment rate during predicted graduation year. Baseline results use an education-specific quartic function. The white markers in Figure B2 show the estimates when I control for unemployment rate using an education-specific categorical piece-wise function: I bin the unemployment rate into 3 categories (low, medium, high), include fixed effects for each of these categories, and allow the fixed effects to vary by educational attainment. The cutoffs for the three categories are 11% and 16% and are based on the worker-level distribution of regional unemployment rates at the time of graduation, roughly dividing workers equally between the three categories. The estimates are very similar to baseline.

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

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

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

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

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

Business cycle conditions until age 35. I estimate an expanded specification which, in addition to controlling for unemployment rates during the year of predicted graduation, also controls for unemployment rates in subsequent years, up until age 35 (i.e., the last year entering the lifetime income measure). This robustness check is meant to allay potential concerns related to persistent regional spillovers. The red markers in Figure B2 show the elasticity estimates using this specification, which are very similar to baseline.

Large employers and growing employers. Measuring employer size at the time the worker joins the firm could conflate having a first job at a large employer with having a first job at an employer which is doing well and growing in size. To address this distinction I estimate an alternative specification using a different measure of first-employer size. Instead of using employer size at the time of joining the firm, I use an average over the four years prior to the year the worker joined.[2] The orange markers in Figure B2 show the estimates using this measure. These elasticities are also very similar to baseline.

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

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

Provincial size of first job. Larger employers are typically located in more populated areas. One could argue that this is part of the set of attributes defining large firms. Alternatively, we would like to know if the first-employer size premium is simply driven by geographical effects of more populated areas. I test for this possibility by estimating equa-

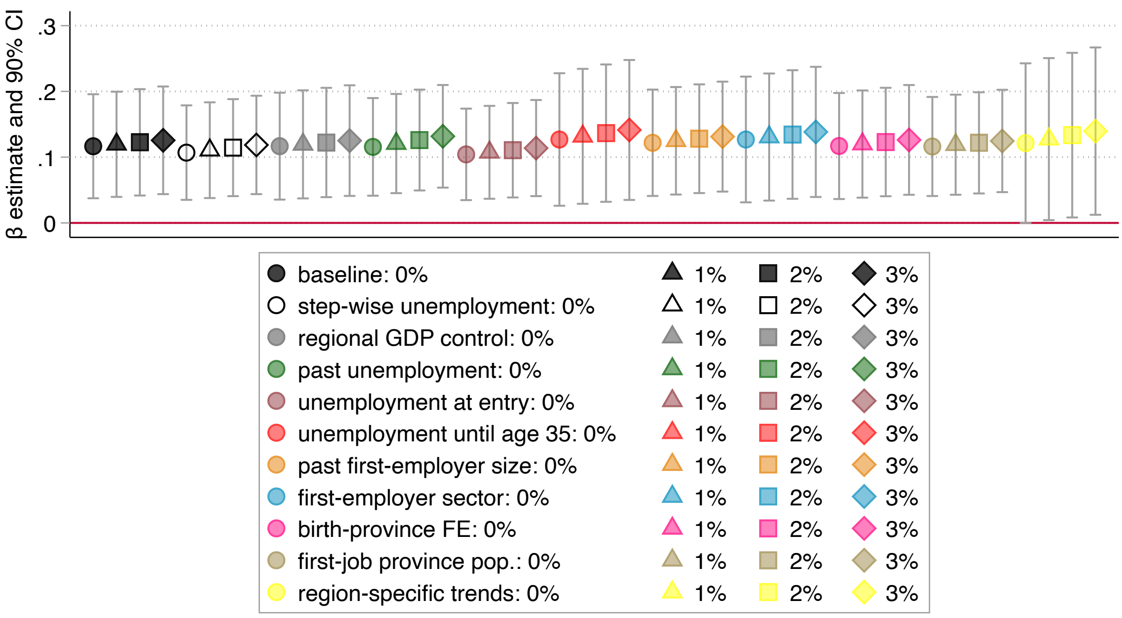
[2] In a small number of cases the data for a given firm does not go back enough. When this happens I average over the amount of prior years of data available.

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

tion (2) with an additional control: (log) population of the province where the worker held his first job. The results from this specification are represented by the brown markers in Figure B2 and are essentially identical to baseline.

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

Figure B2: IV elasticity of lifetime income w.r.t. first-employer size: robustness



Notes: Point estimates and 90% confidence intervals of the IV TSLS elasticity of lifetime income with respect to first employer size using varying specifications of equation (2) in the text. Different marker shapes correspond to different annual discount factors in lifetime income computation. Black markers: baseline results coinciding with those in Table 3, columns (6)–(9). White markers: using a step-wise function of regional unemployment instead of the baseline quartic function. Gray markers: controlling for regional GDP growth during predicted graduation year at region of birth in addition to unemployment. Green markers: controlling for the unemployment rates in years previous to predicted graduation; for college (vocational) workers this includes the unemployment rate present when they would have graduated from high school and vocational (high school) education. Maroon markers: in addition to controlling for a flexible function of the unemployment rate at the time of predicted graduation, I control for a flexible function of the unemployment rate during the actual year of labor market entry. Red markers: In addition to controlling for the unemployment rate at predicted graduation, I control for unemployment rate in all subsequent years until age 35. Orange markers: worker's first employer size measured as the average size over the four years prior to worker's hiring. Blue markers: controlling for sector of first employer (58 sector fixed effects). Pink markers: including province-of-birth fixed effects (instead of region-of-birth). Brown markers: controlling linearly for (log) population of province-year of first job. Yellow markers: including region-specific linear time trends.

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

As I discuss in Section 2, the monthly earnings measure in social security data is censored. I have followed a procedure similar to Bonhomme and Hospido (2017) to impute monthly earnings for censored observations.[3] While censored observations are few (8.7% and 3% of observations in the monthly panel are top- and bottom-coded respectively), one could wonder about the sensitivity of the main results to the imputation procedure.

A feature of the MCVL data is that social security records are also linked to tax data. The benefit of the tax data is that they provide measures of uncensored annual income. The downside is that, as opposed to social security earnings, tax data does not go back in time retrospectively. Tax earnings data are contemporaneous to each MCVL round, and thus available from 2005 onwards. They are also not available for residents of Navarre and the Basque Country since these regions have independent tax authorities.

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

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

Where y_{it} is the income person i earns in year t (in 2016 Euro).[4] The age at which this income is earned will vary across cohorts in my sample. The oldest (youngest) cohort, born in 1968 (1980), earns Y^{05-15} between the ages of 37 and 47 (25 and 35). I am able to compute this measure for 97% (77,754 workers) of my main analysis sample.

I estimate the elasticity of Y^{05-15} with respect to a worker's first employer size by estimating equation (2), using $\ln(Y_i^{05-15})$ as dependent variable.[5] Table B2 shows the OLS, first stage, and IV-TSLS results. It is reassuring to see that the estimated elasticities are very similar in magnitude to those in Table 3. OLS is equal to .0289 compared to .0269–.0276 in Table 3, IV is equal to .1408 compared to .1166–.1255.

The second check I carry out using uncensored tax data is to replicate the elasticity of income at age 35 with respect to first employer size (see Table 6). This replication is directly comparable since the tax data allows me to compute annual income at age 35 for 11 out of the 13 cohorts in my sample.[6] Table B3 shows that the results using tax data are very similar to those using social security data. The OLS is equal to .026 compared to .037 in Table 6. Reassuringly, the IV estimates are practically the same, .085 compared to .089.

[3]This involves grouping worker-month observations into 5,480 cells c { professional category × age × 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.

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

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

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

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

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

Notes: All variables enter regressions in logs. OLS and IV-TSLS estimates of the elasticity of total 2005–15 tax data income with respect to first employer size. 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.

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

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

Notes: All variables enter regressions in logs. OLS and IV-TSLS estimates of the elasticity of tax data annual income at age 35 with respect to first employer size. 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.

B.6 Job security: Temporary and permanent contracts

I test for the existence of relationship between first-employer size and job security later on in the working life. The interpretation of this type of analysis, however, requires some nuance. In particular, young workers could face a trade-off between a job offering high security and a job opening up future opportunities (getting "stuck" in a bad job).

The Spanish social security data include information on the nature of labor contracts. However, type of contract starts being recorded in my data in 1991 and it is missing in large proportions until 1998.[7] The oldest cohort in my sample was born in 1968, which motivates focusing on job security between the ages of 30 and 35.

Figure B3 shows the prevalence of temporary contracts for workers in my sample when they are between ages 30–35.[8] 45% of workers never work under a temporary contract in this period. By contrast, 12% work exclusively under temporary contracts while aged 30–35. The remaining 43% of people work under both types of contract during this period.

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

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

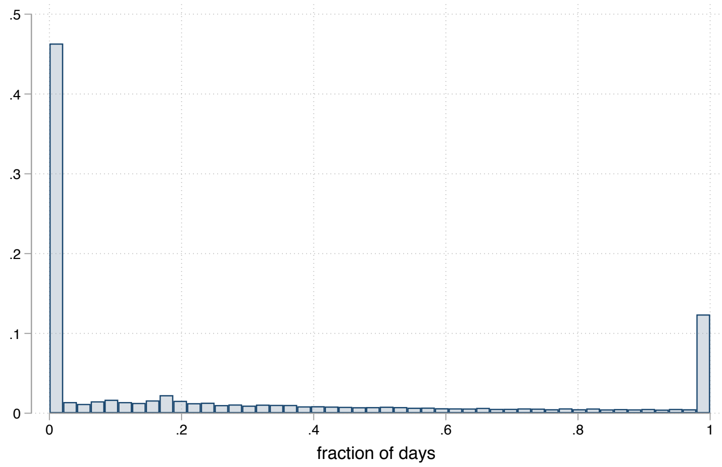
I use these two indices as outcome variables in OLS and IV estimations of equation (2). Table B4 shows the results from this exercise. Columns (1) and (2) show that in OLS first-employer size does a good job at predicting job security experienced between ages 30–35. Starting the working life in a larger employer is significantly correlated with a lower probability of working under temporary contracts during the 30s (column (1)), and a higher probability of experiencing total job security more broadly (column (2)). Columns (4) and (5) show the equivalent IV results. The message is the similar as in OLS, although the estimates are somewhat imprecise. Column (4) indicates a negative causal effect between having a larger first employer and the probability of working later on under temporary contracts. Equivalently, column (5) shows a positive IV effect of first-employer size on the probability of achieving total job security, although the estimate is not statistically significant at conventional levels.

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

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

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

Figure B3: Fraction of days worked under temporary contract between ages 30–35



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

Table B4: Job security between ages 30–35 and first-employer size: OLS and IV-TSLS estimates

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

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

B.7 Job search and human capital in a simple framework

This section complements the discussion in Section 4.2. I provide a simple framework that illustrates how first-employer persistent effects can arise through job search and human capital channels. I first focus on pure search and then add an on-the-job skill component.

Search

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

$$V_t(u_t) = u_t + \beta\Big[\lambda\mathbf{E}\big[\max\{V_{t+1}(u_t), V_{t+1}(u)\}\big] + (1-\lambda)V_{t+1}(u_t)\Big], \tag{B5}$$

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

$$\mathbf{E}\big[\max\{V_{t+1}(u_t), V_{t+1}(u)\}\big] = F(u_t)V_{t+1}(u_t) + \int_{u_t}^{\bar{u}} V_{t+1}(u)f(u)du. \tag{B6}$$

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

$$\mathbf{E}(u_{t+1}|u_t) = \big[(1-\lambda) + \lambda F(u_t)\big]\cdot u_t + \lambda\big(1-F(u_t)\big)\cdot\mathbf{E}(u|u>u_t). \tag{B7}$$

It follows that:[9]

$$\frac{\partial}{\partial u_t}\mathbf{E}(u_{t+1}|u_t) = (1-\lambda) + \lambda F(u_t) > 0. \tag{B8}$$

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

$$V_t(u_t) = u_t + \beta\Big[(1-\delta)\lambda\mathbf{E}\big[\max\{V_{t+1}(u_t), V_{t+1}(u)\}\big] + (1-\delta)(1-\lambda)V_{t+1}(u_t) + \delta D_{t+1}\Big]. \tag{B9}$$

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

$$D_t = \beta\Big[\lambda\mathbf{E}\big[V_{t+1}(u)\big] + (1-\lambda)D_{t+1}\Big]. \tag{B10}$$

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

Human capital

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

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

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

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

$$K_{t+1} = K_t + A_t K_t, \tag{B11}$$

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

$$\begin{aligned} V_t(K_t, A_t) = RK_t + \beta\Big[&(1-\delta)\lambda\mathbf{E}\big[\max\{V_{t+1}(K_{t+1}, A_t), V_{t+1}(K_{t+1}, A)\}\big] \\ &+ (1-\delta)(1-\lambda)V_{t+1}(K_{t+1}, A_t) + \delta D_{t+1}(K_{t+1})\Big]. \end{aligned} \tag{B12}$$

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

$$D_t(K_t) = \beta\Big[\lambda\mathbf{E}\big[V_{t+1}(K_t, A)\big] + (1-\lambda)D_{t+1}(K_t)\Big]. \tag{B13}$$

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

$$K_t = g\big(K_0, \{A_\tau\}_{\tau=0}^{t-1}\big). \tag{B14}$$

Finally, note that the human capital accumulation function (B11) implies that K_t increases proportionally, an example where initial investments (and thus initial draws of A) can be particularly relevant for long-term human capital accumulation. An example of an alternative law of motion explicitly capturing the idea that formative years could be more fruitful for human capital development is

$$K_{t+1} = K_t + A_t f(a_t) K_t, \tag{B15}$$

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

B.8 Career Dynamics

This section examines the relationship between first-employer size and measures of career dynamics: being employed at the first employer at age 35, being employed in the same sector as the first employer at age 35, tenure at the first job, and number of employers up until age 35. Figure B4 shows binned scatterplots that control for the covariates in equation (2). Figure B5 shows the same relationships without adjusting for covariates (unconditional). Since the patterns are similar, I focus the discussion on Figure B4. Table B5 shows OLS and IV estimates of β in equation (2) where, instead of lifetime income, outcome variables are career dynamics measures.

Employment at first-job firm at age 35. Figure B4, panel (a), shows the probability of being employed at age 35 in the firm where a worker held his first job, as a function of first-employer size. The probability is increasing among very small firms, going from 0.025 to 0.08. The probability stays quite flat around 0.08–0.10 for most of the size distribution. The gradient is increasing again for the largest firms, starting at about 1,000 employees. The probability is then much higher, over 0.2, for the very largest sizes. As such, the relationship is increasing overall, but flat for the vast majority of the first-employer size distribution. The OLS estimate in Table B5, column (1), equals 0.0091, in line with the overall slightly positive graphical relationship. The IV point estimate in column (5) is similar to OLS, equal to 0.0125, but not statistically significant.

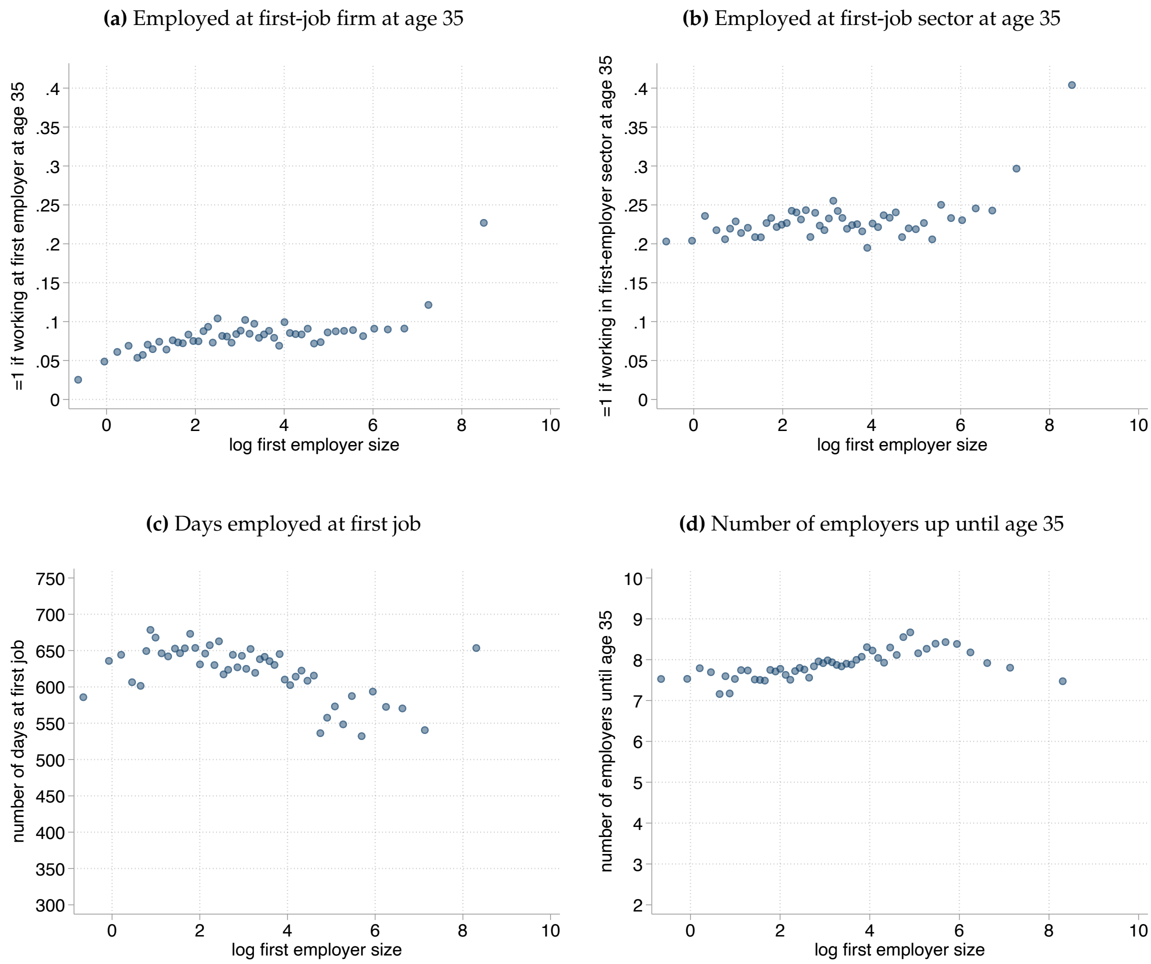
Employment at first-job sector at age 35. Figure B4, panel (b), shows the probability of being employed at age 35 in the sector where a worker held his first job, as a function of first-employer size. The gradient is flat, ranging between 0.2–0.25 for most of the size distribution—from the smallest firms up until about size 1,000. Then, among the very large sizes, the gradient slopes upwards and for workers with the biggest first employers it reaches 0.4. The OLS estimate in Table B5, column (2), equals 0.0084, in line with the overall slightly positive graphical relationship. The IV point estimate in column (6) is similar to OLS, equal to 0.0109, but not statistically significant.

Duration of first-job employment. Figure B4, panel (c), shows the average duration (in days) of workers' first job, as a function of first-employer size. The average is conditional on switching jobs at least once by age 35. The gradient is slightly decreasing, from around 650 days for those starting at small firms, to around 550 days for those starting at firms sized 150–1,000. However, the average jumps back up to 650 among those who start at the very largest firms. Table B5, column (3), shows the OLS parameter estimate from equation (2) when the outcome variable is a dummy equal to one if first job duration is greater than 730 days (i.e., two years, about the 75th percentile). The estimate is rather close to zero, equal to -0.0056. The IV estimate in column (7) is even closer to zero and not statistically significant.

Number of employers up until age 35. Figure B4, panel (d), shows the average number of employers a worker has up until age 35, as a function of first-employer size. The average is conditional on switching jobs at least once by age 35. The gradient is rather flat, with the average mostly ranging between 7.5 and 8.5 employers. Table B5, column (4), shows the OLS parameter estimate from (2) when the outcome variable is a dummy equal to one if a worker has 10 or more employers up until age 35 (10 is the 75th percentile). The estimate is slightly positive, equal to 0.0073. The IV point estimate in column (5) is similar to OLS, equal to 0.0082, but not statistically significant.

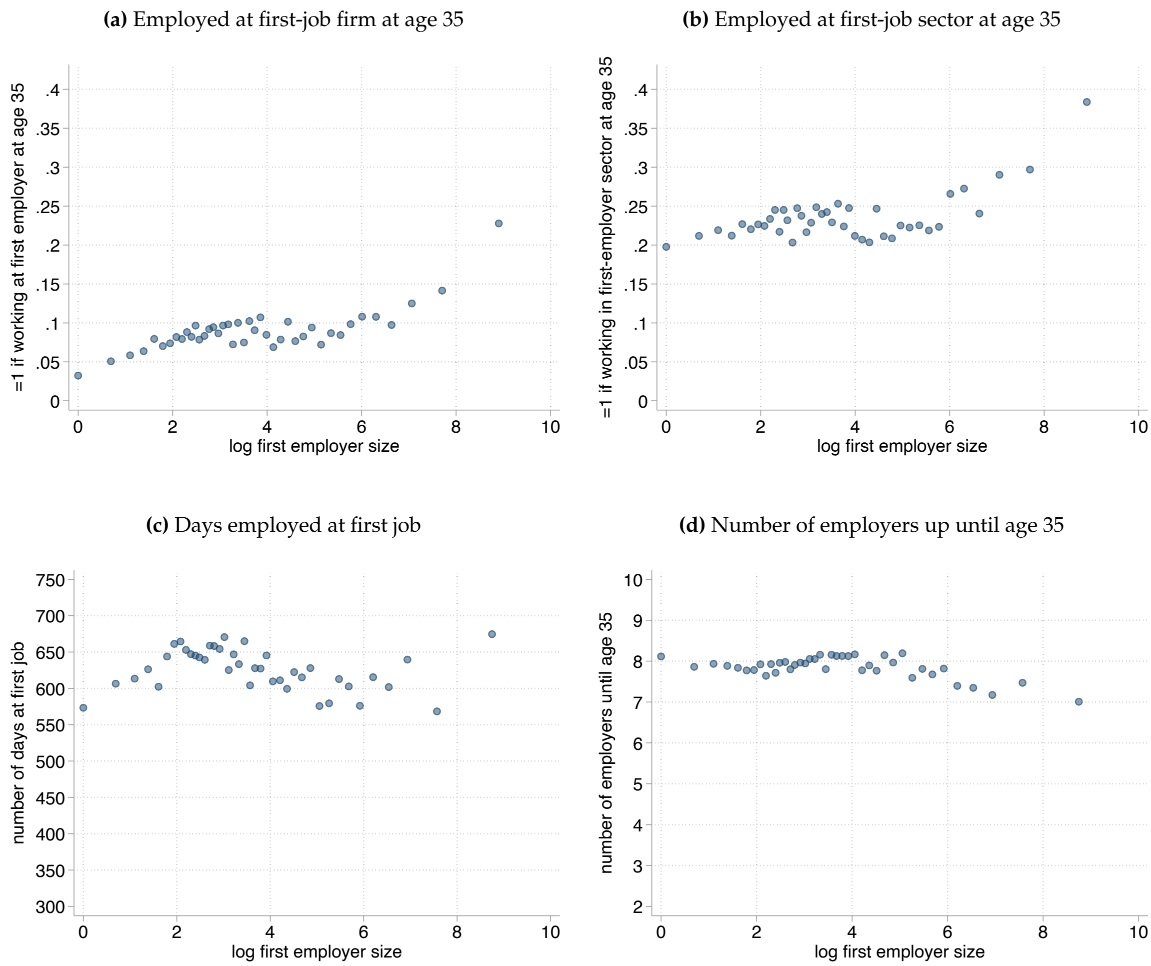
All in all, these measures of career dynamics vary along the first-employer size distribution but not a great deal. The only exception is those who start at the very largest firms, who are significantly more likely to still be found there by age 35, or at least in the same sector. These, however, are a small number of workers. OLS and IV estimates are similar in magnitude and not very large, in line with the graphical patterns. In the case of IV, estimates are not statistically significant.

Figure B4: Career dynamics and first-employer size: binned scatterplots with controls



Notes: All binned scatterplots adjust 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. In all panels horizontal axis represents log size of first employer. Vertical axis varies across panels as follows. Panel (a): dummy equal to 1 if worker is employed at age 35 at his first employer. Panel (b): dummy equal to 1 if worker is employed at age 35 in the same sector as his first employer. Panel (c): number of days employed at the first job. Panel (d) number of employers up until age 35. Panels (c) and (d) restrict attention to workers who have more than one job up until age 35.

Figure B5: Career dynamics and first-employer size: unconditional binned scatterplots



Notes: Unconditional binned scatterplots. In all panels horizontal axis represents log size of first employer. Vertical axis varies across panels as follows. Panel (a): dummy equal to 1 if worker is employed at age 35 at his first employer. Panel (b): dummy equal to 1 if worker is employed at age 35 in the same sector as his first employer. Panel (c): number of days employed at the first job. Panel (d) number of employers up until age 35. Panels (c) and (d) restrict attention to workers who have more than one job up until age 35.

Table B5: Career dynamics and first-employer size: OLS and IV-TSLS estimates

	OLS				IV - TSLS			
	=1 same employer at age 35 (1)	=1 same sector at age 35 (2)	=1 first-job tenure over 730 days (3)	=1 over 10 employers until age 35 (4)	=1 same employer at age 35 (5)	=1 same sector at age 35 (6)	=1 first-job tenure over 730 days (7)	=1 over 10 employers until age 35 (8)
first employer size	0.0091***	0.0084***	-0.0056***	0.0073***	0.0125	0.0109	0.0014	0.0082
	(0.0007)	(0.0012)	(0.0012)	(0.0008)	(0.0230)	(0.0309)	(0.0358)	(0.0326)
Dep. var. mean	0.083	0.231	0.261	0.274	0.083	0.231	0.261	0.274
SE Clusters	661	661	661	661	661	661	661	661
Observations	66,183	60,101	76,156	76,156	66,183	60,101	76,156	76,156

Notes: First employer size enters regressions in logs. IV-TSLS estimates instrument for first-employer size using the labor-demand composition index defined in the text. Outcome in columns (1) and (5) is a dummy equal to 1 if worker is employed at age 35 at his first employer. Outcome in columns (2) and (6) is a dummy equal to 1 if worker is employed at age 35 in the same sector as his first employer. Outcome in columns (3) and (7) is a dummy equal to 1 if worker spent more than 730 days at his first job. Outcome in columns (4) and (8) is a dummy equal to 1 if worker had 10 or more employers until age 35. Regressions at the worker level. Regressions in columns (3), (4), (7), and (8) restrict attention who workers who have more than one job up until age 35. 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.

B.9 Additional evidence on persistent effects

This section provides additional evidence on the persistent of first-employer effects consistent with a human capital channel.

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

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

$$y_{iq} = (\beta_1 + \beta_2 \cdot q + \beta_3 \cdot q^2) \cdot s_{J(i)} + X_{iq}'\gamma + \varepsilon_{iq}. \tag{B16}$$

Where y_{iq} is the log of quarterly income of worker i, q quarters after labor market entry. The β coefficients allow the elasticity with respect to first employer size, $s_{J(i)}$, to follow a quadratic trend. The vector X_{iq} includes a series of controls whose coefficients are also allowed to vary across time.[10] Table B6 shows the implied elasticities at different points in time (Table B7 shows the underlying β estimates). Table B6 allows a quadratic trend as in equation (B16), or imposing a linear trend (assuming $\beta_3 = 0$).

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

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

Notes: Elasticity of quarterly income with respect to first-employer size at different points in time after labor market entry. Based on IV-TSLS estimates of equation (B16) in the text, shown in Table B7. Standard errors clustered at the level of region of birth × education × birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

The β estimates in Table B7 indicate an increasing first-employer size effect. This is true for both the linear and quadratic time trends, and it implies that a larger first employer results in higher earnings growth. Focusing on the linear trend, Table B6 shows that the time-varying elasticity three and six years after labor market entry is 0.026 and 0.056 although imprecisely estimated. Nine years after labor market entry this value is 0.087, and 12 years after it is the same value as the baseline lifetime elasticity, 0.117. The quadratic time trend delivers qualitatively similar results, although the implied elasticity twelve years after entry is somewhat larger, equal to 0.161.

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

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

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

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

Notes: OLS and IV-TSLS estimates of the time-varying elasticity of quarterly income with respect to first-employer size outlined in equation (B16). Regressions at the worker$\times$quarter level. Dependent variable is log total quarterly income, and q is the number of quarters passed since labor market entry. All regressions control for an education-specific quartic function function of the regional unemployment at predicted graduation year, birth-cohort fixed effects, and education fixed effects. All these controls are allowed to vary across quarters. Also control for region-of-birth fixed effects, and quarter fixed effects. Columns (1) and (3) allow a linear time trend while Columns (2) and (4) allow a quadratic one. TSLS estimates in Columns (2)–(3) use as instrument the labor demand instrument described in the text and the same instrument interacted with q and q^2. Standard errors clustered at the level of region of birth $\times$ education $\times$ birth cohort in parentheses. * 0.10 ** 0.05 *** 0.01.

Wage growth between the first and second job

One could still wonder whether persistence results are driven by the small fraction of people who stay with their first employer throughout this time period.[11] To address this, I test whether persistent first employer effects still arise when explicitly taking into account job mobility and initial wages at different jobs. I test whether workers with larger first employers experience greater wage growth when moving to their second job, holding constant first job tenure and second employer size. I do this by estimating

$$g_i^{1,2} = \beta_1 s_{J_1(i)} + \beta_2 s_{J_2(i)} + \rho \ln(\bar{w}_{i1}) + f_1(tenure_i^1) + f_2(tenure_i^2) + g(unemp_i^{1,2}) + X_i'\gamma + \varepsilon_i. \quad \text{(B17)}$$

Where $g_i^{1,2} \equiv \ln(\bar{w}_{i2}) - \ln(\bar{w}_{i1})$ is the growth rate between the average daily wage worker i earned in his second job ($\bar{w}_{i2}$) and the one he earned in his first job ($\bar{w}_{i1}$). $s_{J_1(i)}$ and $s_{J_2(i)}$ are log employer size for the first and second employers, $tenure_i^j$ is the amount of days i worked at his jth employer, $unemp^{1,2}$ controls for the existence and length of an unemployment spell between the first and second jobs, and X_i includes the same controls as equation (2) in addition to start of second job year dummies.[12]

I estimate different specifications of equation (B17). Table B8 shows OLS and IV-TSLS estimates of β_1. The OLS estimates are small, negative, and close to zero (though precisely estimated). The IV estimates are positive indicating an elasticity of between-job wage growth and first employer size of .09–.11. Thus, it seems that returns to a larger first employer already arise in the form of higher wage growth when moving from the first to the second job.

[11] 6.8% of the workers in the sample remain in their first job until the year in which they reach age 35.

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

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

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

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

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

C IV-TSLS Interpretation and Comparison with OLS

This section provides further insight into the instrumental-variable (IV) two-stage least squares (TSLS) estimation of the elasticity of lifetime income with respect to first-employer size. In particular, Section 3.5argues that heterogeneous treatment effects and compliers' characteristics likely explain the difference between the OLS and IV estimates. While this *local average treatment effect* (LATE) logic is well-known and well-understood for the case of binary treatments and binary instruments, it is less straightforward in settings such as mine where the treatment (first-employer size) and the IV (index of labor demand composition) take multiple values.

Here, I follow Angrist and Imbens (1995) (AI95) to clarify what causal effect is TSLS estimating. I then build on these analytic results and, using a distribution regression framework (Chernozhukov et al., 2013), estimate weights from different parts of the first-employer size distribution that feed into the TSLS estimate. Further, by carrying out this exercise across worker subgroups, I get a better understanding of what type of workers are driving the TSLS estimates. Lastly, I document non-linear OLS elasticities separately for groups based on "likely complier" status. The combination of these results provides a better understanding of the OLS-IV comparison.

C.1 Complier Weights: Derivation

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

Setup. Potential outcomes (lifetime earnings) for worker i whose first-employer (log) size is $s = 0, 1, 2, \ldots, J$ are denoted by Y_{si}.[13] The instrument is represented by Z_i and it could be binary $Z_i \in \{0, 1\}$, or multivalued $Z_i \in \{0, 1, 2, \ldots, K\}$. My application involves the latter, but the binary case is simpler to build intuition. Different values of the instrument induce different potential treatment values. S_{Zi} denotes first employer (log) size for each different instrument value. With a binary instrument, each worker i has two potential treatment values, S_{1i} and S_{0i}.

C.1.1 Binary instrument

Consider the following assumptions:

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

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

What causal effect is TSLS estimating? AI95, Theorem 1, shows that TSLS identifies a weighted average of causal responses to a unit change in treatment, $Y_{si} - Y_{(s-1)i}$, for those whose treatment status is affected by the instrument. Compliers in this case are characterized by the base level at which they comply S_{0i}, and by the intensity of compliance, $S_{1i} - S_{0i}$.

[13]Positive integers are not a realistic representation of log firm size, but units are immaterial in this discussion.

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

More specifically, AI95, Theorem 1 shows that

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

where

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

Note that ω_s, the weight attached to the average of $Y_{si} - Y_{(s-1)i}$, is proportional to the amount of workers that the instrument induces to change first employer size from less than s to s or more. That is, the relevant "s-compliers" are those for whom the IV induces their treatment value to "jump over" s.

From which treatment values is β^{TSLS} mostly coming from? The unit-response weights above can be estimated with observables S_i and Z_i since

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

Plotting the weighting function

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

would show which s values have higher weight in β^{TSLS}.

C.1.2 Multivalued Instrument

When both the treatment and the instrument are multivalued the interpretation becomes more involved but the intuitions from above carry forward.

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

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

AI95 show that, similarly as for their Theorem 1,

$$\beta_{k,l} = \sum_{s=1}^{J} \omega_s^{kl} \cdot E[Y_{si} - Y_{(s-1)i} | S_{ki} \geq s > S_{li}], \tag{C3}$$

where

$$\omega_s^{kl} = \frac{Pr(S_{ki} \geq s > S_{li})}{\sum_{m=1}^{J} Pr(S_{ki} \geq m > S_{li})}. \tag{C4}$$

Theorem 2 in AI95 concludes that in the multivalued instrument case

$$\begin{aligned} \beta^{TSLS} &= \sum_{k=1}^{K} \mu_k \beta_{k,k-1} \\ &= \sum_{k=1}^{K} \sum_{s=1}^{J} \mu_k \omega_s^{k,k-1} \cdot E[Y_{si} - Y_{(s-1)i} | S_{ki} \geq s > S_{k-1,i}] \end{aligned} \tag{C5}$$

where

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

and

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

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

C.2 Complier Weights: Estimation

I propose an estimation procedure for weights $\omega_s^{k,k-1}$ that relies on estimating a flexible version of the TSLS first-stage. In a general way, I model the first stage with the conditional distribution function

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

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

After estimating $F(s|Z_i, X_i)$ the first goal will be to study properties of the weights $\omega_s^{k,k-1}$ in equation (C5):

$$\omega_s^{k,k-1} = \frac{Pr(S_{ki} \geq s > S_{k-1,i})}{\sum_{m=1}^{J} Pr(S_{ki} \geq m > S_{k-1,i})}. \tag{C6}$$

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

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

$$F(s|Z_i, X_i) = \Lambda\Big(g\big(Z_i, X_i; \theta(s)\big)\Big) \qquad \text{for all } s \in \mathcal{S}, \tag{C7}$$

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

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

where the controls X_i are the same as in the main IV specification from equation (2) in the main text: a quartic of regional unemployment rate at predicted graduation interacted with

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

educational attainment fixed effects, birth region fixed effects, and birth cohort fixed effects.

Obtaining estimates $\hat{\theta}'(s) = \left[\hat{\gamma}_0(s), \hat{\gamma}_1(s), \hat{\delta}(s)\right]$ for each $s \in \mathcal{S}$ involves estimating the following $|\mathcal{S}|$ logit regressions:

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

Continuous weighting function. Using the parameter estimates $\hat{\theta}(s)$ I compute objects that resemble the weights of of equation (C6). The key idea is to use the estimated distribution first stage and note that for an instrument Z_i that is close to continuous such as mine

$$Pr(S_{ki} \geq s > S_{k-1,i}) = Pr(S_i \geq s|Z_i = k) - Pr(S_i \geq s|Z_i = k-1) \approx \frac{\partial Pr(S_i > s|Z_i = k)}{\partial Z}, \tag{C8}$$

and that the distributional regression model with logistic link function readily provides an expression for the derivative of interest:

$$\frac{\partial Pr(S_i > s|Z_i, X_i)}{\partial Z} = -\gamma_1(s)\cdot\Lambda\Big(\gamma_0(s)+\gamma_1(s)Z_i+X_i'\delta(s)\Big)\cdot\Big[1-\Lambda\Big(\gamma_0(s)+\gamma_1(s)Z_i+X_i'\delta(s)\Big)\Big]. \tag{C9}$$

Taking equations (C6) and (C8) together, I define an estimable two-dimensional weighting function which averages across the distribution of covariates X_i:

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

where:

$$\phi(k) = \left(\sum_{m=1}^{J} \frac{\partial Pr(S_i > m|Z_i = k, X_i = x)}{\partial Z}\right)^{-1}. \tag{C10}$$

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

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

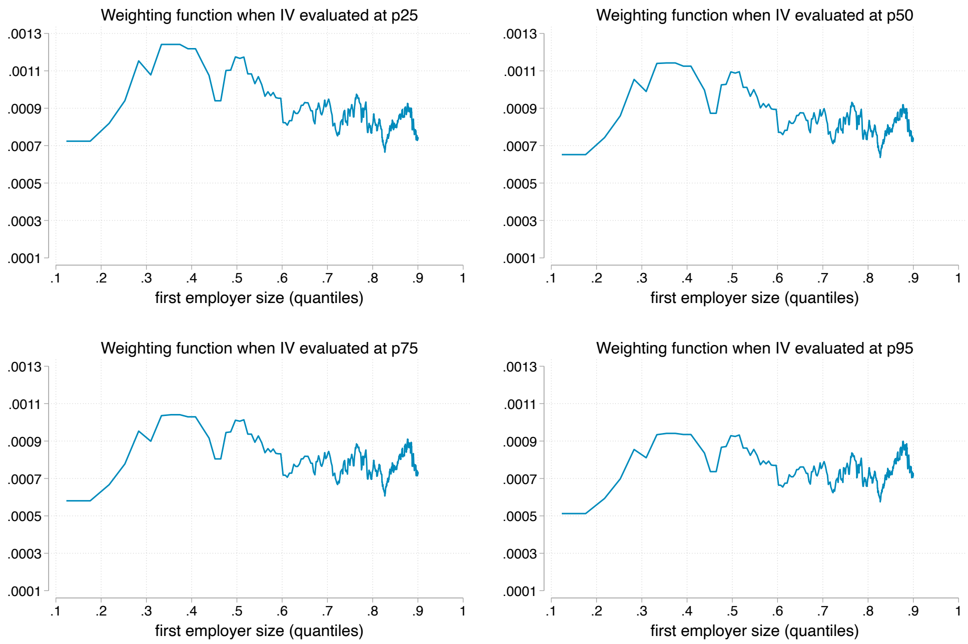
where:

$$\hat{\phi}(k) = \left[\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{J}\Bigg(-\hat{\gamma}_1(s)\cdot\Lambda\Big(\hat{\gamma}_0(s)+\hat{\gamma}_1(s)k+X_i'\hat{\delta}(s)\Big)\cdot\Big[1-\Lambda\Big(\hat{\gamma}_0(s)+\hat{\gamma}_1(s)k+X_i'\hat{\delta}(s)\Big)\Big]\Bigg)\right]^{-1} \tag{C11}$$

From equation (C6) we can interpret weights $\hat{r}(s,k)$ as putting higher values on the levels of s that, induced by the instrument, more people "jump over." Under this interpretation, Figure C1 suggests that, generally speaking, marginal changes in the instrument induce moves across the whole spectrum of the first-employer size distribution.[14] This is because the weighting function is positive and not very steep throughout the size distribution. The pattern is quite homogeneous for different values of the instrument.

[14]An alternative possibility would have been the instrument inducing changes in first-employer size only among local areas of the firm-size distribution, such as very small or very large firms.

Figure C1: Estimated weight function from flexible first stage



Notes: Estimated weight function $\hat{r}(s,k)$ from equation (C11) as a function of first-employer size s. Plotted across panels evaluating the instrument Z at different values k.

C.3 Complier Weights for Subgroups

We can learn something about what are the characteristics of people more responsive to the instrument—characteristics of "compliers"—using the machinery developed above. Consider the weights $\omega_s^{k,k-1}$ in equation (C6) and note that, based on equation (C8) and the exogeneity of the instrument, they can be written as

$$\begin{aligned}\omega_s^{k,k-1} &\approx \phi(k) \cdot \frac{\partial Pr(S_i > s | Z_i = k)}{\partial Z} \\ &= \phi(k) \cdot \left[\frac{\partial Pr(S_i > s | Z_i = k, C_i = 1)}{\partial Z} \cdot Pr(C_i = 1) + \frac{\partial Pr(S_i > s | Z_i = k, C_i = 0)}{\partial Z} \cdot Pr(C_i = 0) \right],\end{aligned} \tag{C12}$$

where C_i is a binary individual characteristic of interest.

Let $m_C(c; s, k) \equiv \frac{\partial Pr(S_i > s | Z_i = k, C_i = c)}{\partial Z}$. Equation (C12) illustrates that $m_C(c; s, k)$ quantifies the proportion of "s-compliers" for whom C_i is equal to c, over and above their unconditional proportion in the population, $Pr(C_i = c)$. That is, if $m_C(1; s, k)$ is greater than $m_C(0; s, k)$, workers for whom $C_i = 1$ are overrepresented among "s-compliers"—i.e., those that the instrument induces to "jump over" first-employer size s. I estimate $m_C(c; s, k)$ for two C_i variables, a dummy equal to one if a worker has a college degree, and a dummy equal to one if a worker was born in one of the more rural provinces in Spain.[15] Figure C2

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

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

shows estimates of $m_C(c; s, k)$ for each of these variables.[16]

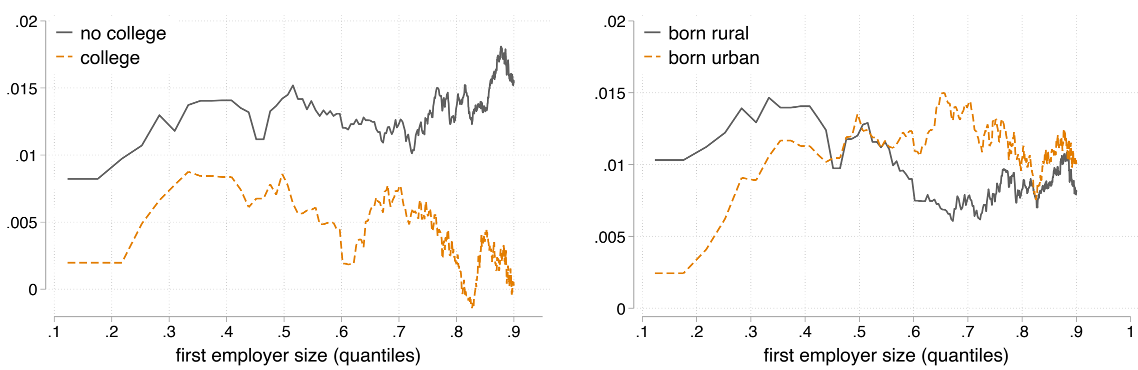
Figure C2, panel (a) shows that, compared to college workers, non-college workers are more responsive to the instrument throughout the first-employer size distribution. This implies that non-college workers are disproportionately represented in the "compliers" group. Panel (b) shows that $m_C(c; s, k)$ for the rural-born has a U-shape relative to the urban-born. $m_C(c; s, k)$ for rural-born workers is higher in the bottom half of the size distribution, lower in the mid-high part, and equal to urban-born workers in the very top. This implies that rural-born workers play an important role in the "compliers" group. They are disproportionately likely to avoid very small first employers as a response to variation in the IV, and they are equally likely to be driven into the very largest employers, even if these are less present in rural areas.

Taken together, these results suggest that young workers with lower earnings potential (less educated, less urban) play an important role in the "compliers" group and disproportionately underpin TSLS estimates in the main analysis.

Figure C2: Complier weights for subgroups

(a) Shares according to college- vs. non-college education

(b) Shares according to urban/rural place of birth



Notes: Estimated share functions $m_C(c; s, k)$, as defined in the text. In panel (a), C is a variable classifying workers into college- and non-college-educated. In panel (b), C is a variable classifying workers as born in the more rural provinces of Spain or not. Both figures evaluate the instrument at k equal to the 95th percentile and plot $m_C(c; s, k)$ as a function of first-employer size s.

C.4 Nonlinear OLS Elasticities for Subgroups

Consider the following equation that allows for a nonlinear elasticity of lifetime income, y_i, with respect to first-employer size, $s_{J(i)}$:

$$y_i = \beta_1 s_{J(i)} + \beta_2 s_{J(i)}^2 + \beta_3 s_{J(i)}^3 + \beta_4 s_{J(i)}^4 + \delta' X_i + \varepsilon_i. \tag{C13}$$

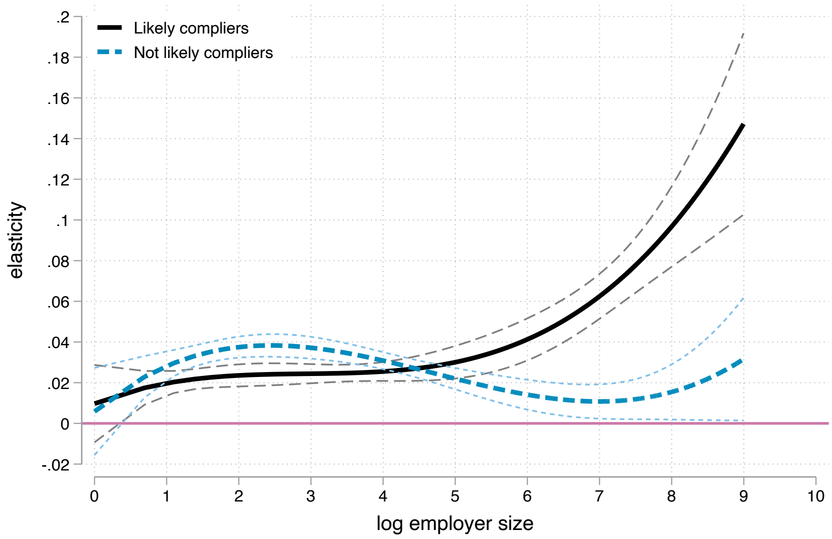
I estimate equation (C13) by OLS separately by "likely complier" status, where I categorize as likely compliers those who do not hold a college degree and were born in the less urban provinces of Spain. Figure C3 shows the estimates of the implied elasticities, as a function of first-employer size. The elasticity is rather constant for both groups of workers in the bottom half of the size distribution. However, starting from first-employer size equal to 150 employees,[17] the elasticity increases steeply for likely compliers, while it decreases sharply and approaches zero for not likely compliers.

[16] These figures evaluate the IV in its 95th percentile.The main takeaways are similar for other IV values.
[17] Note that $\ln(150) = 5$.

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

Keeping in mind the caveats of OLS, these patterns suggest that less educated and less urban workers always benefit from starting at a larger firm, and marginal increases in size are valuable for them even among relatively large firms. More educated and urban workers on the other hand, would benefit from increases in first-employer size among relatively small firms, but not among larger ones.

Figure C3: Nonlinear OLS elasticity, by likely complier status



Notes: OLS elasticity of lifetime income with respect to first-employer size, based on parameter estimates from equation (C13). Elasticity is defined as $\hat{\epsilon}(s) = \hat{\beta}_1 + 2\hat{\beta}_2 s + 3\hat{\beta}_3 s^2 + 4\hat{\beta}_4 s^3$. Standard errors are clustered at the level of region of birth $\times$ education $\times$ birth cohort. Standard error of $\hat{\epsilon}(s)$ computed using the delta-method. 95% confidence intervals for $\epsilon(s)$ are shown.

C.5 Taking Stock and Implications for OLS-IV Comparison

This Appendix has laid out an analytical framework that unpacks what TSLS in the main analysis is estimating in the presence of heterogeneous treatment effects. Building upon this framework and estimating a flexible first stage, I show that variation in the instrument induces shifts in first-employer size across the entire size distribution. Aiming to discern who are "compliers," I find that non-college and rural-born workers likely play an important role in the complier group—non-college throughout the size distribution, and the rural-born especially when avoiding very small firms and accessing the largest ones. Compliers in this type of setting being those who do not re-optimize to labor market conditions in their region of birth and predicted time of entry is an argument put forward by Kahn (2010). Finamor (2022) finds evidence that young people with lower earnings potential are indeed less likely to re-optimize and thus behave more as "compliers."

An OLS elasticity that allows non-linearities, estimated separately for "likely compliers" vs. not, suggests that likely compliers might particularly benefit the most from avoiding small firms and entering large ones. Taking together, these results can explain the larger IV estimates compared to OLS in the main analysis. Under this interpretation, those with lower earnings potential are overrepresented in the group of compliers, and they have the largest gains from starting out at a larger firm.

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

D Distinctive Large-Employer Attributes and Skill Accumulation

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

D.1 Formal Training and Education

Large employers engage in higher amounts of training and in a more structured way. A reason for doing this might be the spreading of fixed costs associated with worker training; another reason might be the higher likelihood of large employers to benefit from training through internal labor markets. Lynch and Black (1998) show that training programs are more prevalent at larger employers, and that these include teaching of general skills such as computing and basic education.[18]

Table D1 uses survey data to show the positive relationship between firm size and employer-provided training in Spain. Workers at employers with 250+ employees are twice as likely to be engaged in informal workplace education than workers at employers with 1–10 employees (3.49% vs. 1.68%), around six times more likely to be engaged in formal workplace education (4.33% vs. 0.75%), and three times more likely to be engaged in either formal or informal workplace education (6.66% vs. 2.30%).[19]

Table D1: Workplace training and education across employer size

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

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

D.2 Organizational Structure

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

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

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

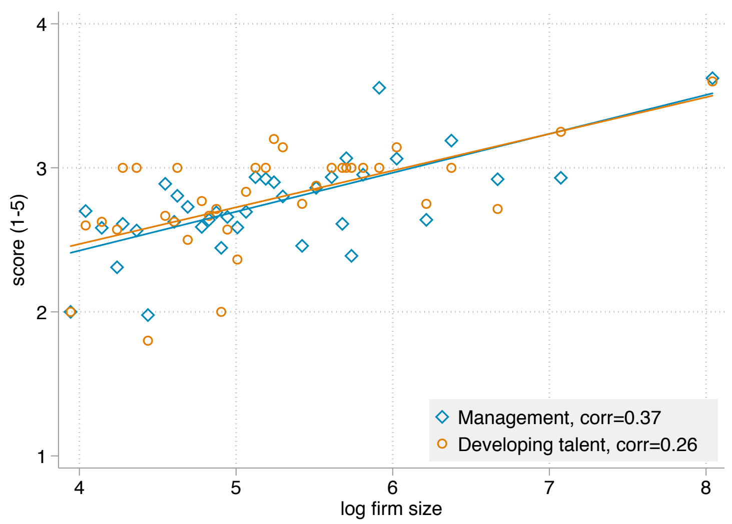
procesing" or "organizational socialization" (Van Maanen, 1978)—how internal processes impact the way in which new workers learn the necessary skills at their new jobs. Many "people processing" practices that could impact a young worker's initial experiences in the firm can only be carried out successfully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

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

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

Managerial and coworker quality. The hierarchical production literature provides complementary theoretical and empirical evidence on the relationship between organizational structure, employer size, and skill-development opportunities for workers (Garicano, 2000; Garicano and Rossi-Hansberg, 2015; Caliendo et al., 2015). Robust predictions of these models are that the marginal return of a worker is linked to the the characteristics of other workers in her team, and that better managers lead better and larger teams (Lucas, 1978). This suggests an opportunity to learn from better peers and better managers at larger employers (see Caicedo et al., 2019; Nix, 2017; Jarosch et al., 2021). Bloom and Van Reenen (2006) show that larger firms tend to be better managed. Using data from the World Management Survey, Figure D1 shows that the correlation between size and management quality is present for Spanish employers.

Figure D1: Firm size and managerial quality in Spain



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

Supplemental Material (not copyedited or formatted) for: Jaime Arellano-Bover. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size."
Journal of Labor Economics 42(2). DOI: https://doi.org/10.1086/723500.

D.3 Firm Production and Activities

Larger employers are more likely to be exporters and, similarly to size, this is a firm attribute the literature has associated with higher wages (Bernard et al., 1995; Bernard and Jensen, 1999).[20] Using data from Italy, Macis and Schivardi (2016) argue that export wage-premia are most important for workers with *previously existing* export-related experience. This is suggestive of a type of skill developed on the job, more likely to be acquired at large employers, and that could be valuable throughout workers' careers. Skills related to exporting activities could be particularly relevant in the context I study, given the undergoing modernization and internationalization of the Spanish economy at the time.

Kugler and Verhoogen (2012) document a strong correlation between manufacturing plants' size and the quality of their inputs and outputs. This complements the fact that larger employers tend to be more productive (e.g. Leung et al., 2008; Moral-Benito, 2018), and evidence suggesting that they are faster to adopt new technologies (e.g. Fabiani et al., 2005). Working with higher quality inputs, adhering to higher quality standards, being involved in more efficient processes, or using more sophisticated technology are channels through which workers might develop higher-value skills at large employers. Table D2 shows that during the 1990s and early 2000s, larger employers in Spain were more likely to invest on R&D and foreign technology transfers.

Table D2: R&D investment, foreign technology transfer payments, and firm size

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

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

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

E Differential Returns to Large-Employer Experience

This section provides additional evidence consistent with human capital mechanisms. I test for a differential return to large-employer experience using the data in its panel dimension, exploiting the richness of its monthly frequency. Observing employer transitions at the daily level and workers' histories since entry allows me to quantify actual experience at different employers measured in days. Experience at large firms could be correlated with unobserved worker characteristics and attributes of the current employer that affect wages.[21] To address this endogeneity problem, the empirical approach features worker fixed effects controlling for worker unobserved heterogeneity, and controls for observable characteristics of the current job (employer size, sector, location, type of contract). The remaining variation that I use compares workers who contemporaneously work for observably similar employers and have the same amount of experience, but acquired this experience in different—large or small—firms.

E.1 Empirical Approach

I estimate the following monthly wage equation:

$$\ln w_{it} = \alpha_i + \psi_{s(i,t)} + \gamma_1 bigExp_{it} + \gamma_2 (bigExp_{it} \cdot Exp_{it}) + X_{it}'\delta + \varepsilon_{it}. \tag{E1}$$

Where w_{it} is the monthly wage of worker i in month t, α_i are worker fixed effects, $\psi_{s(i,t)}$ are size-category fixed effects for worker i's employer at month t, $bigExp_{it}$ is the amount of actual experience (in days) that worker i has accumulated up until month t at employers with 250 or more employees, and Exp_{it} is the amount of total experience (in days, including both large and small employers).[22] X_{it} includes time-varying controls: a quadratic term for total experience, tenure at current employer (quadratic), age (quadratic), regional unemployment level (quadratic), size of municipality or urban area where employer is located (six categories), type of labor contract (permanent or fixed-term), sector fixed effects, and time (year-month) fixed effects.[23]

The parameters γ_1 and γ_2 capture the differential value of experience at large firms and how this differential varies over the working life. Let $Exp_{it} = bigExp_{it} + smallExp_{it}$ and Z_{it} be equation (E1) regressors, then

$$\frac{\partial \mathbf{E}(\ln w_{it}|Z_{it})}{\partial bigExp_{it}} - \frac{\partial \mathbf{E}(\ln w_{it}|Z_{it})}{\partial smallExp_{it}} = \gamma_1 + \gamma_2 Exp_{it}. \tag{E2}$$

Worker fixed effects α_i prevent (time-invariant) unobserved worker heterogeneity (e.g. innate ability) to bias the differential return estimates. Controls for current-employer size, $\psi_{s(i,t)}$ (together with sector and location controls), imply that γ_1 and γ_2 are identified com-

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

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

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

paring workers who have different experience profiles but have the same amount of total experience and are currently working for similar employers. Comparing estimates of γ_1 and γ_2 with and without including $\psi_{s(i,t)}$ is an informative exercise. Intuitively, differential returns to large employer experience *not* controlling for current employer size will combine returns to skill *and* job search. Keeping constant current employer characteristics controls for returns to job search (at least among the observed employer characteristics). That is, if part of the benefits of past experience at a large firm comes from human capital *and* the possibility to be working at a large firm today, specifications including $\psi_{s(i,t)}$ will keep constant the latter channel, making the estimated returns more plausibly attributed to skill accumulation.[24]

E.2 Findings

Columns (1) and (2) of Table E1 show estimates of equation (E1). Column (1) does not include current-employer size category fixed effects $\psi_{s(i,t)}$, while column (2) does. In both cases $\hat{\gamma}_1$ and $\hat{\gamma}_2$ indicate that large-employer experience has higher returns than other experience, and that the differential slowly decreases over time. The fact that $\hat{\gamma}_1$ from column (1) is significantly larger than that from column (2), indicates how a job ladder effect can be of importance. While $\hat{\gamma}_2$ indicates a decreasing differential, the rate of decline is small. Figure E1 helps understand the magnitude implied by the coefficients and its evolution over time. The figure plots the differential return to one year of large-employer experience (specification including $\psi_{s(i,t)}$). Concretely, it plots $365 \cdot 100(\hat{\gamma}_1 + \hat{\gamma}_2 Exp)$ for different levels of Exp, together with 95% confidence intervals. One year of large-employer experience is associated with a return that is between 2–3 percentage points higher than a year of experience elsewhere. As benchmark, the average annual wage growth during the first eight years in the labor market is 10%. These results suggest that there is a differential return to large-employer experience, its magnitude is economically significant, and seems to be more relevant at the beginning of the working life.[25]

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

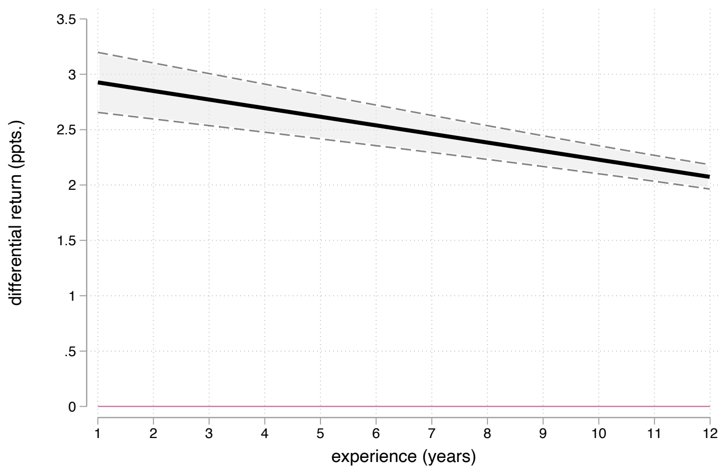
[25] Figure E1 displays marginal effects up until 12 years of (actual) experience since that is close to the average level of experience for workers in my sample at age 35, which is where the panel I use to estimate equation (E1) ends.

Table E1: Differential returns to experience at large employers: Monthly wages

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

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

Figure E1: Differential wage return to one year of large employer experience, by total experience



Notes: Monthly wage differential return to one year of experience at a large employer (250+ employees) with respect to a year of experience elsewhere (<250 employees) for different overall experience levels. Uses estimates of equation (E1) (Table E1, column (2)) and plots $365 \cdot 100(\hat{\gamma}_1 + \hat{\gamma}_2 Exp)$ and a 95% level confidence interval computed using the delta method. Exp is measured in days, x-axis re-scaled for readability. Standard errors are clustered at the worker level.

Potential threats

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

Signaling. I have interpreted the differential wage return to large-employer experience as evidence of differential human capital acquisition at large employers. Consider an alternative interpretation. Working at a large or small employer makes no difference in terms of human capital development. However, big-firm experience serves as a signal of high unobserved ability for subsequent employers.[26] Then, workers with big-firm experience are paid more not because of what they have learnt at these jobs, but because employers believe these workers are of high productivity.

I test for this possibility following the logic of Altonji and Pierret (2001). The idea is that under the pure signal interpretation, the importance of large-employer experience should diminish over time as the worker's true ability is revealed to the employer.[27]

I estimate specifications of equation (E1) that allow for the differential value of large-employer experience to vary by current employer tenure. In particular I augment equation (E1) by estimating

$$\ln w_{it} = \alpha_i + \psi_{s(i,t)} + \gamma_1 bigExp_{it} + \gamma_2(bigExp_{it} \cdot Exp_{it}) + \gamma_3(bigExp_{it} \cdot Tenure_{it}) + X'_{it}\delta + \varepsilon_{it} \tag{E3}$$

This specification allows a differential return to experience in large employers that can vary by experience and tenure. That is, letting Z_{it} be equation (E3) regressors,

$$\frac{\partial \mathbf{E}(\ln w_{it}|Z_{it})}{\partial bigExp_{it}} - \frac{\partial \mathbf{E}(\ln w_{it}|Z_{it})}{\partial smallExp_{it}} = \gamma_1 + \gamma_2 Exp_{it} + \gamma_3 Tenure_{it} \tag{E4}$$

A large and negative $\hat{\gamma}_3$ would be consistent with the idea of large-employer experience serving as a signal for unobserved ability. Columns (3) and (4) of Table E1 show estimates of equation (E3) without and with $\psi_{s(i,t)}$, respectively. Focusing on column (4), the table shows that $\hat{\gamma}_1$ is essentially unchanged with respect to that of column (2). $\hat{\gamma}_3$ is negative, consistent with signaling playing some role. Understanding the magnitude of the implied decay by $\hat{\gamma}_3$ will be informative of the extent to which pure signaling drives the differential return to big-firm experience.

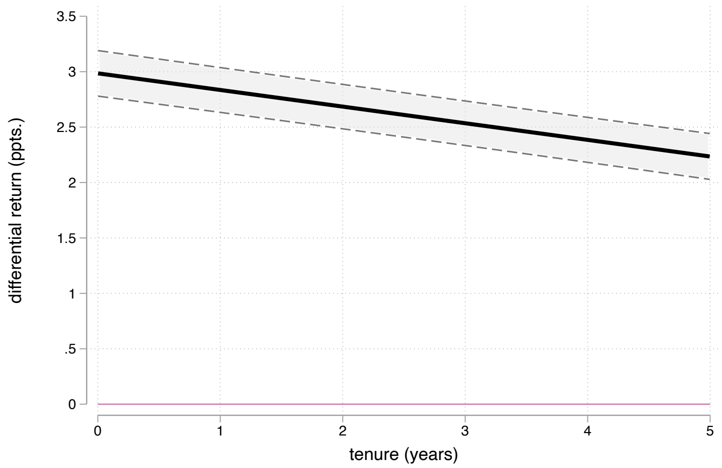
Figure E2 shows the rate of decay as tenure increases, holding constant experience at five years. The data is consistent with large-employer experience having some signaling value, but far from explaining all of the differential return. Given the estimates of $\{\gamma_1, \gamma_2, \gamma_3\}$, a worker should stay at the same employer for over 20 years before the large-employer experience differential vanishes, which is a level of tenure not present in this sample of relatively young, mobile workers.[28]

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

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

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

Figure E2: Differential wage return to one year of large employer experience, by current employer tenure



Notes: This figure plots the monthly wage differential return to one year of experience at a large employer (250+ employees) with respect to a year of experience elsewhere (<250 employees) for different current employer tenure levels, holding overall experience fixed. Uses estimates of equation (E3) (in Table E1, column (4)) and plots $365 \cdot 100(\hat{\gamma}_1 + \hat{\gamma}_2 \overline{Exp} + \hat{\gamma}_3 Tenure)$ and a 95% level confidence interval computed using the delta method. $\overline{Exp}$ set at 1825 days (5 years). $Tenure$ is measured in days, x-axis re-scaled for readability. Standard errors are clustered at the worker level.

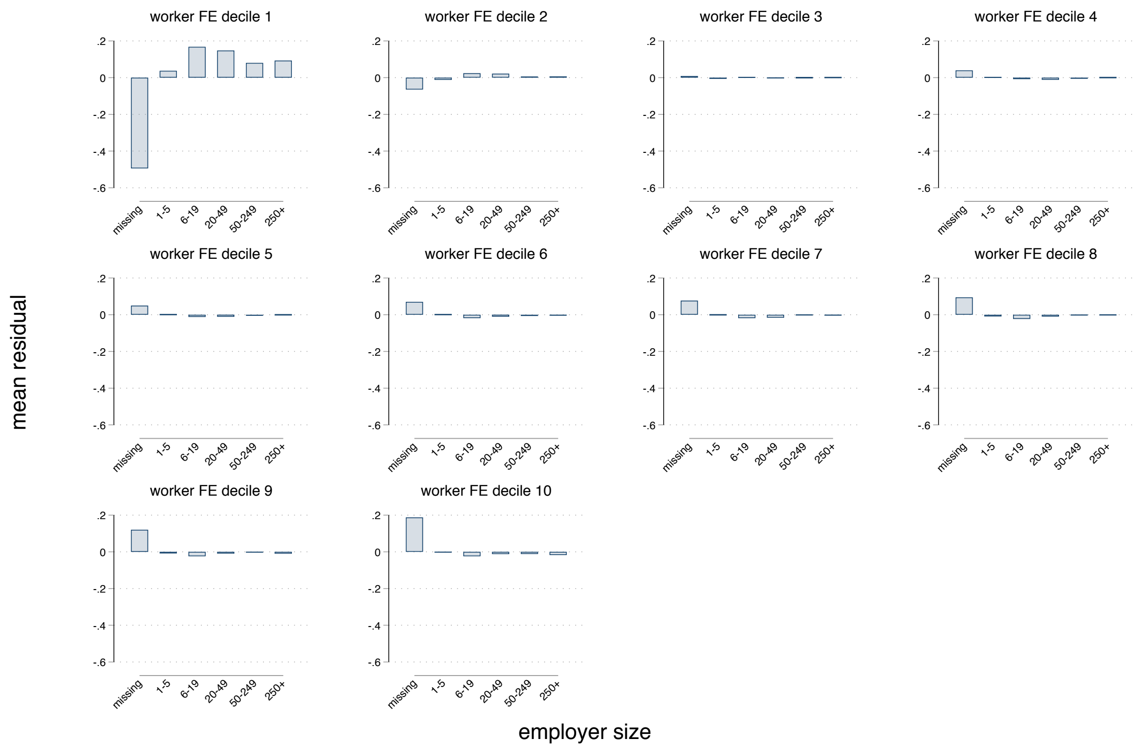
Additive separability. Another concern that could introduce bias in the differential experience return estimates is model misspecification arising in the form of employer-size premia that vary across worker types. This would mean that the assumption of common proportional employer-size premia for all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

Card et al. (2018) discuss how the violation of additive separability in firm and worker effects is a common concern in the AKM literature and provide specification tests that support this assumption in their context. I follow Card et al. (2018) and check the plausibility of the additive specification in equation (E1) by checking the distribution of mean residuals for different employer-size categories and worker types. The logic is that if the additive model is correct, residuals should have mean close to zero for all employer size/worker type combinations. On the other hand if the employer size premiums vary systematically across worker types we should see systematic departures from zero.

Figure E3 plots the mean residual for each cell based on the six employer size categories and ten deciles of estimated worker effects. Mean residuals are relatively close to zero. The largest mean residuals are those corresponding to the lowest paid (1st decile) workers, a finding consistent with Card et al. (2018) which could be explained by minimum wage policies.[29]

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

Figure E3: Mean residuals by worker effect decile/employer size



Notes: Figure shows mean residuals from estimated equation (E1) with cells defined by decile of estimated worker effects (α_i) interacted with employer size category.

E.3 Promotions

Having found a differential wage premium for large-employer experience, I study its relationship to career progression through promotions. The literature has emphasized the connection between promotions and workers' ability or human capital (see Gibbons and Waldman, 1999). A differential return to experience in terms of an increased arrival rate of promotions would further support the hypothesis that skills learned at large employers are more valuable over the working life.[30]

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

$$Prom_{it} = \alpha_i + \psi_{s(i,t)} + \phi_{p(i,t-1)} + \lambda_1 bigExp_{it} + \lambda_2 (bigExp_{it} \cdot Exp_{it}) + X'_{it}\delta + \varepsilon_{it}. \quad \text{(E5)}$$

Where $Prom_{it}$ is a dummy variable that equals one if worker i experienced a promotion on month t, α_i are worker fixed effects, $\psi_{s(i,t)}$ are current-employer size category fixed effects, $\phi_{p(i,t-1)}$ are indicators for the professional category worker i was holding on month $t-1$, $bigExp_{it}$ is the amount of actual experience (in days) that worker i has accumulated up until month t at employers with 250 or more employees, and Exp_{it} is the amount of total ac-

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

tual experience (in days, including large and small employers). X_{it} includes time-varying controls: a quadratic term for duration in current professional category, total experience (quadratic), tenure at current employer (quadratic), age (quadratic), regional unemployment level (quadratic), type of labor contract (permanent or fixed-term), sector fixed effects, and time (month) fixed effects.

In an analogous way to γ_1 and γ_2 in equation (E1) in the text, λ_1 and λ_2 capture the differential impact of large-employer experience in the promotion probability, and how it varies over the working life. Let $Exp_{it} = bigExp_{it} + smallExp_{it}$ and Z_{it} be equation (E5) regressors, then

$$\frac{\partial Pr(Prom_{it} = 1|Z_{it})}{\partial bigExp_{it}} - \frac{\partial Pr(Prom_{it} = 1|Z_{it})}{\partial smallExp_{it}} = \lambda_1 + \lambda_2 Exp_{it}. \tag{E6}$$

Table E2: Differential returns to experience at large employers: Promotion arrival rate

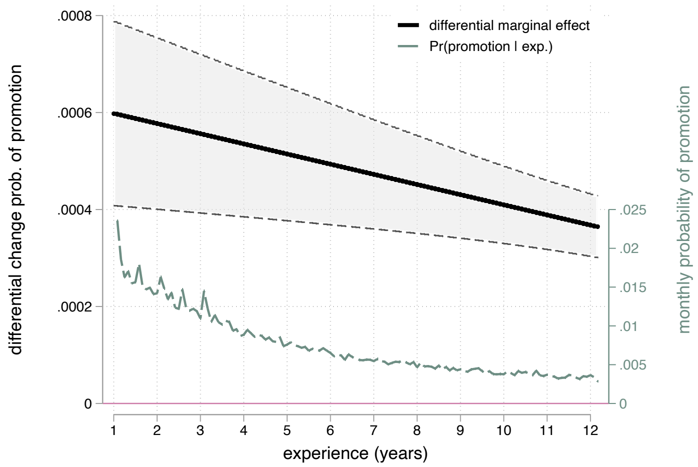
	(1)	(2)
$bigExp$	1.4458***	1.6964***
	(0.2675)	(0.2847)
$bigExp \cdot Exp$	-0.0001**	-0.0002***
	(0.0001)	(0.0001)
Exp	-6.6838***	-6.7633***
	(0.2043)	(0.2046)
Exp^2	0.0013***	0.0013***
	(0.0000)	(0.0000)
$Prof.cat. - duration$	16.7451***	16.7498***
	(0.1343)	(0.1343)
$Prof.cat. - duration^2$	-0.0028***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

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

Columns (1) and (2) of Table E2 show estimates from equation (E5). Column (1) does not include current employer size category fixed effects $\psi_{s(i,t)}$, while column (2) does. In both cases $\hat{\lambda}_1$ and $\hat{\lambda}_2$ indicate that large-employer experience has higher returns in terms of promotion probability that slowly decrease over time. Figure E4 helps understand the relevant magnitude implied by the coefficients and its evolution over time. On the left y-axis, it plots the differential change in the probability of promotion from one year of large-employer experience vs. one year of experience elsewhere together with a 95% confidence interval.[31] To interpret the magnitude of this differential, the right y-axis plots the relevant baseline: the monthly probability of promotion conditional on experience. It ranges from .023 when workers have one year of (actual) experience to .003 when they have twelve. The figure implies that the differential return to one year of large-employer experience amounts to 2.6% of the baseline probability when workers have one year of experience, 8.3% when they have six, and 11.6% when they have twelve.

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

Figure E4: Differential change in probability of promotion to one year of large employer experience, by total experience



Notes: Differential increase in the monthly probability of promotion of one year of experience at a large employer (250+ employees) with respect to a year of experience elsewhere (<250 employees) (left y-axis), and the monthly probability of promotion (right y-axis) for different levels of experience. Left y-axis uses estimates of equation (E5) (Table E2, column (2)) and plots $365 \cdot (\hat{\lambda}_1 + \hat{\lambda}_2 Exp)$ and a 95% level confidence interval computed using the delta method. Exp is measured in days, x-axis re-scaled for readability. Standard errors are clustered at the worker level.

The promotion results suggest that time spent at a large employer is more valuable than that spent elsewhere in terms of future career progression. I interpret this as further supportive evidence for the hypothesis that workers learn differentially valuable skills at large employers that pay off in terms of higher wages and faster career progression.

Construction of promotion variable

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

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

References to Online Appendices

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg (2021). Learning from coworkers. *Econometrica 89*(2), 647–676.

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

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

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

Leung, Danny, Césaire Meh, and Yaz Terajima (2008). Firm size and productivity. Working Paper 2008–45, Bank of Canada.

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.

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

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

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

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

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

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