

Employer-Provided Training, Wage Growth, and Motherhood Penalty: Evidence from Taiwan*

Kuan-Ming Chen[†] Bo-Yan Huang[‡] Ming-Jen Lin[§] Alice Wu[¶]

October 28, 2026

Please click [here](#) for the latest version.

Abstract

Employer-provided training is widely viewed as an important driver of human capital accumulation and wage growth, but a lack of high-quality data has made it difficult to study. Using administrative tax records from Taiwan, we characterize the firms that provide training and examine how training relates to wage growth and to labor market outcomes around first childbirth. Firms that provide more training pay higher wages, invest more in R&D, and offer better benefits. Conditional on these characteristics, working for a high-training firm predicts stronger wage growth over the next decade, so the role of training is not fully absorbed by other firm attributes. The returns are higher for women. Short-term returns accrue only to stayers, but female movers from high-training firms also see higher wage growth five years later, suggesting that on-the-job training is portable across employers. High-training firms also buffer the child penalty in employment and earnings. New mothers from high-training firms stay more attached to the labor market and experience a smaller wage drop than their counterparts from low-training firms. The differences are small for new fathers. High-training firms thus matter more for reducing the child penalty for women, and thereby for narrowing gender inequality.

*We thank Ting-Chun Lin for outstanding research assistance, and the seminar audience of UW Household Finance for helpful comments. acknowledges the support from the National Science and Technology Council grant NSTC 111-2628-H-002-019, 114-2628-H-002-019, and the Yushan Fellow Program by the Ministry of Education (MOE) NTU-114V1037-3. Chen and Lin thank the support from the E. Sun Academic Award. The research data are drawn from tax records provided by the Fiscal Information Agency, Ministry of Finance. The data are used under a memorandum of understanding signed between the data center and Academia Sinica, with strict safeguards in place to protect individual privacy.

[†]National Taiwan University. Email: kuanmingchen@ntu.edu.tw.

[‡]National Taiwan University. Email: b10303005@ntu.edu.tw.

[§]National Taiwan University. Email: mjlin@ntu.edu.tw.

[¶]Department of Economics, University of Wisconsin - Madison. hwu526@wisc.edu

1 Introduction

Technological change, labor market reforms, and individual career concerns all require workers to update their skills throughout their careers. A better understanding of on-the-job training is important for designing policies that help workers meet changing skill demands, but research on this topic has been limited by several data challenges (Lynch 1998). Matched employer-employee data with information on training are rare, and the definition of training varies from survey to survey (Black, Skipper, and Smith 2023). This variation makes it difficult to test theories of training provision (Becker 1964; Stevens 1994; Acemoglu and Pischke 1999) and to provide large-scale evidence on how much training matters for wage growth (Brown 1989; Lynch 1991).

In this paper, we use matched employer-employee data from Taiwan, linked to firms' training expenditures reported in business tax filings, to answer three questions. Which firms train, and who works for them? What are the wage returns to working at a training firm, and do they differ by gender? Does employment at a training firm change how workers, especially mothers, fare around the birth of their first child?

Training expenses are uniformly defined by the tax bureau (Article 86-1 of the Regulations Governing Assessment of Profit-Seeking Enterprise Income Tax; Ministry of Finance 2023), which reduces the concern that what counts as employer-provided training varies across firms or surveys (Lynch 1998). With access to balance sheets and the firm registry, we study the relationship between employer-provided training and a wide range of firm characteristics, including wages, R&D expenses, and firm size. We find a U-shaped relationship between training intensity and the firm-specific wage premium (Abowd, Kramarz, and Margolis 1999; Card, Cardoso, Heining, and Kline 2018). High-training firms are not necessarily high-wage firms. Some employers may set current pay low but provide more training that raises workers' human capital in the future (Becker 1964; Manning 2003). Training is positively related to R&D expenses and employee benefits, but these characteristics, together with industry fixed effects, explain less than 15% of the variation in training intensity. On-the-job training is thus a distinct firm attribute that is not easily predicted from other observables.

Life-cycle patterns in selection into training firms reveal important gender differences. In their 20s, women are more likely than men to start at firms that provide training, but the pattern reverses after the early 30s. The reversal is driven by a drop in the share of women at firms in the bottom 75% of training intensity among firms that train (low-training firms). The female lead at the top 25% (high-training firms) also narrows as workers age. These gender differences in sorting to training may reflect anticipated career changes (Royalty 1996; Blau and Kahn 2017), especially around parenthood (Blundell, Costa Dias, Goll, and Meghir 2021; Ma, Nakab, Navajas-Ahumada, and Vidart 2025). Sorting across firms is an important source of the gender gap in early-career

wage growth, accounting for a fifth of the gap in Italy ([Card, Devicienti, Rossi, and Weber 2025](#)). If women benefit at least as much as men from employer-provided training, their selection out of training firms as they age could widen the gender wage gap through the same channel. We test the premise by estimating gender-specific returns to employment at training firms.

To understand how much employer-provided training matters for human capital accumulation and career progression, we compare wage growth across workers at high-, low-, and no-training firms. Conditional on worker and firm characteristics, including the leave-out mean log wage at the firm and industry fixed effects, men at high-training firms experience 0.9 log points higher one-year wage growth than men at no-training firms, roughly 10% of mean wage growth. Among younger cohorts, whose early careers we observe more completely, the difference between high- and no-training firms is 1.6 log points for men and 2.4 log points for women. Men at low-training firms see wage growth similar to men at no-training firms, whereas younger women at low-training firms see 1.0 log point higher wage growth than their counterparts at no-training firms.

Over longer horizons, employment at a high-training firm continues to predict higher wage growth for both men and women. The estimated differences across training tiers are larger for women than for men, suggesting that exposure to high-training firms matters more for women's career progression, especially early in their careers. Because we control for the leave-out mean log wage, this relationship is not driven by working at a high-wage firm. We also run a horseshoe between high training and two other firm characteristics: R&D expenses, a proxy for reliance on new technologies and technical skills, and employee benefits, a proxy for amenities. The wage return to high-training firms is stable across specifications, further evidence that the relationship between employer-provided training and wage growth is not an artifact of other firm attributes.

An important question in the training literature is whether on-the-job training builds general or firm-specific human capital. In imperfectly competitive labor markets, firms may invest in general skills that are portable across employers, because search and information frictions reduce turnover and allow firms to capture part of the return to training ([Stevens 1994](#); [Acemoglu and Pischke 1998](#); [Acemoglu and Pischke 1999](#); [Manning 2003](#)). Our expense-based measure does not distinguish general from firm-specific training, but we can test the portability of training indirectly by estimating the relationship between employment at training firms and wage growth separately for job stayers and movers.

In the short run, we find evidence consistent with [Lynch \(1991\)](#) that employer-provided training is largely firm-specific: movers from high-training firms see lower one-year wage growth than movers from no-training firms. Over five or ten years, however, younger women who leave high-training firms experience 1.6 to 5.2 log points higher wage growth than those who leave no-training firms, and they also outpace those who leave low-training firms. This suggests that some of the training is portable across employers in the longer run. The pattern is weaker for men who leave

high-training firms, which again points to women benefiting more from exposure to high-training firms.

Employer-provided training may offset part of the labor market penalty of having children (Blundell et al. 2021). We provide new evidence on this hypothesis by studying how workers at different training tiers fare around the birth of their first child. We find that mothers from high-training firms are also more likely to stay with their pre-birth employer.

Following Kleven, Landais, and Søgaaard (2019), we estimate event studies around the first birth separately for women and men and find a child penalty in Taiwan similar to those documented elsewhere: five years after the birth, the earnings penalty for mothers is about 20%, close to the estimate for Denmark. We then compare parents by the training tier of their employer two years before the birth. Before the birth, the employment paths of parents from high- and low-training firms run parallel, although every parent in the sample is employed two years before the birth. After the birth, mothers who worked at a high-training firm lose less. The year after the birth, their full-time employment is 7 percentage points higher than that of mothers from low-training firms, and among those still working, their earnings fall 8 log points less. Both gaps narrow by five years after the birth but do not close. Fathers' outcomes differ little by tier: the gap in full-time employment is about 1 percentage point, and their earnings do not differ.

Mothers from high-training firms are also more likely to stay with their pre-birth employer: one year after the birth, 12 percentage points more of them are still there than mothers from low-training firms. Retention does not fully explain their stronger attachment, however. Fathers from high-training firms also stay with their employer more often, by 8 percentage points, yet their full-time employment barely differs by tier. The difference-in-differences in retention, mothers relative to fathers and high- relative to low-training firms, is about 4 percentage points, roughly two-thirds the size of that in full-time employment. Retention is also far from complete: roughly 40 percent of new mothers from high-training firms have left their pre-birth employer one year after the birth, while their employment falls by only 24 percentage points, so many of them move to another employer. One explanation is that mothers whose wages grow faster at high-training firms have more to lose by leaving. Another is sorting: women who plan to stay in the labor force after childbirth may choose high-training firms in the first place. In the next draft, we will examine these mechanisms using external surveys on training activities, to see whether high-training firms build general skills that help mothers remain employed after childbirth.

This paper contributes to two literatures. First, we provide new evidence on how employer-provided training relates to career progression, with attention to the differential returns for women and men and to the portability of training across employers. The existing literature relies on surveys to measure the incidence and extent of training, but the survey questions vary widely, and whether a worker or a manager responds also matters (Black et al. 2023). We measure training from training

expenses in tax data, which all employers in Taiwan must report. Linking training expenses to employment and wage records also allows us to estimate the returns to training at scale. We find a robust positive relationship between exposure to high-training firms and both short- and long-term wage growth, consistent with [Brown \(1989\)](#) and [Lynch \(1991\)](#). The gender differences in the returns to training, and in how parents fare around the first birth, complement the evidence in [Blundell et al. \(2021\)](#) that employer-provided training can reduce part of the child penalty for women.

Second, we contribute to the literature on the child penalty by showing that the firm a parent worked for before the birth predicts labor market outcomes after it. Family policies such as parental leave and childcare have done little to close the gap between mothers and fathers ([Kleven et al. 2024](#)), which has turned attention to employers. Recent work finds that the penalty is larger at firms with less generous pay or weaker peers ([Casarico and Lattanzio 2023](#)), but no smaller at female-dominated or female-led firms ([Almond et al. 2023](#)), and that employer-provided childcare reduces it for high-wage mothers ([Costas-Fernández et al. 2026](#)). Mothers also receive less training after the first birth ([Ma et al. 2025](#)). We show that a different employer attribute, investment in workers’ skills, predicts a smaller penalty for mothers and makes little difference for fathers.

2 Data

Taiwan offers a useful setting for studying firm-provided training. With 23.4 million residents, Taiwan has more people than Denmark, Norway, and Sweden combined, which had 22.1 million in 2024 ([World Bank 2025](#)). Administrative registers from these three countries underlie much of the evidence on families and the labor market, including the child penalty ([Kleven et al. 2019](#)). Its wage-setting institutions, however, resemble those of the United States rather than those of Scandinavia. Wages are set by individual firms, and collective bargaining is almost absent: only 919 collective agreements were in force nationwide at the end of 2024 ([Ministry of Labor 2025a](#)).¹ Because union-driven wage compression is a leading explanation for why firms pay for general training ([Acemoglu and Pischke 1998](#); [Acemoglu and Pischke 1999](#)), a setting without it is informative about training in decentralized labor markets such as the United States.

2.1 Data Sources

Our data come from administrative records held by Taiwan’s Fiscal Information Agency (FIA), the agency of the Ministry of Finance that processes the country’s tax filings, similar to the Internal Revenue Service in the United States. We work with de-identified records on site in a secure facil-

¹The Ministry of Labor reports the number of agreements but not the number of workers they cover. The count also overstates how widespread bargaining is. Of the 919 agreements, 670 were signed by industrial unions, and most of these trace to a single industrial union in Kaohsiung, which signed agreements with 339 employer groups in 2015 and with another 275 in 2022 ([Ministry of Labor 2025b](#)).

ity. The FIA data are organized as separate files for each tax and registry system, and individuals and firms carry encrypted identifiers that are consistent across files. We link three of these files over the years 2007 to 2023: income tax records, business income tax returns, and the household registry.

Income tax records. Employers report every salary payment they make to the tax authority, with the firm’s tax identification number, the worker’s identifier, and the amount paid. They file these reports whether or not the worker ends up owing income tax, so the records cover nearly all formal wage employment in Taiwan. Summing payments by worker, firm, and year gives a matched employer–employee panel of annual earnings.

Business income tax returns. Profit-seeking enterprises file an annual business income tax return. The return reports the firm’s industry, location, net profit, and operating expenses, including salaries, employee benefits, and R&D. It also has a separate line for training expenses, which is our measure of employer-provided training (Section 2.2). Government agencies do not file these returns, so our analysis covers workers whose main employer is a private-sector firm.

Household registry. The household registry covers every registered resident of Taiwan and records gender, dates of birth and death, marriage, and the identifiers of both parents. Through the parent links we identify each worker’s children and the year of the first birth.

2.2 Measuring Employer-Provided Training

What the training fee measure covers. Article 86-1 of the Regulations Governing Assessment of Profit-Seeking Enterprise Income Tax defines training expenses for all firms (Ministry of Finance 2023; Ministry of Finance 2017), and Appendix A.3 gives its full text. Training expenses are costs a firm pays to train its employees, either by running training activities related to its business or by sending employees to them. They include instructor fees and travel, trainees’ travel and course fees, teaching materials, meals and venues during training, skill-certification fees, overseas study for selected employees, and payments to schools under industry–academia cooperation agreements. The measure captures what firms spend on formal training. Informal learning on the job or from coworkers does not appear on the return, and the line records spending rather than hours of training. Because the definition is set by the tax authority, it avoids a common problem with survey measures, in which each survey defines training differently (Lynch 1998).

Reporting incentives. Training expenses are deducted at their verified amount and must be supported by original vouchers under the same rules as other expenses (Article 86-1). Training is an ordinary deductible expense, so a firm saves the same tax whether it reports a dollar as training or

under another expense line. A firm could still overstate training to lower its taxable income, but this is no different from overstating any other expense. The National Taxation Bureaus audit business income tax returns, and a firm that underreports its taxable income is fined up to twice the tax it evaded ([Ministry of Finance 2026](#), Article 110). We therefore expect training to be misreported no more often than other deductible expenses.

Training tiers. Training is concentrated in a minority of firms. About 16% of firm-years have positive training expenses over a three-year window (Table 1, column 1). Training firms are larger, however, so 57.6% of worker-years are spent at firms with positive training. Among firms that train, spending per worker is highly skewed. We therefore group firms into three tiers by how much they train, which lets us compare worker characteristics, wage growth, and other outcomes across firms that train heavily, firms that train a little, and firms that do not train.

We divide each firm’s training expense by its number of workers, counting each worker at the firm that pays them the most in the year (Section 2.3). Training spending varies over time, so we average training per worker over a rolling three-year window, $t - 2$ to t , and require the firm to be observed in all three years. Firms with no training expense in the window form the no-training tier. Among firms with a positive three-year average, those at or above the 75th percentile form the high-training tier and the rest form the low-training tier. The 75th-percentile cutoff is computed separately for each year. A worker at a high-training firm receives about US\$300 of training per year, roughly 2% of annual earnings.

2.3 Sample Construction

For each worker and year, we sum payments by firm, assigning branch establishments to their headquarters, and define the worker’s main employer as the firm that paid the most. A worker is full-time in a year if annual pay from the main employer is at least 12 times the monthly minimum wage, and part-time if it is between 6 and 12 times. The tax data do not record hours of work, so we use these thresholds to separate regular employment from marginal attachment.

Our main sample consists of full-time private-sector workers aged 22 to 50 whose main employer files a business income tax return and has a training tier in that year. It contains 64 million person-year observations from 2007 to 2023. The wage-growth regressions in Section 4 use a broader sample that also includes part-time workers (at least 6 months of minimum wage), and report separate estimates for workers born in 1987 or later, whose careers we observe from labor market entry. The event-study sample in Section 5 consists of workers with a first birth who were employed two years before the birth. Firm characteristics in that sample are measured in that base year. Appendix A.2 reports the number of observations at each step of the sample construction.

2.3.1 Variables

Our main outcome is log annual earnings from the main employer, and wage growth is the change in log earnings over horizons of 1, 3, 5, and 10 years. Future earnings are taken from years in which the worker is employed full-time. A worker is a stayer at horizon s if the main employer in year $t + s$ is the same as in year t . A worker is a mover if they move directly to a new employer, whose employment spell starts no earlier than the last year at the old firm. Earnings are in nominal NTD; year fixed effects absorb inflation in the regressions.

Firm controls are measured from the business income tax return and the matched worker records. They are an indicator for location in one of the six special municipalities, industry (19 sections of Taiwan’s Standard Industrial Classification), log number of workers, mean worker age, share of men among workers, and return on capital (net profit over capital). To separate training from firm pay, we control for the firm’s leave-one-out mean log wage, which excludes the worker’s own earnings. As a robustness check, we replace it with an AKM firm wage premium ([Abowd et al. 1999](#)), estimated on the largest connected set of firms (as in [Chen et al. 2026b](#)).

3 Which Firms Train, and Who Works for Them?

We begin by characterizing which firms train and then document selection into training firms by age and gender.

3.1 Characterization of Training Firms

Business tax records and balance sheets allow us to measure several firm characteristics at the firm-year level. We examine the relationship between training expenses and these characteristics at both the extensive and intensive margins.

We regress whether a firm trains, and how much it trains, on firm size, urban status, mean log wage, employee benefits, and R&D expenses, conditional on industry and year fixed effects (Table 1). Column (1) reports the extensive margin: whether a firm provides any training. Larger firms and firms in urban areas are more likely to train. Training is also positively correlated with characteristics of “better employers”: firms that pay higher wages, spend on R&D, and provide employee benefits are more likely to train. Yet these characteristics, together with industry fixed effects, explain only about 21% of the variation in training provision across firms. Training thus appears to be a distinct feature of employers that wages and other amenities do not capture.

TABLE 1. Relationship between Training and Other Firm Characteristics

Dep. Var.:	(1) Any Training	(2) Log Training per Worker	(3) High-Training
log(firm size)	0.0864*** (0.0003)	-0.4180*** (0.0036)	-0.0719*** (0.0008)
Urban	0.0046*** (0.0007)	0.1349*** (0.0095)	0.0364*** (0.0021)
Mean log wage	0.0094*** (0.0002)	0.1077*** (0.0047)	0.0162*** (0.0011)
Has employee benefits	0.1527*** (0.0007)		0.0457*** (0.0015)
Has R&D fee	0.2893*** (0.0023)		0.0765*** (0.0032)
log(empl. ben. + 1)		0.0549*** (0.0008)	
log(R&D fee + 1)		0.0566*** (0.0014)	
Mean of dep. var.	0.15944	6.3730	0.24896
<i>N</i>	7,590,772	1,210,280	1,210,280
Adj. <i>R</i> ²	0.21274	0.14157	0.07988

Notes: This table reports regressions of training on observable firm characteristics at the firm-year level. All training measures use a three-year rolling window (Section 2.2). Column (1) covers all firm-years; the dependent variable is an indicator for any training. Columns (2) and (3) are restricted to firms with positive training. In column (2), the dependent variable is log training expenses per worker, and employee benefits and R&D expenses enter as log per-worker amounts. In column (3), the dependent variable is an indicator for high-training firms, defined as the top 25% of training firms ranked by training expenses per worker, and benefits and R&D enter as indicators. All columns control for return on capital, average employee age, male share of employees, and industry and year fixed effects. Standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Columns (2) and (3) restrict the sample to firms with any training and relate the intensity of training to the same set of characteristics. Higher wages, employee benefits, and R&D expenses remain positively associated with training. Firms with employee benefits are 4.6 percentage points (p.p.) more likely to be a high-training firm, an 18% increase relative to the mean. Firms with R&D expenses, a proxy for reliance on new technologies and technical skills, are 7.6 p.p. or 31% more likely to be in the high-training tier. The one sign reversal is firm size: among firms that train, larger firms spend less per worker, with an elasticity of -0.42 , and are less likely to be in the high tier. Together, these characteristics explain less than 15% of the variation in training intensity across employers, further evidence that training is a distinct feature of firms.

We next look more closely at the relationship between firm pay and training, which is positive at both the extensive and intensive margins. The left panel of Figure 1 plots mean log training per worker against mean log wage in a binned scatter plot. Training intensity is roughly flat across the bottom three quartiles of the wage distribution and rises steeply in the top quartile.

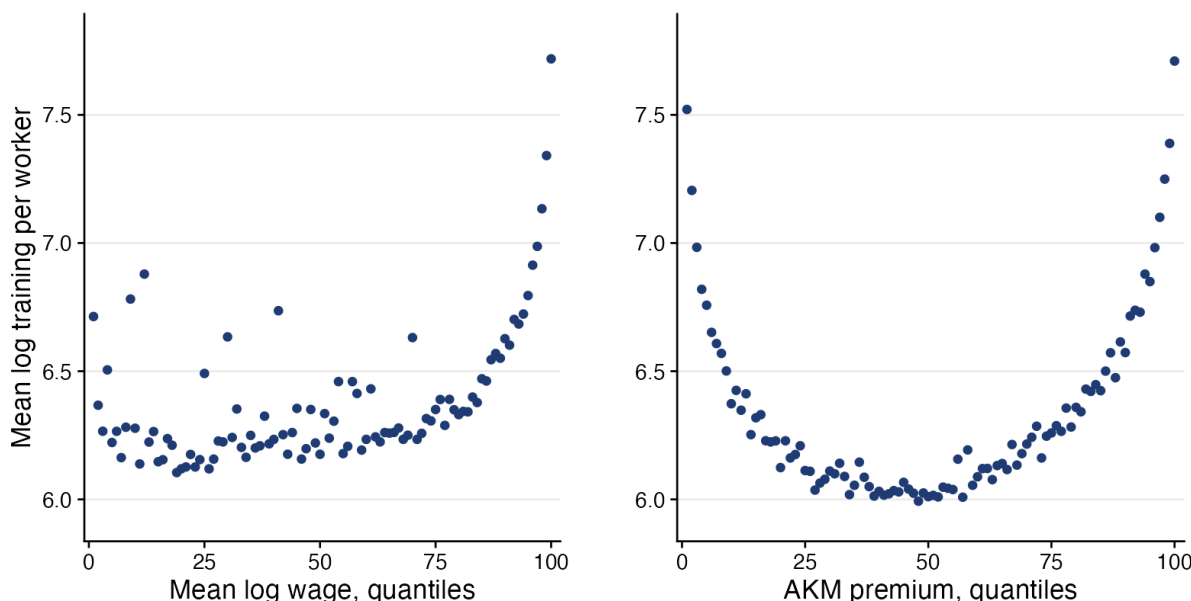


FIGURE 1. Training per worker, mean log wage, and AKM firm fixed effect

Notes: This figure shows the relationship between mean log training per worker and two measures of firm pay: mean log wage in the left panel and the AKM firm fixed effect in the right panel. The sample comprises 1.2 million firm-year observations with positive training expenses over a three-year rolling window. We sort firm-years into 100 equal-sized bins by the pay measure on the x -axis and plot the mean training intensity within each bin on the y -axis. To estimate the AKM firm effects, we regress log annual earnings of full-time workers on worker, firm, and year fixed effects. The firm effects are identified only within the largest set of firms connected by worker moves, so firms outside that set drop out: 40,580 of the 1.2 million firm-years, or 3.4 percent.

Mean log wage is shaped by both the selection of workers into the firm and how the firm sets pay for all its employees. To separate the two, we estimate an AKM model (Abowd, Kramarz, and Margolis 1999) on the largest connected set of firms and use the estimated firm fixed effects as a measure of the firm-specific wage premium (Card, Cardoso, Heining, and Kline 2018).

The right panel of Figure 1 shows that the relationship with the AKM premium is U-shaped. Training per worker is highest at firms with the lowest and the highest wage premiums and lowest at firms in the middle. The high training at low-premium firms is consistent with workers paying for general training through lower current pay in exchange for faster wage growth later (Becker 1964). High-premium firms, in contrast, worry less about turnover and are willing to provide more

training because they expect their employees to stay longer ([Acemoglu and Pischke 1999](#)). They may also train more when training builds firm-specific human capital that complements the firm's capital.

To better understand the U-shaped relationship, we will compare the industry composition at the two tails of the AKM distribution and use external survey data to examine whether low- and high-wage firms spend on similar types of training.

3.2 Selection into Training by Age and Gender

Having shown that training is a distinct attribute of the firm, not fully explained by wages and proxies for productivity and amenities, we turn to study who works at training firms. About 16% of firm-years involve any training, but because training firms are larger, 57.6% of worker-years are spent at them.

Figure 2 plots the share of full-time workers at no-, low-, and high-training firms by age and gender. Young women are more likely than young men to work at a training firm. At age 22, 65% of women work at a (low- or high-) training firm, compared with 58% of men. The gap closes by age 30, when about 73% of both women and men are at a training firm, and reverses in the early 30s. From then on, the share of women at no-training firms rises faster than the share of men, and by age 50 it is 40% for women and 38% for men.

Women are also more likely than men to work at a high-training firm at every age until the late 40s. The share of women at high-training firms peaks at 21% around age 31 and declines to 16% by age 50. The share for men stays between 16% and 18% from the late 20s on. The female lead at high-training firms is therefore largest in the late 20s and early 30s, around the median age at first birth for women (31 for the 1980s birth cohorts), and narrows afterward. Training intensity (training expenses per worker) shows similar patterns ([Appendix Figure A.1](#)).

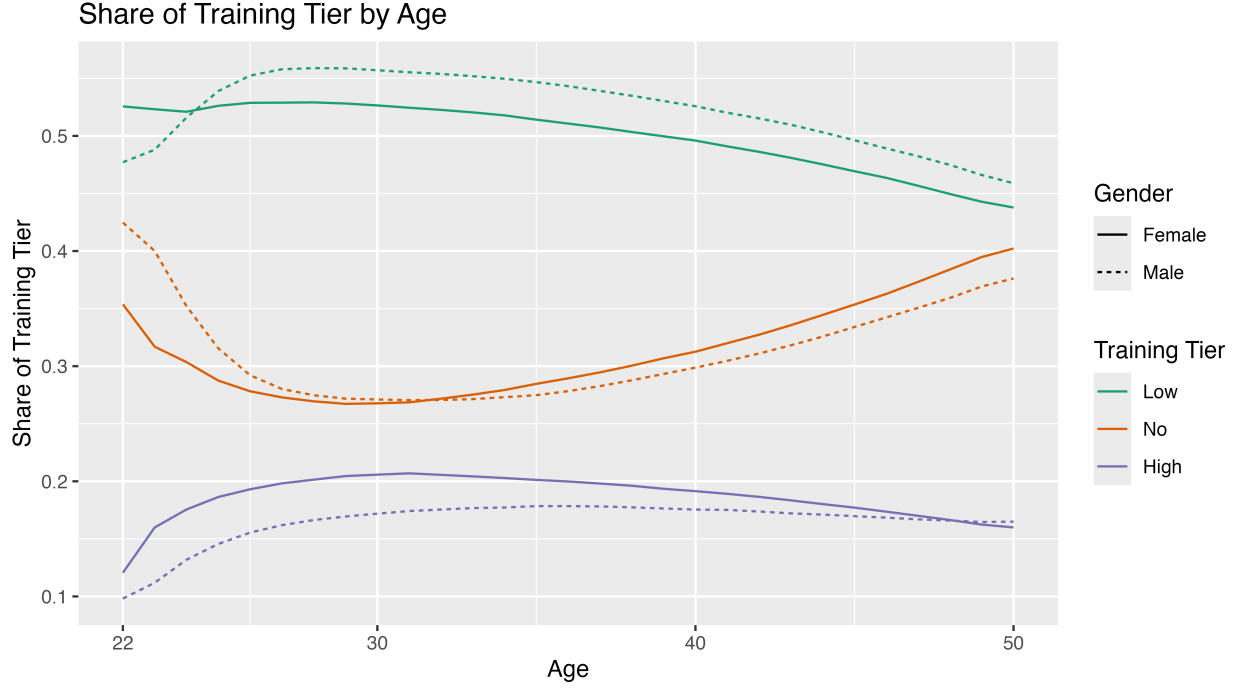


FIGURE 2. Employment at No-, Low-, and High-Training Firms by Age and Gender

Notes: This figure shows the share of full-time women and men employed at no-, low-, and high-training firms at each age from 22 to 50. Firms are classified into the three tiers using a three-year rolling window of training expenses per worker (Section 2.2). The sample is restricted to full-time workers with annual earnings above 12 months of the minimum wage.

These patterns reflect the life cycle rather than differences across cohorts. Appendix Figure A.2 shows the shares of workers at each tier of firms within birth cohorts. Following the 1980–1989 cohorts through their 20s and 30s, we find the same pattern: women’s higher exposure to training firms vanishes in the early 30s, and their lead in exposure to high-training firms shrinks as they age.

Averaged across ages, women are 0.4 p.p. less likely than men to work at a training firm. Among workers at training firms, however, women are more often at high-training firms, as Figure 2 shows. As a result, average training intensity at women’s employers is about 14 log points higher than at men’s.

To examine how much of the gender gaps at the two margins reflects sorting across industries, we apply an Oaxaca–Blinder decomposition (Oaxaca 1973; Blinder 1973):

$$\bar{y}^F - \bar{y}^M = \sum_k \beta_k^F p_k^F - \sum_k \beta_k^M p_k^M = \underbrace{\sum_k \beta_k^M (p_k^F - p_k^M)}_{\text{between industry}} + \underbrace{\sum_k p_k^F (\beta_k^F - \beta_k^M)}_{\text{within industry}}, \quad (1)$$

where p_k^g is the share of workers of gender g in industry k and β_k^g is the mean exposure to training of gender g in that industry, obtained from a regression of exposure on industry indicators estimated separately by gender. The first term in (1) is the gap that would arise if women and men had the same exposure to training within each industry and differed only in their distribution across industries. The second term captures gender differences in sorting to firms with different training within industries.

TABLE 2. Decomposition of the Gender Gap in Exposure to Training

	Any training (extensive margin)	Log training per worker (intensive margin)
Gap (Women-Men)	−0.0044	0.1359
<i>Reference group: Men, as in equation (1)</i>		
Between industry	0.0094	0.1585
Within industry	−0.0138	−0.0225
<i>Reference group: Women</i>		
Between industry	0.0096	0.1483
Within industry	−0.0140	−0.0124

Notes: This table decomposes the gender gap in exposure to training into between- and within-industry components using equation (1). The sample is full-time workers aged 22 to 50. The intensive margin is among workers at firms with positive training. The first panel uses men’s industry means as the reference, as in equation (1); the second uses women’s.

Table 2 shows that industry composition works in women’s favor on both margins. Women are concentrated in industries that train more, such as wholesale, retail, finance, and insurance. Industry composition alone would make women 0.9 p.p. more likely than men to be at a training firm and would raise training per worker at their employers by 16 log points. Within industries, the gaps are negative: women are 1.4 p.p. less likely than men to be at a training firm and are at firms that train 2 log points less per worker. The raw gaps therefore understate the disadvantage women face within industries, consistent with the negative female coefficient in regressions with industry controls (Appendix Table A.1). Using women’s industry means as the reference instead yields the same conclusion.

4 Returns to Training

A long-standing question is how much on-the-job training matters for human capital accumulation and career progression (Brown 1989; Lynch 1998). We study this question by comparing wage

growth across high-, low-, and no-training firms. Given the gender differences in selection into training firms (Section 3.2), we examine whether exposure to training firms benefits women and men differently. We also compare wage growth for stayers and movers. In the short run, we find evidence consistent with Lynch (1991) that employer-provided training is largely firm-specific. Over longer horizons, however, women who leave high-training firms still see stronger wage growth, which suggests that some of the training is portable across employers.

4.1 Differences in Wage Growth Across Training Tiers

We regress log wage growth on the training tier of a worker’s current employer, denoted by $j(i, t)$, conditional on a rich set of observables:

$$\log(w_{i,t+s}) - \log(w_{it}) = \beta_0 + \sum_{k \in \{\text{High, Low}\}} \underbrace{\beta_k \cdot 1[j(i, t) \in k]}_{\text{return to training tier } k} + x'_{it} \gamma + \psi_{\text{industry}(i,t)} + \phi_t + u_{it}, \quad (2)$$

where the training tier k is defined over three-year rolling windows as in Section 2.2, and no-training firms are the omitted category. The controls x_{it} include a cubic in age, which captures life-cycle patterns in wage growth; an urban indicator; and characteristics of the current employer $j(i, t)$: log firm size, mean employee age, male share of employees, and return on capital. Controlling for leave-out mean log wage at the employer allows us to test whether the relationship between employer-provided training and wage growth is fully explained by working for a high-wage firm. Industry and year fixed effects absorb common factors affecting wage growth within industries and over time. The horizon $s \in \{1, 3, 5, 10\}$ captures immediate as well as longer-term career progression.

Table 3 reports the results for one-year wage growth ($s = 1$). In column (1), men at low-training firms experience about the same wage growth as men at no-training firms, while men at high-training firms experience 0.9 log points higher wage growth, roughly 10% of the mean. The bottom panel reports the estimates for women. Women at high-training firms experience about 1 log point higher wage growth than women at no-training firms, and women at low-training firms again show no difference.

TABLE 3. One-Year Log Wage Growth and Training Tier

	All Cohorts			Younger Cohorts		
	(1) All	(2) Stayer	(3) Mover	(4) All	(5) Stayer	(6) Mover
Men						
Low Training	-0.0003 (0.0029)	0.0007 (0.0030)	-0.0163*** (0.0026)	0.0055 (0.0058)	0.0074 (0.0062)	-0.0132*** (0.0035)
High Training	0.0093** (0.0039)	0.0107*** (0.0040)	-0.0237*** (0.0046)	0.0163* (0.0098)	0.0193* (0.0105)	-0.0185*** (0.0049)
Mean Log Wage	-0.0284*** (0.0025)	-0.0157*** (0.0025)	-0.1424*** (0.0024)	0.0093 (0.0058)	0.0281*** (0.0061)	-0.1095*** (0.0029)
<i>N</i>	32,117,795	29,440,093	2,677,702	5,389,428	4,745,838	643,590
Adj. R^2	0.05036	0.04927	0.07208	0.04806	0.05723	0.05910
Mean Dep. Var.	0.08984	0.07821	0.21766	0.18498	0.16553	0.32847
Women						
Low Training	-0.0007 (0.0019)	0.0002 (0.0018)	-0.0182*** (0.0032)	0.0104*** (0.0034)	0.0127*** (0.0036)	-0.0106*** (0.0031)
High Training	0.0099*** (0.0025)	0.0121*** (0.0024)	-0.0298*** (0.0058)	0.0244*** (0.0058)	0.0288*** (0.0061)	-0.0162*** (0.0043)
Mean Log Wage	-0.0275*** (0.0018)	-0.0195*** (0.0017)	-0.1062*** (0.0030)	-0.0070* (0.0037)	0.0058 (0.0039)	-0.0759*** (0.0024)
<i>N</i>	24,858,145	23,063,390	1,794,755	4,449,581	3,907,837	541,744
Adj. R^2	0.05490	0.04931	0.06738	0.05966	0.06510	0.05339
Mean Dep. Var.	0.09194	0.07871	0.26205	0.18528	0.16188	0.35407

Notes: This table reports estimates of equation (2) for one-year log wage growth, $\log(w_{i,t+1}) - \log(w_{it})$, on indicators for working at a low- or high-training firm in year t , estimated separately by gender. The omitted category is no-training firms. Columns (1)–(3) include all workers who earn at least six months of the minimum wage; columns (4)–(6) restrict to younger cohorts born in 1987 or later. Stayers remain with the same employer from t to $t + 1$; movers change employers. All columns control for a cubic in age, mean log wage at the firm, log firm size, urban status, the firm’s mean employee age and male share, return on capital, and industry and year fixed effects. Standard errors are clustered at the worker level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

For workers born in 1987 or later, we observe early career progression more completely than for older cohorts. Column (4) shows that younger men at high-training firms experience 1.6 log points higher wage growth than their counterparts at no-training firms, and about 1.1 log points

higher than those at low-training firms. Younger women at high-training firms experience 2.4 log points higher wage growth than those at no-training firms, and younger women at low-training firms also gain 1 log point relative to no-training firms. Wages of younger workers grow about twice as fast as those of the full sample (18 log points on average, against 9 log points in column (1)), so the returns to working at a high-training firm relative to the mean are similar in columns (1) and (4).

Employment at a high-training firm also predicts stronger wage growth over longer horizons for both men and women (Appendix Tables A.2, A.3, and A.4 for $s = 3, 5$, and 10). Younger women at high-training firms consistently show higher wage growth than their peers at low- or no-training firms. The corresponding estimates for younger men are smaller and imprecise, suggesting that women benefit more from working at high-training firms, especially early in their careers.

Because the regressions control for mean log wage, the relationship between exposure to high-training firms and wage growth is not driven by working at a high-wage firm. We also run a horseshoe between high training and two other firm characteristics: R&D expenses, which proxy for a firm's reliance on new technologies and the skill upgrading they demand, and employee benefits, which proxy for amenities. Appendix Table A.5 reports the results for three-year wage growth among women at training firms. Employment at a high-training firm predicts 1.7–1.8 log points higher wage growth than at a low-training firm, and the estimate is stable across specifications. R&D intensity is positively related to wage growth, but adding it barely moves the coefficient on high training. Employee benefits have no relationship with wage growth.

4.2 Portability of Employer-Provided Training

In a perfectly competitive labor market, firms pay for firm-specific human capital but not for general human capital that workers can carry elsewhere (Becker 1964). In imperfectly competitive labor markets, firms may also invest in general human capital, because search and information frictions reduce turnover and allow firms to capture part of the return (Acemoglu and Pischke 1998; Acemoglu and Pischke 1999; Manning 2003). Our measure of training is based on employer spending and does not distinguish general from specific training. We can, however, examine whether the training is portable by estimating equation (2) separately for stayers and movers: if training raises general human capital, workers who leave high-training firms should also see faster wage growth.

Columns (2) and (3) of Table 3 report the estimates for stayers, who remain with the same employer in $t + 1$, and movers, who are at a new employer in $t + 1$. The one-year wage returns to employment at a high-training firm accrue only to stayers, for both men and women. Movers from low- or high-training firms, in contrast, experience lower wage growth than movers from no-training firms. Movers see large wage gains on average, 22–26 log points in one year against 8 log

points for stayers, but those leaving high-training firms gain 2–3 log points less than those leaving no-training firms. This finding suggests that training does not raise general human capital, at least in the short run.

The same patterns hold for the younger cohorts, shown in columns (5) and (6): stayers at high-training firms experience 2–3 log points higher wage growth in one year, while movers experience about 2 log points lower wage growth, relative to their counterparts at no-training firms. The negative estimates for movers could reflect negative selection: because high-training firms pay higher wages and offer better amenities on average, the workers who leave them may differ from those who stay. In the next draft, we will separate movers who move up the job ladder, to higher-wage or more productive firms, from those who move

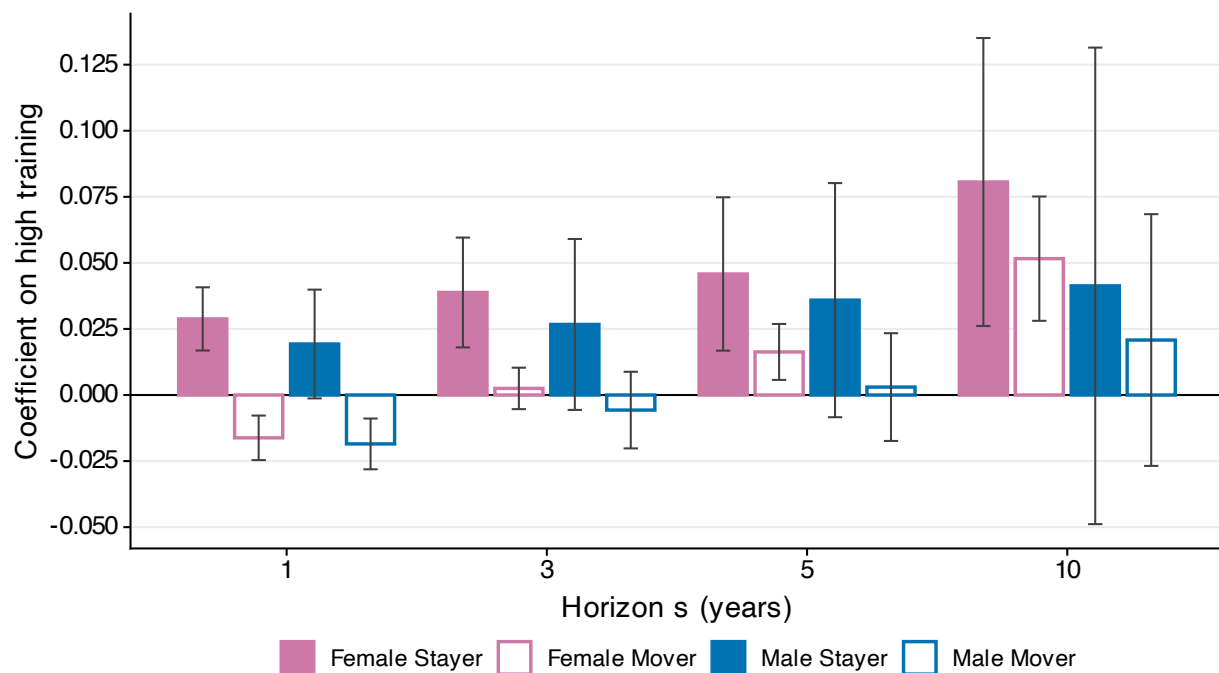


FIGURE 3. Wage Growth at High-Training Firms by Horizon: Younger Cohorts

Notes: This figure plots the coefficient on high training, β_{High} in equation (2), for wage growth $\log(w_{i,t+s}) - \log(w_{it})$ over horizons $s \in \{1, 3, 5, 10\}$, relative to workers at no-training firms. The sample comprises younger cohorts born in 1987 or later. Each bar is from a separate regression by gender and by stayer or mover status; the estimates correspond to columns (5) and (6) of Table 3 and Appendix Tables A.2–A.4. Bars show 95% confidence intervals based on standard errors clustered by firm.

Figure 3 summarizes the coefficients on high training across horizons for younger women and men, by stayer and mover status. Among stayers, the coefficient is larger for women than for men at every horizon, and the gap widens at the ten-year horizon. In the short run, both women and

men who leave high-training firms experience lower wage growth than those who leave no-training firms. We interpret these findings as evidence that, in the short run, on-the-job training primarily raises firm-specific rather than portable human capital, consistent with [Lynch \(1991\)](#).

Over longer horizons, the pattern for movers reverses. Women who leave high-training firms experience 1.6 log points higher wage growth over five years, and 5 log points higher over ten years, than women who leave no-training firms. Men who leave high-training firms also show higher ten-year wage growth than their counterparts from no-training firms, but the estimate is not statistically significant. These findings suggest that having worked at a high-training firm benefits the longer-term career progression of young workers, and more so for women. Through the lens of human capital, this is evidence that employer-provided training raises some general human capital, although the return does not show up in wages immediately.

In summary, working at a high-training firm predicts higher wage growth among stayers over the next decade. The relationship is stronger for younger workers and for women. In the short run, movers out of high-training firms experience lower wage growth than movers out of no-training firms, which may reflect negative selection of workers leaving high-quality employers, an issue we will explore further. Over longer horizons, the relationship between employment at a high-training firm and wage growth is positive even among movers, and it is consistently stronger for women.

5 Motherhood Penalty

Women gain more than men from working at a high-training firm ([Section 4](#)), yet from their early 30s they sort out of training firms, and the industries they work in do not explain the gap ([Section 3.2](#)). Women thus leave the firms where they would gain the most, and they start to leave around the median age at first birth. In this section, we ask how workers fare around the birth of their first child, and whether working at a training firm before the birth changes that.

Childbirth can reduce women's training in two ways. First, time out of work, or in part-time work, to care for a child interrupts training as well as work. Mothers receive less training after the first birth ([Ma et al. 2025](#)), and part-time work builds far less human capital than full-time work, so the lost experience lowers their wages for years afterwards ([Blundell et al. 2021](#)). Second, women who expect to work less after a birth have less reason to invest in training or to join a firm that trains, because career interruptions shorten the period over which training pays off ([Royalty 1996](#); [Blundell et al. 2021](#)). Training can also keep mothers in work. A firm that has trained a worker has reason to keep her, and a worker whose wages grow faster has more to lose by leaving. We find that mothers who worked at a high-training firm before the birth lose less employment and earnings than mothers who worked at a low-training firm, while fathers' outcomes differ little by tier.

5.1 Empirical Strategy

We follow the event-study design of [Kleven et al. \(2019\)](#), the standard approach in the child-penalty literature. Although fertility is a choice, labor market outcomes change sharply at the first birth, while their other determinants should evolve smoothly, so the break identifies the effect of the child ([Kleven et al. 2019](#)). The design also traces out how these effects evolve over time, which lets us compare the paths of mothers from different training tiers. For worker i in calendar year t , let $l = t - b_i$ be event time, where b_i is the year of i 's first birth. We estimate

$$Y_{it} = \beta_0 + \sum_{l \neq -2} \gamma_l \cdot D_{it}^{(l)} + x'_{it} \lambda + \psi_{\text{industry}(i,t)} + \phi_t + v_{it}, \quad (3)$$

where $D_{it}^{(l)}$ indicates that year t is l years from i 's first birth. The year $l = -2$ is omitted, so each γ_l measures the change in the outcome relative to two years before the birth.² The controls x_{it} include age fixed effects and characteristics of the worker's employer at $l = -2$: urban location, industry fixed effects, the leave-out mean log wage, log size, mean employee age, share of men among employees, and return on capital. Industry and year fixed effects absorb common factors affecting labor market outcomes within industries and over time. Age and year effects are identified from variation in the age at first birth, as in [Kleven et al. \(2019\)](#).

We estimate equation (3) separately by gender, and, for Figures 5 and 6, separately by the training tier of the worker's employer at $l = -2$. Assigning the tier before the pregnancy keeps it from responding to the birth itself. Each regression has its own baseline, so the tier comparisons are about changes from that baseline. The sample consists of parents whose first child was born between 2010 and 2018, who were at least 25 at the birth, who are alive at least five years after it, and who worked at least part time at $l = -2$. Workers who never have a child are not in the sample. The outcomes are full-time employment (annual earnings from the main employer of at least 12 months of the minimum wage), any employment (positive earnings), and log annual earnings. The earnings sample keeps parents with positive earnings in at least one year from $l = 1$ to 5, and the log drops the years with zero earnings.

²[Kleven et al. \(2019\)](#) omit $l = -1$. We use $l = -2$ because the year before the birth is the pregnancy year, in which employment and earnings already start to fall, and because the employer whose training tier we use is measured in that same year.

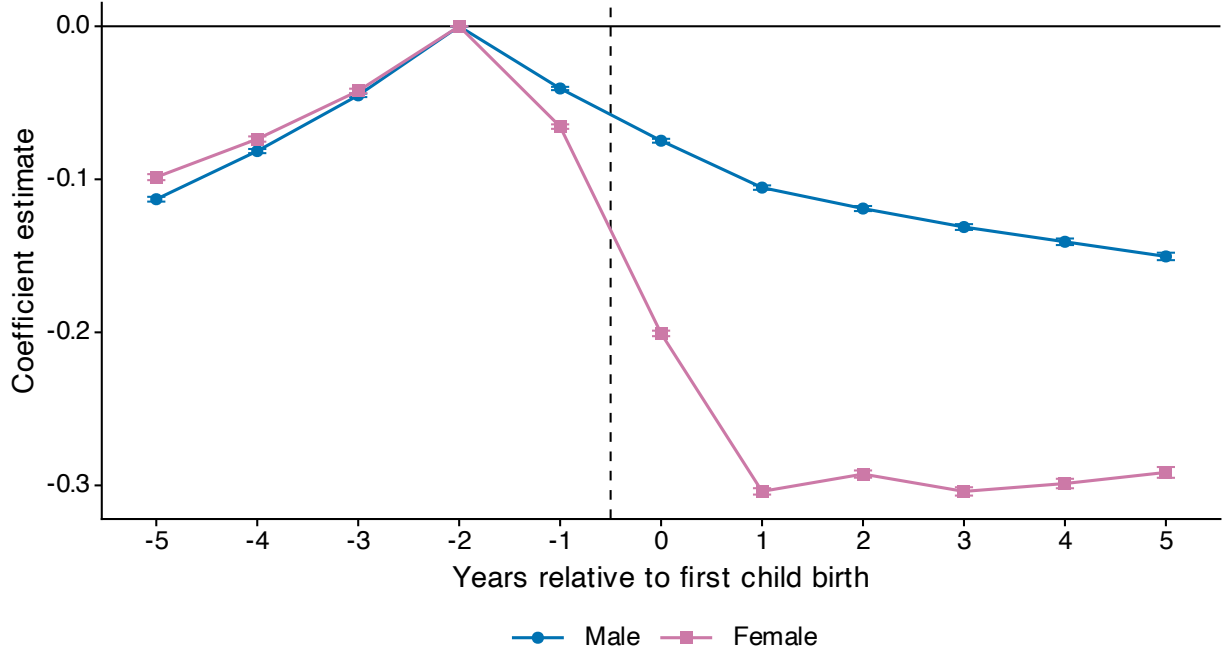


FIGURE 4. Full-Time Employment around First Childbirth by Gender

Notes: This figure plots the event-time coefficients $\hat{\gamma}_l$ from equation (3), with 95% confidence intervals based on standard errors clustered by individual. Coefficients are relative to $l = -2$, two years before the first birth, and the dashed vertical line separates the year of the birth from the years before it. Equation (3) is estimated separately for mothers and fathers. The outcome is an indicator for full-time employment: annual earnings from the main employer of at least 12 times the monthly minimum wage. The sample consists of parents whose first child was born between 2010 and 2018, who were at least 25 at the birth, and who worked at $l = -2$ (annual earnings from the main employer of at least six times the monthly minimum wage). Controls are age and calendar-year fixed effects and characteristics of the $l = -2$ employer: urban location, industry fixed effects, leave-out mean log wage, log size, mean employee age, share of men among employees, and return on capital. The regressions use event times -13 to 13 ; the figure shows -5 to 5 . Data: income tax records and household registry, Fiscal Information Agency, 2005–2023.

5.2 Child Penalty in Taiwan

Figure 4 replicates the standard result. Relative to $l = -2$, women’s full-time employment falls by 20 percentage points (p.p.) in the year of the birth and by 30 p.p. the year after, and it does not recover. Men’s full-time employment falls by 7.5 p.p. at birth and 15 p.p. by $l = 5$. Most of the decline for men is mechanical. Because the sample must be employed at $l = -2$, employment for both genders peaks near $l = -2$ and drifts down afterwards, and the negative pre-period coefficients reflect the same selection.

Following Kleven et al. (2019), we scale the gap between men’s and women’s coefficients by women’s counterfactual outcome and evaluate it five years after the birth, the last year in our

figures (Kleven et al. 2019 use ten years). The penalties are similar in size to those in the literature: our long-run earnings penalty is close to the 20% that Kleven et al. (2019) find for Denmark. For Taiwan, Chen et al. (2026a) use health-insurance payroll records and a matched group of never-pregnant women and also find large and persistent earnings losses after the first birth, part of which begins with the pregnancy itself. Across countries, the Child Penalty Atlas places Taiwan’s penalty above that of the United States and below those of Japan and Korea (Kleven et al. 2025).³

Fathers’ outcomes show no clear break at the birth. Their full-time employment after the birth roughly mirrors its path before it, falling by about as much in the years after $l = -2$ as it rose in the years leading up to it, which is the pattern the selection on employment at $l = -2$ produces. As in Kleven et al. (2019), the fathers’ estimates are mainly a comparison for mothers.

³Our employment penalties are smaller than the Atlas estimate for Taiwan, and the two differ in design and data. We condition on employment at $l = -2$, which keeps mothers who were attached to the labor market before the birth; the Atlas covers all women. The Atlas also uses pseudo-event studies on cross-sectional surveys, while we follow the same parents over time, and its employment measure includes self-employment, which our tax records do not cover. For earnings, we observe annual earnings but not hours or months worked, so a drop in earnings combines fewer hours, fewer months, and lower pay. Parents on parental leave appear to have low or no earnings: they may take up to two years of job-protected unpaid leave before the child turns three, and the leave allowance, paid by the employment insurance system rather than the employer, does not appear in the salary records.

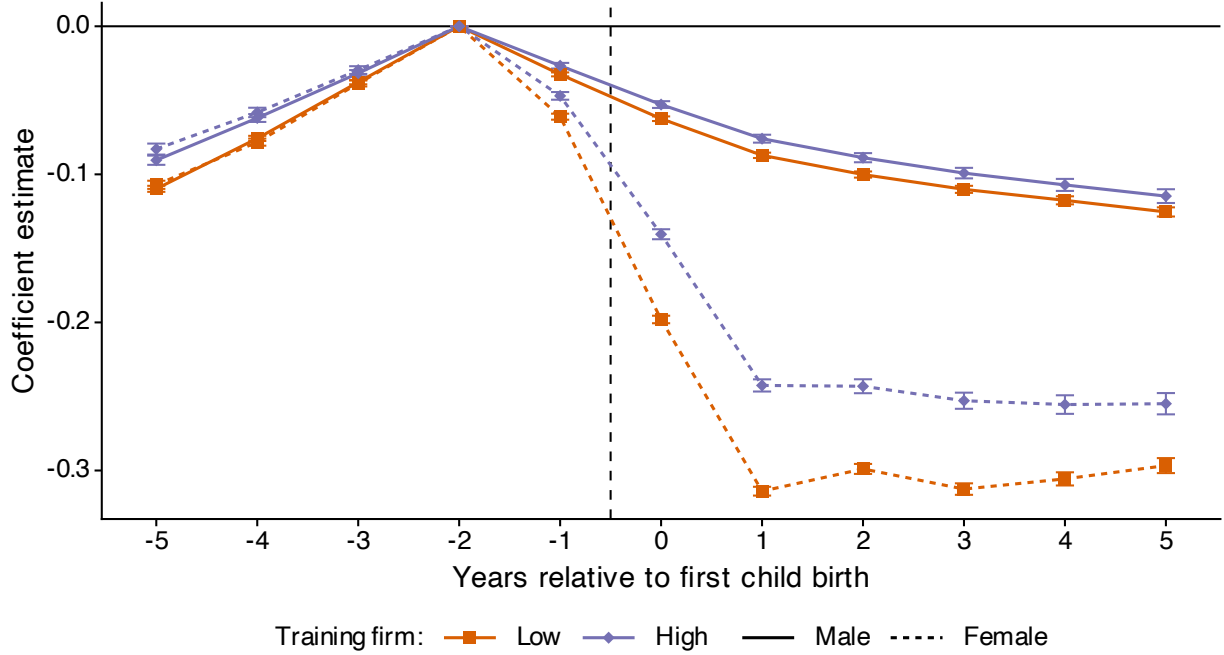


FIGURE 5. Full-Time Employment around First Childbirth by Training Tier

Notes: This figure plots the event-time coefficients $\hat{\gamma}_l$ from equation (3), with 95% confidence intervals based on standard errors clustered by individual. Coefficients are relative to $l = -2$, two years before the first birth, and the dashed vertical line separates the year of the birth from the years before it. Each line comes from a separate regression by gender and by the training tier of the worker's employer at $l = -2$. High-training firms are those at or above the 75th percentile of three-year average training expense per worker among firms with positive training; low-training firms are the other firms with positive training (Section 2.2). Estimates for workers at no-training firms are not shown. Colors and markers distinguish tiers; solid lines are fathers and dashed lines are mothers. The outcome is an indicator for full-time employment: annual earnings from the main employer of at least 12 times the monthly minimum wage. The sample consists of parents whose first child was born between 2010 and 2018, who were at least 25 at the birth, and who worked at $l = -2$ (annual earnings from the main employer of at least six times the monthly minimum wage). Controls are age and calendar-year fixed effects and characteristics of the $l = -2$ employer: urban location, industry fixed effects, leave-out mean log wage, log size, mean employee age, share of men among employees, and return on capital. The regressions use event times -13 to 13 ; the figure shows -5 to 5 . Data: income tax records and household registry, Fiscal Information Agency, 2005–2023.

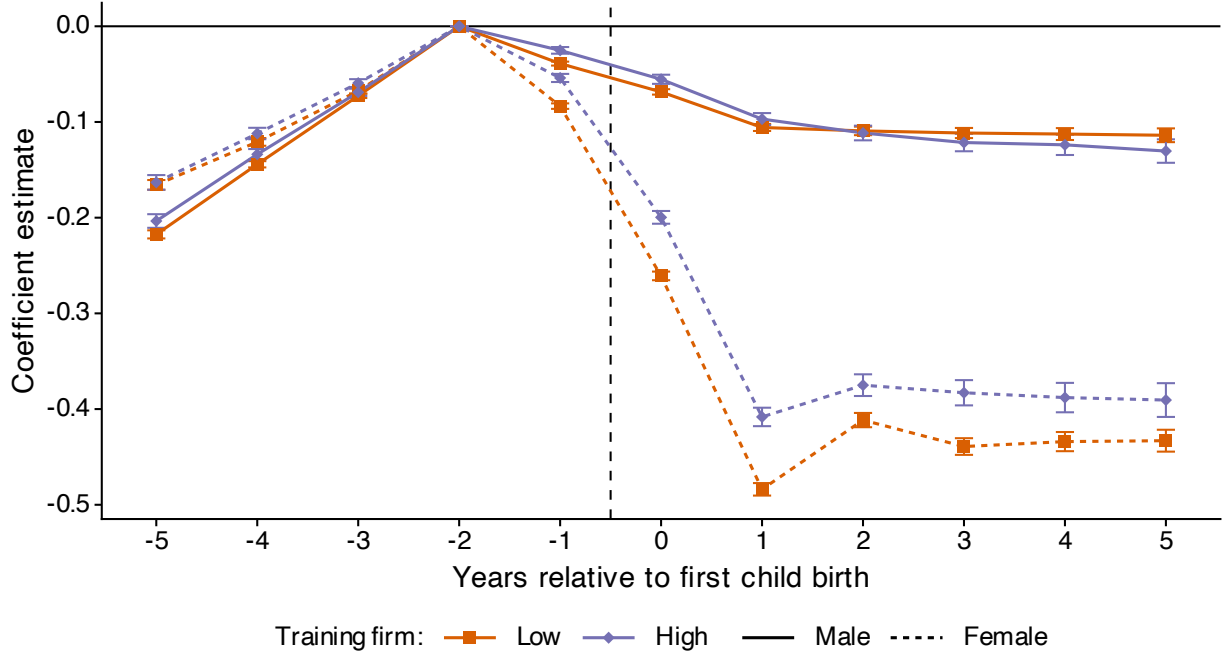


FIGURE 6. Log Earnings around First Childbirth by Training Tier

Notes: This figure plots the event-time coefficients $\hat{\gamma}_l$ from equation (3), with 95% confidence intervals based on standard errors clustered by individual. Coefficients are relative to $l = -2$, two years before the first birth, and the dashed vertical line separates the year of the birth from the years before it. Each line comes from a separate regression by gender and by the training tier of the worker's employer at $l = -2$. High-training firms are those at or above the 75th percentile of three-year average training expense per worker among firms with positive training; low-training firms are the other firms with positive training (Section 2.2). Estimates for workers at no-training firms are not shown. Colors and markers distinguish tiers; solid lines are fathers and dashed lines are mothers. The outcome is log annual earnings from the main employer, so years with zero earnings drop out. The sample consists of parents whose first child was born between 2010 and 2018, who were at least 25 at the birth, and who worked at $l = -2$ (annual earnings from the main employer of at least six times the monthly minimum wage). Controls are age and calendar-year fixed effects and characteristics of the $l = -2$ employer: urban location, industry fixed effects, leave-out mean log wage, log size, mean employee age, share of men among employees, and return on capital. The regressions use event times -13 to 13 ; the figure shows -5 to 5 . The sample is further restricted to parents with positive earnings in at least one year from $l = 1$ to 5 . Data: income tax records and household registry, Fiscal Information Agency, 2005–2023.

5.3 Child Penalty by Training Tier

Figure 5 splits the sample by the training tier of the $l = -2$ employer. Relative to their own pre-birth levels, mothers who worked at a high-training firm before the birth lose less full-time employment than mothers who worked at a low-training firm. This comparison alone does not show that the firm causes the smaller loss. The year after the birth, full-time employment of mothers from high-training firms is 7 p.p. higher than that of mothers from low-training firms. The gap shrinks to 4

p.p. by $l = 5$. For fathers the gap is about 1 p.p. throughout. Before the birth, the paths of the two tiers run parallel. From $l = -5$ to $l = -1$, the gap in full-time employment between tiers is at most 2.4 p.p. for mothers and 1.9 p.p. for fathers, and it moves in the same way for both, although every parent in the sample is employed at $l = -2$. The gap is larger and more persistent on the extensive margin: any employment is 6 p.p. higher for mothers from high-training firms at $l = 1$ and still 5 p.p. higher at $l = 5$, against 1 to 2 p.p. for fathers.

Figure 6 shows the same comparison for log earnings among those still working. Mothers from high-training firms see a smaller earnings drop, by 8 log points at $l = 1$ and 4 log points at $l = 5$. For fathers the tiers do not differ.

Mothers from high-training firms are also more likely to stay with their pre-birth employer (Appendix Figure A.4). At $l = 1$, the share still with the $l = -2$ employer is 11.6 p.p. higher for mothers from high-training firms than for mothers from low-training firms. Retention does not fully explain these mothers' stronger attachment, however. Fathers from high-training firms also stay with their employer more often, by 8.0 p.p. at $l = 1$, yet their full-time employment differs from that of fathers at low-training firms by only 1.1 p.p.

In summary, the child penalty in Taiwan is similar in size to those documented elsewhere: five years after the first birth, the earnings penalty is about 20%, close to the estimate for Denmark. Mothers who worked at a high-training firm before the birth stay more attached to the labor market and see a smaller earnings drop than mothers from low-training firms, while fathers' outcomes differ little by tier.

These comparisons are descriptive. In [Blundell et al. \(2021\)](#), most of the wage cost of motherhood comes from the experience mothers lose in part-time work and out of work. Our results suggest that the first step, whether a mother stays in full-time work at all, differs with her employer's training. Other employer attributes also matter for the penalty: it is larger at firms with less generous pay ([Casarico and Lattanzio 2023](#)) and smaller where employers provide childcare ([Costas-Fernández et al. 2026](#)), but not smaller at female-dominated or female-led firms ([Almond et al. 2023](#)). Mothers who work at a high-training firm before the birth may also differ from other mothers in ways that affect their attachment to work after it.

6 Conclusion

Using matched employer-employee data and business tax filings from Taiwan, we study the role of employer-provided training in wage growth and in how workers fare around the birth of their first child. We identify training firms from the training expenses reported in business tax filings. This approach measures training consistently for the whole economy of Taiwan, which has 23.4 million people, more than Denmark, Norway, and Sweden combined.

Firms that provide training pay higher wages, spend more on R&D, and offer better employee

benefits. Conditional on these job attributes, substantial variation in training intensity remains within industries. We therefore interpret training as a distinct attribute of the employer that pay and amenities do not absorb. We will continue to investigate the U-shaped relationship between the firm-specific wage premium and training intensity.

We find important gender differences in exposure to training and in its returns. Women are more likely than men to start at training firms when they are young, but they are less likely to work at training firms after their early 30s, and their lead at high-training firms also shrinks as they age. Linking exposure to training firms to wage growth, we find that women at high-training firms experience higher wage growth than men. Over five or ten years, younger women who leave high-training firms also see higher wage growth than women who leave low- or no-training firms, suggesting that at least some employer-provided training builds general human capital. For men, the returns to training accrue mostly to stayers.

The child penalty in Taiwan is similar in size to those documented elsewhere, but it is smaller for mothers from high-training firms. After the birth, they are more likely to remain in full-time employment than mothers from low-training firms, and among those still working, their earnings fall by less. Fathers' outcomes differ little by tier. Higher retention at high-training firms accounts for part of the difference. Mothers whose wages grow faster at high-training firms may have more to lose by leaving, or women who plan to stay in the labor force after childbirth may choose high-training firms in the first place. We will continue to examine these mechanisms.

Together, these findings suggest that women benefit more than men from exposure to high-training firms. Policies that increase employer-provided training, especially at firms that hire women, could help narrow the gender wage gap and the child penalty.

References

- John M Abowd, Francis Kramarz, and David N Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999.
- Daron Acemoglu and Jorn-Steffen Pischke. Why do firms train? theory and evidence. *The Quarterly Journal of Economics*, 113(1):79–119, 1998.
- Daron Acemoglu and Jorn-Steffen Pischke. Beyond Becker: Training in imperfect labour markets. *The Economic Journal*, 109(453):F112–F142, 1999.
- Douglas Almond, Yi Cheng, and Cecilia Machado. Large motherhood penalties in US administrative microdata. *Proceedings of the National Academy of Sciences*, 120(29):e2209740120, 2023.
- Gary Becker. *Human Capital*. University of Chicago Press, 1964.
- Dan A. Black, Lars Skipper, and Jeffrey A. Smith. Chapter 5 - firm training. volume 7 of *Handbook of the Economics of Education*, pages 287–468. Elsevier, 2023.
- Francine D. Blau and Lawrence M. Kahn. The gender wage gap: Extent, trends, and explanations. *Journal*

- of Economic Literature*, 55(3):789–865, 2017.
- Alan S. Blinder. Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4):436–455, 1973.
- Richard Blundell, Monica Costa Dias, David Goll, and Costas Meghir. Wages, experience, and training of women over the life cycle. *Journal of Labor Economics*, 39(S1):S275–S315, 2021.
- James N. Brown. Why do wages increase with tenure? on-the-job training and life-cycle wage growth observed within firms. *The American Economic Review*, 79(5):971–991, 1989.
- David Card, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):S13–S70, 2018.
- David Card, Francesco Devicienti, Mariacristina Rossi, and Andrea Weber. The gender gap in career trajectories: Do firms matter? Working Paper 33730, National Bureau of Economic Research, May 2025.
- Alessandra Casarico and Salvatore Lattanzio. Behind the child penalty: Understanding what contributes to the labour market costs of motherhood. *Journal of Population Economics*, 36(3):1489–1511, 2023.
- Kuan-Ming Chen, Yen-Fang Chien, Hui Ding, and Warn N. Lekfuangfu. Miscarriage and delays in child-bearing: Implications for motherhood penalties. Working paper, National Taiwan University, 2026a.
- Kuan-Ming Chen, Lin-Tung Tsai, and Hau-Hung Yang. Labor market sorting in taiwan. *Taiwan Economic Review*, 54(2):267–289, 2026b. doi: 10.6277/TER.202606_54(2).0003.
- Julián Costas-Fernández, Sebastian Findeisen, Anna Raute, and Uta Schönberg. Family-friendly workplace policies. CEPR Discussion Paper 21499, Centre for Economic Policy Research, 2026.
- Henrik Kleven, Camille Landaïs, and Jakob Egholt Søgaaard. Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4):181–209, 2019.
- Henrik Kleven, Camille Landaïs, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller. Do family policies reduce gender inequality? evidence from 60 years of policy experimentation. *American Economic Journal: Economic Policy*, 16(2):110–149, 2024.
- Henrik Kleven, Camille Landaïs, and Gabriel Leite-Mariante. The child penalty atlas. *The Review of Economic Studies*, 92(5):3174–3207, 2025.
- Lisa M. Lynch. The role of off-the-job vs. on-the-job training for the mobility of women workers. *The American Economic Review*, 81(2):151–156, 1991.
- Lisa M. Lynch. A needs analysis of training data: What do we want, what do we have, can we ever get it? In John Haltiwanger, Marilyn E. Manser, and Robert Topel, editors, *Labor Statistics Measurement Issues*, volume 60 of *Studies in Income and Wealth*, chapter 12. University of Chicago Press, Chicago, 1998.
- Xiao Ma, Alejandro Nakab, Camila Navajas-Ahumada, and Daniela Vidart. The motherhood training penalty. Working papers 2025-04, University of Connecticut, Department of Economics, Apr 2025.
- Alan Manning. *Human Capital and Training*, pages 301–322. Princeton University Press, 2003.
- Ministry of Finance. How Are Training Expenses of Profit-Seeking Enterprises Recognized? Tax Q&A No. 2242, 2017. URL <https://www.etax.nat.gov.tw/etwmain/tax-info/>

- [understanding/tax-q-and-a/national/profit-seeking-enterprise-income-tax/operating-expense-calculation/PVPLDxq](#). Ministry of Finance Taxation Portal. Updated March 20, 2017. Accessed October 19, 2026.
- Ministry of Finance. Regulations Governing Assessment of Profit-Seeking Enterprise Income Tax, 2023. URL <https://law.moj.gov.tw/LawClass/LawSingle.aspx?pcode=G0340051&flno=86-1>. Article 86-1, Training Expenses. Last amended December 11, 2023. Laws and Regulations Database of the Republic of China (Taiwan). Accessed October 19, 2026.
- Ministry of Finance. Income Tax Act, 2026. URL <https://law.moj.gov.tw/LawClass/LawAll.aspx?pcode=G0340003>. Laws and Regulations Database of the Republic of China (Taiwan). Accessed October 25, 2026.
- Ministry of Labor. Yearbook of Labor Statistics, 2024, 2025a. URL <https://statdb.mol.gov.tw/html/year/year13/33080.pdf>. Table 3-8, Status of Collective Agreement and Labor-management Committees of Enterprises by Area, end of 2024. Accessed October 24, 2026.
- Ministry of Labor. Yearbook of Labor Statistics, 2024: Analysis, Section 3-1, Labor Relations, 2025b. URL <https://statdb.mol.gov.tw/html/year/year13/3-1%20%E5%8B%9E%E5%8B%95%E9%97%9C%E4%BF%82.pdf>. In Chinese. Accessed October 25, 2026.
- Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709, 1973.
- Anne Beeson Royalty. The effects of job turnover on the training of men and women. *Industrial and Labor Relations Review*, 49(3):506–521, 1996.
- Margaret Stevens. A theoretical model of on-the-job training with imperfect competition. *Oxford Economic Papers*, 46(4):537–62, 1994.
- World Bank. Population, Total (SP.POP.TOTL), 2025. URL <https://data.worldbank.org/indicator/SP.POP.TOTL>. World Development Indicators. Denmark, Norway, and Sweden, 2024. Accessed October 25, 2026.

A Appendix

A.1 Appendix Tables

TABLE A.1. Selection into Training by Gender and Parenthood

	Any training		Log training per worker	
	(1)	(2)	(3)	(4)
Female	-0.0127*** (0.0003)	0.0124*** (0.0004)	-0.0934*** (0.0012)	0.0103*** (0.0015)
Kid		-0.0076*** (0.0004)		0.0854*** (0.0016)
Female $\times$ Kid		-0.0480*** (0.0005)		-0.2140*** (0.0022)
<i>N</i>	67,298,077	67,298,077	46,685,863	46,685,863

Notes: This table reports worker-level regressions of exposure to training on gender and parenthood. The sample is full-time workers aged 22 to 50 at firms with a three-year training tier. In columns (1)–(2), the dependent variable is an indicator for working at a low- or high-training firm; columns (3)–(4) restrict to workers at training firms, and the dependent variable is log training expenses per worker. Kid is an indicator for having had a first child by year t . All columns control for a cubic in age, urban status, and industry and year fixed effects. Standard errors clustered by worker in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.2. Three-Year Log Wage Growth and Training Tier

	All Cohorts			Younger Cohorts		
	(1) All	(2) Stayer	(3) Mover	(4) All	(5) Stayer	(6) Mover
Men						
Low Training	-0.0049 (0.0045)	-0.0004 (0.0049)	-0.0145*** (0.0034)	0.0029 (0.0092)	0.0082 (0.0106)	-0.0065 (0.0067)
High Training	0.0109* (0.0056)	0.0174*** (0.0060)	-0.0189*** (0.0051)	0.0166 (0.0132)	0.0267 (0.0165)	-0.0057 (0.0074)
Mean Log Wage	-0.0396*** (0.0037)	-0.0093** (0.0041)	-0.1246*** (0.0025)	0.0141* (0.0074)	0.0631*** (0.0095)	-0.0644*** (0.0030)
<i>N</i>	26,723,787	20,579,717	6,144,070	3,766,276	2,388,280	1,377,996
Adj. R^2	0.07760	0.06525	0.09236	0.05682	0.07443	0.05133
Mean Dep. Var.	0.18564	0.15383	0.29220	0.36905	0.30857	0.47387
Women						
Low Training	-0.0066** (0.0033)	-0.0030 (0.0036)	-0.0124*** (0.0026)	0.0089* (0.0052)	0.0169*** (0.0065)	0.0009 (0.0028)
High Training	0.0121*** (0.0042)	0.0203*** (0.0044)	-0.0155*** (0.0049)	0.0238*** (0.0085)	0.0388*** (0.0106)	0.0025 (0.0040)
Mean Log Wage	-0.0418*** (0.0031)	-0.0223*** (0.0033)	-0.0995*** (0.0027)	-0.0113** (0.0051)	0.0204*** (0.0065)	-0.0536*** (0.0021)
<i>N</i>	20,442,543	16,040,260	4,402,283	3,102,717	1,904,536	1,198,181
Adj. R^2	0.07386	0.05268	0.08107	0.06646	0.07122	0.05536
Mean Dep. Var.	0.18078	0.14785	0.30075	0.34276	0.28248	0.43857

Notes: This table reports regressions of one-year log wage growth, $\log(w_{i,t+3}) - \log(w_{it})$, on indicators for working at a low- or high-training firm in year t , estimated separately by gender. The omitted category is no-training firms. See the notes under Table 3 for sample and controls. Standard errors are clustered at the worker level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.3. Five-Year Log Wage Growth and Training Tier

	All Cohorts			Younger Cohorts		
	(1) All	(2) Stayer	(3) Mover	(4) All	(5) Stayer	(6) Mover
Men						
Low Training	-0.0072 (0.0053)	-0.0012 (0.0061)	-0.0124*** (0.0040)	0.0029 (0.0117)	0.0058 (0.0150)	-0.0019 (0.0089)
High Training	0.0137** (0.0070)	0.0238*** (0.0079)	-0.0134** (0.0059)	0.0214 (0.0158)	0.0359 (0.0226)	0.0030 (0.0104)
Mean Log Wage	-0.0464*** (0.0044)	-0.0092* (0.0054)	-0.1141*** (0.0029)	0.0237*** (0.0081)	0.0844*** (0.0128)	-0.0320*** (0.0041)
<i>N</i>	21,997,135	14,989,490	7,007,645	2,455,093	1,203,801	1,251,292
Adj. R^2	0.10149	0.08174	0.11231	0.05663	0.08236	0.04662
Mean Dep. Var.	0.26001	0.21417	0.35806	0.52265	0.43382	0.60811
Women						
Low Training	-0.0106*** (0.0040)	-0.0063 (0.0047)	-0.0117*** (0.0025)	0.0117* (0.0062)	0.0206** (0.0090)	0.0073** (0.0036)
High Training	0.0145** (0.0057)	0.0261*** (0.0069)	-0.0081* (0.0043)	0.0277*** (0.0102)	0.0458*** (0.0148)	0.0163*** (0.0054)
Mean Log Wage	-0.0524*** (0.0039)	-0.0279*** (0.0046)	-0.1007*** (0.0026)	-0.0112** (0.0057)	0.0290*** (0.0082)	-0.0428*** (0.0028)
<i>N</i>	16,738,809	11,674,799	5,064,010	2,069,382	961,215	1,108,167
Adj. R^2	0.08904	0.05797	0.09773	0.06805	0.07404	0.05880
Mean Dep. Var.	0.25075	0.20609	0.35371	0.47285	0.38941	0.54521

Notes: This table reports regressions of one-year log wage growth, $\log(w_{i,t+5}) - \log(w_{it})$, on indicators for working at a low- or high-training firm in year t , estimated separately by gender. The omitted category is no-training firms. See the notes under Table 3 for sample and controls. Standard errors are clustered at the worker level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.4. Ten-Year Log Wage Growth and Training Tier

	All Cohorts			Younger Cohorts		
	(1) All	(2) Stayer	(3) Mover	(4) All	(5) Stayer	(6) Mover
Men						
Low Training	-0.0146** (0.0074)	-0.0065 (0.0095)	-0.0103* (0.0054)	-0.0027 (0.0208)	0.0039 (0.0251)	-0.0040 (0.0224)
High Training	0.0163* (0.0086)	0.0329*** (0.0102)	-0.0059 (0.0083)	0.0260 (0.0277)	0.0413 (0.0460)	0.0208 (0.0243)
Mean Log Wage	-0.0520*** (0.0061)	-0.0015 (0.0088)	-0.1054*** (0.0040)	0.0518*** (0.0148)	0.1546*** (0.0304)	0.0118 (0.0077)
<i>N</i>	11,505,282	6,351,703	5,153,579	445,485	125,724	319,761
Adj. <i>R</i> ²	0.13474	0.11483	0.14478	0.04318	0.11544	0.02915
Mean Dep. Var.	0.41469	0.35239	0.49148	0.86656	0.74234	0.91541
Women						
Low Training	-0.0222*** (0.0063)	-0.0188** (0.0086)	-0.0135*** (0.0032)	0.0152 (0.0094)	0.0239 (0.0150)	0.0194** (0.0081)
High Training	0.0247*** (0.0081)	0.0405*** (0.0115)	0.0066 (0.0056)	0.0536*** (0.0163)	0.0806*** (0.0278)	0.0516*** (0.0120)
Mean Log Wage	-0.0669*** (0.0056)	-0.0300*** (0.0078)	-0.1111*** (0.0031)	0.0001 (0.0085)	0.0691*** (0.0156)	-0.0265*** (0.0050)
<i>N</i>	8,670,611	4,930,379	3,740,232	440,766	115,216	325,550
Adj. <i>R</i> ²	0.10406	0.07157	0.11575	0.05001	0.08097	0.04706
Mean Dep. Var.	0.39779	0.34280	0.47028	0.77361	0.68152	0.80620

Notes: This table reports regressions of one-year log wage growth, $\log(w_{i,t+10}) - \log(w_{it})$, on indicators for working at a low- or high-training firm in year t , estimated separately by gender. The omitted category is no-training firms. See the notes under Table 3 for sample and controls. Standard errors are clustered at the worker level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.5. Horserace: Training vs. R&D and Employee Benefits

	(1) Base	(2) Benefits	(3) R&D	(4) Both
High training	0.0174*** (0.0031)	0.0174*** (0.0031)	0.0176*** (0.0031)	0.0175*** (0.0031)
log(empl. benefits p.w. + 1)		5.85×10^{-5} (0.0004)		1.44×10^{-5} (0.0004)
log(R&D fee p.w. + 1)			0.0014*** (0.0003)	0.0014*** (0.0003)
Mean log wage	0.0001 (0.0032)	6.34×10^{-5} (0.0034)	-0.0026 (0.0034)	-0.0026 (0.0035)
log(firm size)	-0.0015 (0.0015)	-0.0015 (0.0015)	-0.0026 (0.0016)	-0.0026 (0.0017)
Return on capital	-6.27×10^{-6} (0.0001)	-6.91×10^{-6} (0.0001)	4.32×10^{-5} (0.0001)	4.31×10^{-5} (0.0001)
<i>N</i>	14,008,586	14,008,537	14,008,586	14,008,537
Adj. R^2	0.03151	0.03151	0.03182	0.03182
Mean of dep. var.	0.11232	0.11232	0.11232	0.11232

Notes: This table reports regressions of three-year log wage growth, $\log(w_{i(t+3)}) - \log(w_{it})$, for female workers aged 22 to 50 at training firms. The reference group is low-training firms. Column (1) is the baseline; columns (2)–(4) add log employee benefits per worker, log R&D expenses per worker, or both. Movers and stayers are pooled. All columns control for mean log wage, log firm size, and return on capital, as shown, along with the controls in Table 3. Standard errors clustered by worker in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Figures

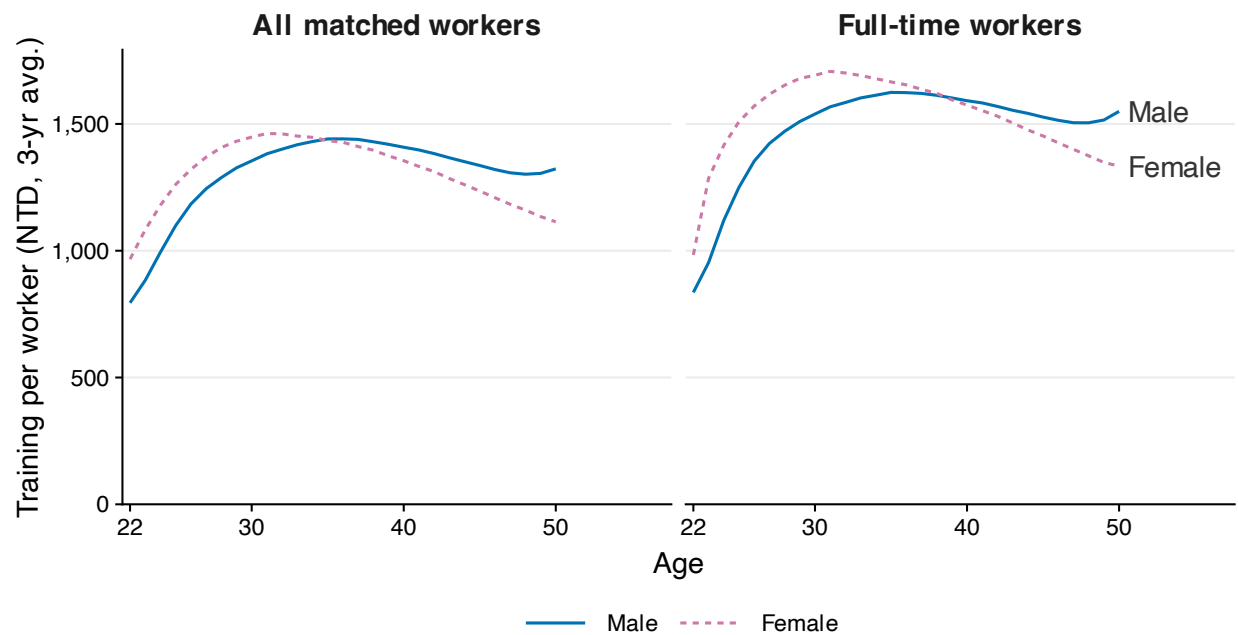


FIGURE A.1. Mean training by gender and age

Notes: Training intensity is measured by the mean of the employer's 3-year average training per worker (NTD). The sample is restricted to workers aged between 22 and 50, and the right panel is further restricted to full-time workers who earn at least 12 months of minimum wage.

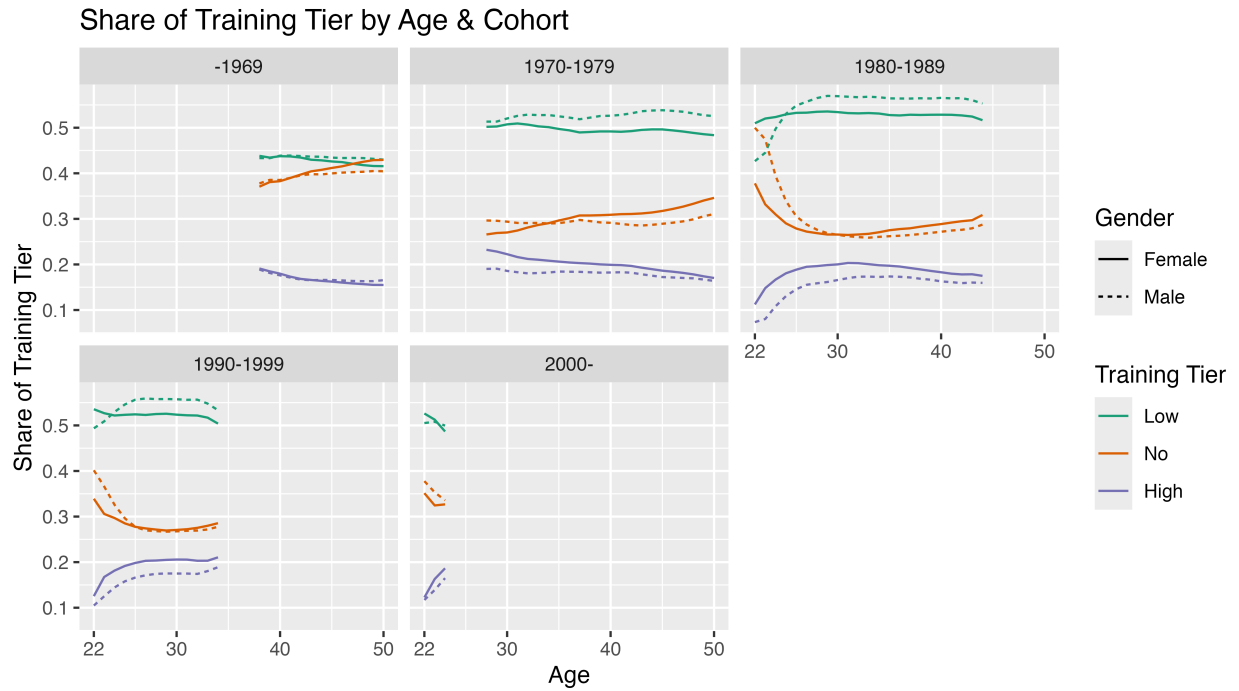


FIGURE A.2. Employment by No, Low, High Training Firms by Age, Gender and Cohort

Notes: This figure replicates Figure 2 by showing the the share of women and men employed by no, low, and high-training firms by age, separately by birth cohorts.

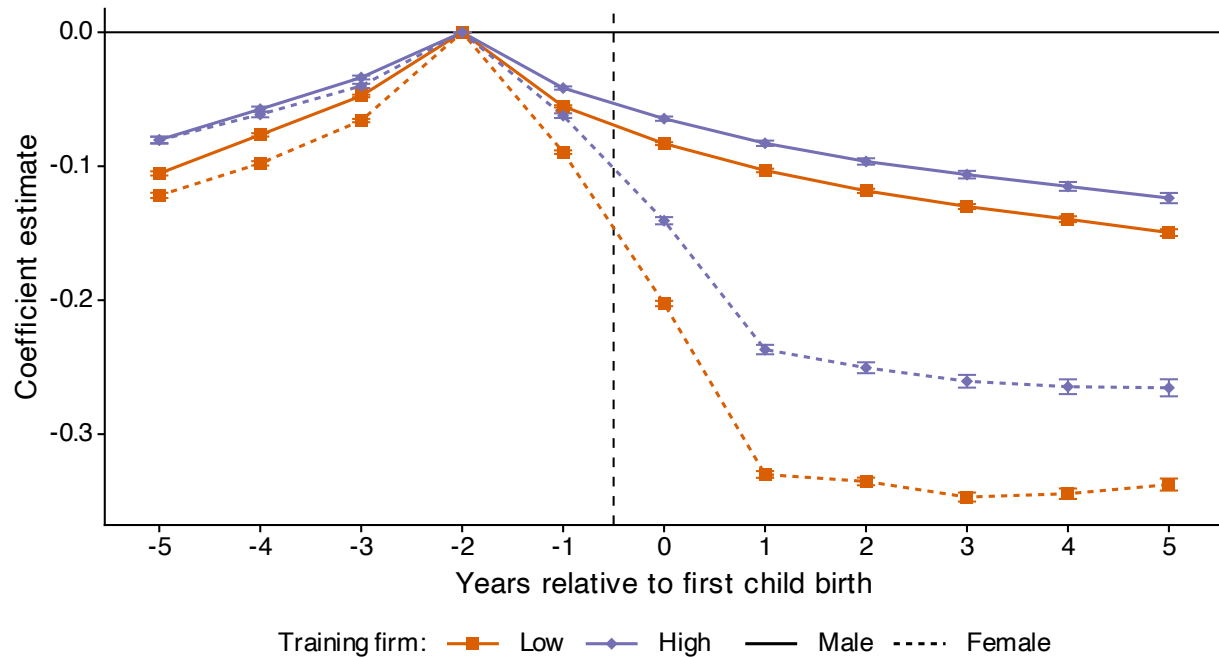


FIGURE A.3. Attachment to the labor market: full-time or part-time

Notes: Working (full- or part-time), high- and low-training firms at $l = -2$. Coefficients (ref. $l = -2$), 95% CI.

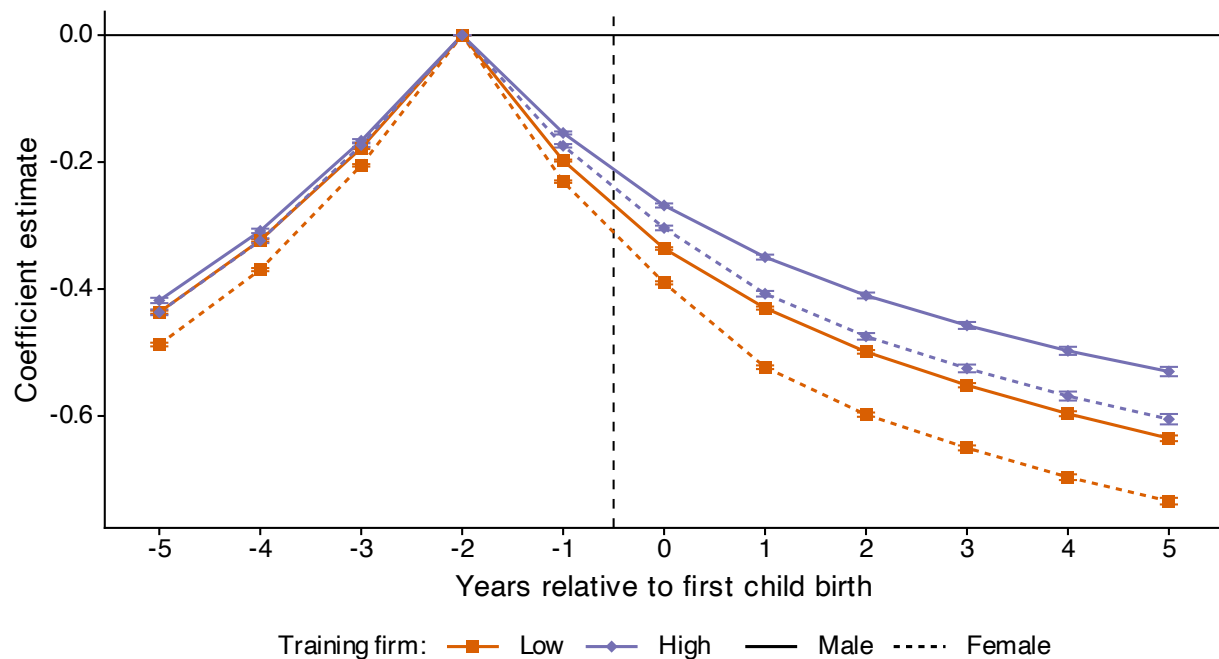


FIGURE A.4. Retention: Still with the $l = -2$ Employer

Notes: Still with the $l = -2$ employer (0 if not working), by its training tier. Coefficients (ref. $l = -2$), 95% CI.

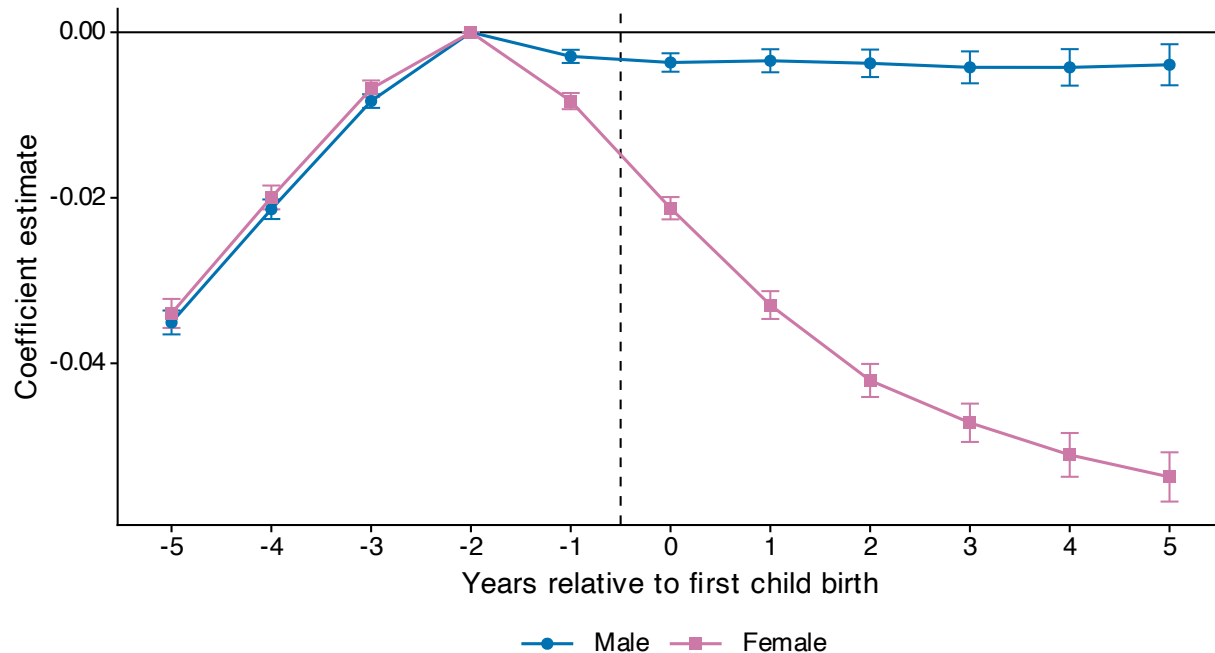


FIGURE A.5. Employment at a High-training Firm

Notes: Employer in year l is high-training (0 if not working); positive wage in some l from 1 to 5. 95% CI.

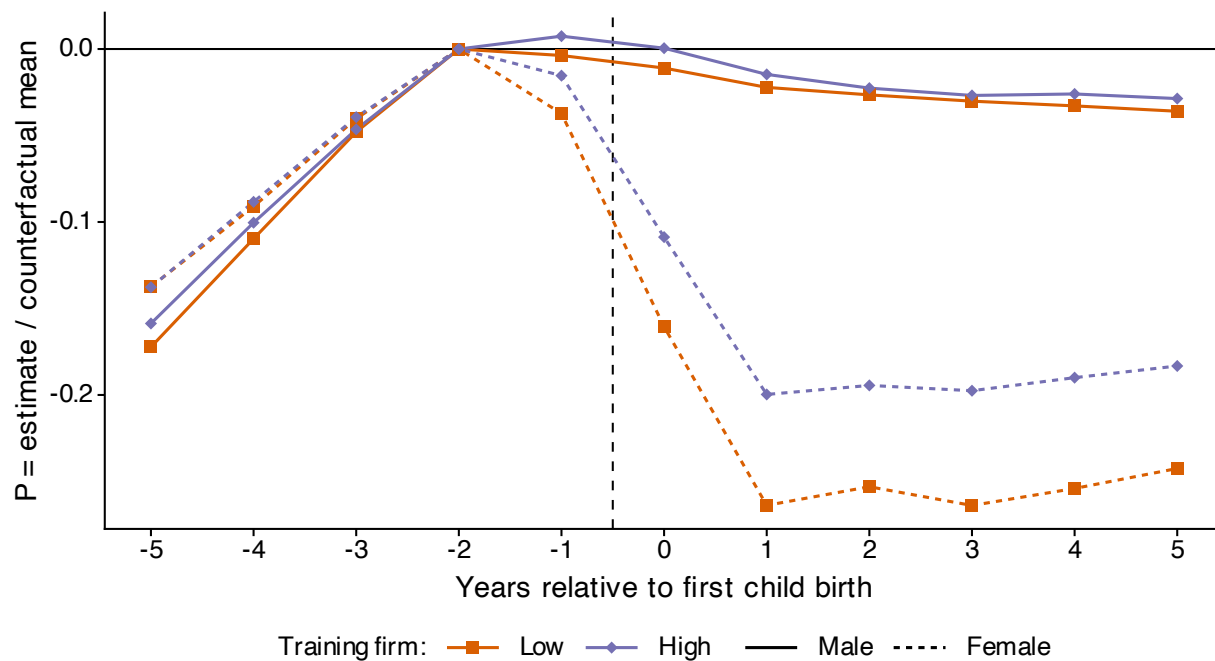


FIGURE A.6. Earnings relative to counterfactual around first childbirth, by training tier

Notes: Annual earnings (NTD) divided by counterfactual mean (no CI), positive-wage years, high- and low-training firms at $l = -2$.

- Following Kleven et al. (2019), $P_l = \hat{\gamma}_l / \mathbb{E}[\tilde{Y}_{it} \mid l]$: the earnings effect at l as a share of predicted earnings absent the birth.

A.2 Data Appendix

Sample construction. Table A.6 reports the number of observations at each step. The salary records begin in 2004 and the business income tax returns in 2005. Because a training tier requires three consecutive years of returns, the first year with a tier is 2007, which is where our analysis samples begin. Workers whose main employer cannot be matched to a firm with a training tier, about 29 percent of person-years, are mostly employed in the public sector or by organizations that do not file business income tax returns.

TABLE A.6. Sample construction

	Observations
<i>Workers</i>	
Salary payment records, 2004–2023	288,434,129
Person-years, main employer only	174,062,877
Main employer matched to a firm with a training tier	123,327,528
Main employer not matched	50,735,349
<i>Matched person-years, by attachment</i>	
Full-time (≥ 12 months of minimum wage)	81,517,095
Part-time (6 to 12 months of minimum wage)	17,504,787
Below 6 months of minimum wage	24,305,646
<i>Analysis samples</i>	
Main sample: full-time, aged 22 to 50	63,773,258
Wage growth, 1-year horizon	56,975,940
Wage growth, 3-year horizon	47,166,330
Wage growth, 5-year horizon	38,735,944
Wage growth, 10-year horizon	20,175,893
<i>Firms</i>	
Firm-years with a business income tax return, 2005–2023	14,582,324
Firm-years with a training tier, 2007–2023	7,593,498

Notes: Wage-growth samples include full-time and part-time workers (≥ 6 months of minimum wage) aged 22 to 50, and drop observations with missing return on capital or leave-one-out mean log wage (firms with one worker). Firms with a training tier are observed in three consecutive years with at least one worker.

Variable details.

- *Earnings.* We keep payments classified as salary or professional-practice income (fees paid to individuals for professional services) to resident individuals.
- *Main employer.* When two firms pay a worker the same amount in a year, we keep one of them.
- *Firm size.* The number of workers for whom the firm is the main employer, counted before the age and full-time restrictions.
- *Winsorization.* Annual earnings are winsorized at the 99.5th percentile within each year. Training, employee benefits, and R&D per worker are winsorized at the 99.5th percentile, and return on capital at the 0.5th and 99.5th percentiles, pooling all years. Six firm-years with negative training expense are set to missing.
- *Industry.* Industry comes from the six-digit business code on the tax return, mapped to the 19 sections of Taiwan's Standard Industrial Classification. The tax system's business codes changed around 2008, so for 2005 to 2007 we assign each firm the section it reports in later years.
- *Event-study sample.* First births between 2010 and 2018, with the parent alive at least five years after the birth.

A.3 Definition of Training Expenses

The training expense line on the business income tax return follows Article 86-1 of the Regulations Governing Assessment of Profit-Seeking Enterprise Income Tax. The full text of the article follows, in our translation from the official Chinese text ([Ministry of Finance 2023](#)).

Article 86-1 (Training expenses)

1. Expenses a profit-seeking enterprise pays to develop its employees, by organizing training activities related to its business or by sending employees to such activities, are recognized at their verified amounts according to the following rules:
 - (a) Instructors' hourly fees and travel expenses.
 - (b) Trainees' travel expenses and fees paid to the training provider.
 - (c) Teaching materials, practice materials, stationery, medical expenses, insurance, observation visits, books and magazines, meals during training, venue fees, and training equipment with a useful life of less than two years.
 - (d) Fees for skill certification tests.

- (e) Other expenses that the central competent authority for the industry and the Ministry of Finance recognize as training expenses on a case-by-case basis.
 - (f) Expenses for sending selected employees abroad for study or research. The enterprise must have written rules for employees' overseas study, and the employees must return to work for the company when the program ends. Any part paid by foreign institutions may not be deducted.
 - (g) Fees or subsidies paid to a partner school under industry–academia cooperation, which require a cooperation plan or contract.
2. Vocational training fees, or the shortfall, that a profit-seeking enterprise pays under the vocational training laws and regulations may be deducted as an expense in the current year.
 3. Original vouchers for training expenses are handled according to the relevant articles of these Regulations.