The Dual Labour Market and the Motherhood Employment Penalty in Japan: Evidence from Household Panel Data

Beatriz Gietner*

March 26, 2026

Abstract

Japan has among the most generous parental leave statutes in the OECD, yet one of the widest motherhood employment gaps. Using an event-study design applied to 662 first births in the JHPS and KHPS panels (2004–2022), I show that the penalty is strongly stratified by labour-market tier. Employment falls 13.2 percentage points at childbirth, with a clean near-lead null at $t = -2$. Among women employed before birth, non-regular contract status predicts non-employment at $t = 0$ with an odds ratio of 7.50 ($p < 0.001$), while small-firm effects are positive but imprecise. Post-birth "recovery" is largely an extensive-margin illusion - returning mothers shift from full-time to part-time work, and conditional hours, earnings, and implied hourly wages remain durably below pre-birth levels. Government training subsidies designed to support re-entry face a three-layer access barrier: roughly half of mothers have never heard of the programme, and among those who are aware, most report being ineligible or unsure whether they qualify, consistent with the loss of employment-insurance-linked eligibility after childbirth exit. These patterns are consistent with weaker effective access to job-continuity protections in non-regular and small-firm segments; statutory entitlements alone appear insufficient where labour-market dualism channels mothers into employment tiers with limited institutional support.

Keywords: motherhood penalty, child penalty, dual labour market, Japan, event study, non-regular employment

JEL Codes: J13, J16, J21, J31, J41

*University College Dublin. I am grateful to Professor Midori Wakabayashi (Tohoku University) for her guidance and support, and for facilitating access to the JHPS/KHPS panel data through the appropriate institutional process; to Naoko Caheny for generous interviews and discussions on women's labour-market conditions in Japan; and to Dr. Yuki Takahashi and participants at the Young Early-Career Japanese Researchers Seminar for valuable comments. All remaining errors are my own.

1 Introduction

Japan presents a puzzle. Over the past three decades, successive governments have enacted increasingly generous work-family policies: parental leave now covers the vast majority of employees, childcare capacity has expanded, and equal employment opportunity legislation has been in place since 1986 (OECD, 2024a; Ministry of Health, Labour and Welfare, 2019). Yet Japanese women's labour force participation continues to exhibit the characteristic M-curve - a sharp decline during prime childbearing years followed by partial recovery - that has defined the country's gender employment gap for half a century (Ochiai, 1997; Yu, 2009). If the policy infrastructure exists, why do women still leave?

A large literature studies the labour-market consequences of childbirth for women. Recent work by Kleven et al. (2019a,b) popularized the child-penalty framework and documented its importance across OECD countries. However, this framework identifies when employment changes occur, but not why (see also Angelov et al., 2016; Adda et al., 2017; Waldfogel, 1998). In institutionally homogeneous labour markets like Denmark's, the distinction may matter less. In Japan, where over 50 percent of employed women hold non-regular contracts with fundamentally different protections than regular workers (OECD, 2024a), the mechanism behind the penalty cannot be inferred from aggregate event-study plots. Despite extensive descriptive work on Japan's dual labour market and gender inequality (Brinton, 1993; Yamaguchi, 2019; Dumauli, 2019; Houseman and Osawa, 2003), to the best of my knowledge, existing studies have not quantified how the regular/non-regular divide and firm size independently predict maternal employment exit using multivariate models with small-sample corrections.

Isolating the role of labour market structure from individual preferences is difficult. Women who hold non-regular jobs may differ from regular workers in unobserved dimensions such as career commitment, family planning intentions, and risk tolerance, that independently predict exit. The low incidence of observed exits in household panel data limits the precision of the mechanism estimates. I address this in three ways: first, by comparing quitters and stayers on observable pre-birth characteristics; second, by using Firth-corrected logit models and bootstrap inference in the low-event setting; and third, by adopting a descriptive design that emphasizes the magnitude and composition of employment change rather than claiming strict causal identification.

Using an event study design applied to 662 first births observed in the JHPS and KHPS from 2004 to 2022, I document that Japan's motherhood employment penalty is strongly stratified by labour market tier: exits are concentrated among women in jobs with weaker continuity through motherhood - non-regular contracts, small employers, and weaker effective leave access. In this work, I reframe the motherhood penalty as an institutionally patterned trajectory rather than a uniform response to childbearing. In this sample, the timing is consistent with a broader family-formation pattern: exits rise before birth around the marriage-to-birth transition, and a sharper employment break appears at childbirth. The estimates are descriptive event-time trajectories in the analytic panel, not within-person causal treatment effects; they combine within-person change with composition change from differential observation across event time and are interpreted with explicit support and composition constraints.

I focus first on the extensive margin. The employment rate at $t = -1$ (the year before first birth) is 50.3 percent. Relative to this baseline, the event-study estimates show no evidence of a near pre-birth decline at $t = -2$ (coefficient -0.012, $p = 0.70$), followed by a 13.2-percentage-point drop at $t = 0$. In back-of-the-envelope terms, that corresponds to about 87 of the 662 mothers in the sample, or roughly one in four women who were working at baseline, moving from employment to non-employment in the birth year. Employment moves back toward its pre-birth level by $t = +3$, but that recovery is misleading: among women who were full-time at $t = -1$, fewer than half (45.7 percent) remain full-time at $t = +1$, and by $t = +5$ only 42.2 percent do. In other words, observed recovery in employment does not imply recovery in job quality.

I turn next to predictive risk stratification at the childbirth margin. Among women who were employed at $t = -1$ and are observed again at $t = 0$ (N = 330; 124 exits), non-regular status at $t = -1$ is a very strong predictor of non-employment at childbirth (OR = 7.50, $p < 0.001$), while small-firm effects are positive but imprecisely estimated. The model-implied absolute risk of exit ranges from 14.6 percent for regular workers in large firms or government employment to 66.5 percent for non-regular workers in small firms; the corresponding raw cell rates are 13.8 and 64.8 percent. The pre-birth exit models at $t = -2$ are retained only as lower-power supporting evidence and are interpreted as predictive stratification, not causal effects.

Lastly, I turn to the intensive margin among women who remain employed. The estimates are conditional-on-employment event-study coefficients, and the reference period is $t = -1$. The mean weekly hours among employed mothers at $t = -1$ are 34.1. The coefficient at $t = +5$ is -10.1. This means that employed mothers work about 10 fewer hours per week than in the pre-birth reference year. The annualized equivalent is about 525 fewer hours of paid work per year. The mean annual earnings among employed mothers at $t = -1$ are 268.9 man-yen. The coefficient at $t = +5$ is -96.2 man-yen. This implies a loss of about 962,000 yen per employed mother per year relative to the pre-birth reference period, and I do not find a clear sign of recovery. I interpret these estimates as conditional intensive-margin losses rather than full-population effects. The pattern is directionally consistent with recent administrative evidence from local tax records (Fukai and Kondo, 2025). The regular/non-regular gradient also closely parallels the heterogeneity documented by Kikuchi (2026) using the larger JPSED panel. The father-side adjustment remains limited throughout the window. Mothers provide roughly five times more childcare than fathers through $t = +5$, and in the partner-linked analytic sample only 1 of 661 partners reports taking childcare leave (0.15%).

The chapter makes four contributions, with each one calibrated to the strength of the underlying evidence. The strongest result is descriptive: employment drops sharply at childbirth, while the near pre-birth lead is essentially flat. This birth-year break is stable across cohort splits, survey sources, IPW reweighting, balanced-panel restrictions, and a matched childless placebo. A second result concerns the composition of post-birth recovery. Aggregate employment moves back toward its pre-birth level, but the recovery masks a persistent shift toward lower-quality work: full-time attachment, hours, and earnings do not return to baseline within five years. A third result is predictive rather than causal. Labour-market tier, and especially non-regular contract status, strongly predicts exit at the childbirth margin, with an odds ratio of 7.50 in the main risk set. The final contribution is interpretive. Taken together, these patterns are consistent with unequal effective access to job continuity across contract types and firm sizes, although that institutional interpretation is not separately identified by the panel design. This logic structures the chapter as a whole: the abstract, results, and conclusion use stronger language for stronger evidence and explicitly mark more tentative claims as interpretive. Table 1 summarises this evidence hierarchy near the outset, and Appendix D.14 provides a consolidated robustness table for the core employment coefficients ($t = -2$, $t = 0$, $t = +5$) together with a denominator map linking each headline statistic to its underlying risk set.

Table 1: Evidence hierarchy used throughout the chapter

Level	Core evidence	Interpretation rule
High	Employment break at $t = 0$ with near lead null; stable across major sensitivity checks	Treated as the chapter's strongest descriptive fact
Moderate	Childbirth-margin risk stratification by labour-market tier; post-birth compositional downgrade	Treated as robust association evidence, not causal mechanism identification
Interpretive	Institutional channels (care logistics, fiscal thresholds, workplace practices)	Used to discipline interpretation, not as separately identified effects

Notes: Detailed evidence-to-claim mapping and full diagnostics are reported in Section 8 and Appendix D.

The paper's comparative advantage is not the best causal estimate of Japan's child penalty - that will require linked administrative data. Rather, it provides an institutionalised reading of the child penalty in a segmented labour market, with transparent support and composition diagnostics, and a well-documented extensive-to-intensive margin transformation that standard aggregate event studies obscure (Kleven et al., 2019b,a; Angelov et al., 2016; Bertrand et al., 2010; OECD, 2024a; Yamaguchi, 2019).

The remainder of the paper proceeds as follows. Section 2 describes the institutional context of Japan's dual labour market and gendered work culture. Section 3 presents the data and empirical strategy. Section 4 documents the aggregate birth-year break and the composition of pre-birth exits. Section 5 presents predictive risk stratification by labour-market tier and reports heterogeneity by job type and firm size. Section 6 examines the intensive margin and the transformation of the penalty over time. Section 7 presents father-side descriptive evidence. Section 8 reports robustness checks. Section 9 discusses policy implications, external validity, limitations, and concludes.

2 Institutional context: Japan's dual labour market and the gendered household

Understanding why Japanese women exit employment at motherhood requires understanding two features of the institutional environment that standard child penalty models do not capture: a labour market divided into structurally distinct tiers, and a household division of labour that assigns virtually all caregiving to mothers. This section describes both.

2.1 The regular/non-regular divide

Japan's labour market is organised around a distinction that has no direct equivalent in most Western economies. Regular employees (*seishain*) hold open-ended contracts with a single employer, receive seniority-based wages, firm-specific training, and strong de facto protections against dismissal. Non-regular employees (*hi-seiki*) hold fixed-term, part-time (*pāto*), or dispatch contracts with limited tenure, flat wage profiles, and few of the implicit guarantees that define regular employment (Brinton, 1993, 2001; Yamaguchi, 2019; Houseman and Osawa, 2003; Kambayashi and Kato, 2017). The contract category "part-time" (*pāto*) is a legal employment classification and does not necessarily imply fewer working hours; in this sample, 36 percent of non-regular workers work full-time hours (35+ per week) and 12 percent of regular workers work part-time hours. As of 2020, approximately 37 percent of the total workforce and over 50 percent of employed women held non-regular positions (OECD, 2024a).

This distinction is important for the motherhood penalty because it determines access to the institutional accommodations that allow continued employment through and after childbirth. Regular employees at large firms benefit from internal labour markets: their positions are held during leave, their return is expected, and seniority continues. Japan's Child Care and Family Care Leave Act formally extends parental leave eligibility to all employees who have worked for their employer for at least one year, with generous replacement rates (Ministry of Health, Labour and Welfare, 2019). On paper, non-regular workers are covered. In practice, this formal coverage is often fragile. Fixed-term contracts can expire before or around leave periods, renewal decisions are uncertain, and small firms face tighter staffing margins and higher administrative costs when workers take extended leave (OECD, 2024a; Yamaguchi, 2019). As a result, legal entitlement (de jure access) and usable entitlement (de facto access) diverge for many women, but especially for those in non-regular and small-firm employment.

Firm size reinforces this divide. Large employers can usually redistribute tasks, maintain staffing buffers, and preserve positions during leave spells. Small and medium-sized firms, by contrast, often operate with thin margins and limited replacement capacity, so even short absences are harder to accommodate. This creates a practical gradient in continuity: the larger and more formalized the employer, the more feasible leave-taking and post-leave return become. The patterns in my data are consistent with this firm-size gradient.

A third institutional layer reinforces segmentation through fiscal incentives. Japan's tax and social-insurance system creates sharp earnings thresholds for secondary earners: the well-known 1.03 million yen tax threshold, and additional thresholds around 1.06 and 1.30 million yen linked to mandatory social-insurance enrollment (Nagase, 2012; OECD, 2024a). In this sample, earnings bunching at these thresholds is not unique to mothers: among childless married employed women, 49.7% earn at or below 103 man-yen and 60.0% at or below 130 man-yen, compared with 47.8% and 55.9% respectively for returning mothers at $t = +3$ to $t = +5$ (see Section 6 and Figure 11). This comparison suggests that the thresholds are a broader structural feature constraining married women's earnings. What motherhood adds is a mechanism that pushes women into the part-time earnings range where these thresholds bind - from full-time positions with earnings above the cutoffs to reduced-hours jobs - often on non-regular contracts - where threshold-constrained hours become the norm.

2.2 The corporate warrior and the gendered household

Japanese men work among the longest hours in the OECD. The figure of the *salaryman* or corporate warrior remains a defining feature of Japanese corporate life despite recent policy efforts to reduce overtime (Nemoto, 2016; North, 2009; OECD, 2024a). The 2018 Work Style Reform Act introduced overtime caps, but enforcement remains uneven - men substitute paid overtime for unpaid overtime, leaving total hours largely unchanged (Burdin et al., 2024) - and the culture of long hours has deep organisational roots (Kuroda and Yamamoto, 2013).

For households, this has a direct implication. When husbands work 50 or more hours per week, they are less available for routine childcare during weekdays, and the domestic care burden shifts toward mothers. The time-use data in this sample show a wife-to-husband childcare ratio of approximately 5-to-1 to 6-to-1, in line with national surveys (OECD, 2024a; Nemoto, 2016). This pattern is consistent with a binding time-allocation constraint in which long-hours norms and limited job flexibility restrict the household's scope for reallocation after childbirth.

2.3 Marriage, motherhood, and regional norms

A third institutional dimension is the continuing role of marriage as a transition point in women's employment. In rural Japan and smaller cities, traditional household norms can create pressure to exit employment at marriage, independently of childbirth (Ochiai, 1997; Yu, 2009; Raymo and Iwasawa, 2005). In metropolitan areas, this norm is weaker, while the motherhood penalty appears more closely associated with labour-market segmentation and corporate work culture. This geographic heterogeneity suggests that aggregation can mask different local compositions of marriage-related and childbirth-related exits.

2.4 Policy perspectives

Japan's work-family policy regime has expanded steadily since the 1990s, with the explicit aim of reducing employment discontinuity around childbirth. The Child Care Leave Act introduced paid leave with income replacement for eligible workers, with extensions up to two years when childcare placement is unavailable (Ministry of Health, Labour and Welfare, 2019). In parallel, the state expanded childcare capacity and even encouraged firms to increase women's workforce participation (OECD, 2024a; Nishitateno and Shikata, 2017; Yamaguchi, 2019). Evidence from policy reforms suggests these measures did improve maternal employment continuity among workers who are eligible and able to use them (Asai, 2015). The central question for this paper is why these expansions have not produced comparable continuity across labour-market tiers.

Policies therefore do not operate uniformly across labour-market tiers. In practice, they are most effective for women already in regular employment. Income-replacement leave requires employment-insurance eligibility, which depends on stable contracts and minimum hours. Childcare expansion can support return-to-work only when a job match still exists to return to. For women whose fixed-term contracts lapse around childbirth, the binding issue is employment continuity rather than childcare access alone. The policy architecture thus aligns more closely with *seishain* careers than with non-regular trajectories, so aggregate effects are attenuated as non-regular employment expands (Genda, 2005; Piotrowski et al., 2015).

A practical issue is timing and affordability of care by child age. In Japan, care for ages 0-2 is largely provided through nursery-type settings (*hoikuen*/certified centres), while age 3 onward expands into a broader early-education menu (*yochien* plus integrated centres), with substantial policy support from age 3-5 (Children and Families Agency, Government of Japan, 2024b,a). This age-structured transition is important for the dynamics of maternal employment. The first years are typically the tightest care-constraint period, and the feasibility of sustained employment often improves only once children move into the 3+ system. Cost also remains relevant despite subsidies, because out-of-pocket burdens vary by household income, municipality, care type, and ancillary fees (OECD, 2025b, 2024a). In short, the institutional environment is a sequence of regimes over the child life-cycle.

3 Data and empirical strategy

3.1 Data sources

I draw on two complementary household panel surveys conducted by the Panel Data Research Center at Keio University.[1] The Keio Household Panel Survey (KHPS) has been fielded annually since 2004, sampling households nationwide with a focus on economic behaviour, employment, and family structure. The Japan Household Panel Survey (JHPS) began in 2009 with a broader scope and a larger initial sample. Both surveys collect detailed individual-level information on employment status, work hours, earnings, job characteristics (including contract type and firm size), commute time, household composition, reasons for job changes, and many other sections depending on the year considered. The surveys use a common questionnaire structure that facilitates the pooling of observations across both panels.

The access to JHPS/KHPS microdata is provided under approved institutional conditions via the Panel Data Research Center at Keio University.[2] The replication materials therefore include only code and derived outputs, but not raw respondent-level microdata, in accordance with the data-use agreement and confidentiality requirements.

The official JHPS/KHPS construction program harmonizes variable numbers across surveys through dedicated integration scripts and a published correspondence file. I follow that harmonization framework and implement additional variable-level checks for all core outcomes and mechanisms used in this paper (employment, hours, earnings, contract type, and firm size). I also report robustness with a dataset indicator (KHPS vs JHPS), which leaves the key near-lead and at-birth estimates substantively unchanged.

I estimate unweighted models. The released integration program and variable lists do not provide a single harmonized longitudinal survey weight for the pooled JHPS/KHPS panel used here. Accordingly, the coefficients are interpreted as sample-average trajecto-

[1] KHPS was launched in 2004 and JHPS in 2009. The surveys were initially managed separately, and their questionnaires were unified from the 2014 survey onward in the integrated JHPS/KHPS framework. Both are nationwide panel surveys built from stratified sampling designs, and the official release includes harmonisation documentation (including variable correspondence files and integration scripts) as well as sample-weight materials.

[2] Panel Data Research Center, Keio University: `https://www.pdrc.keio.ac.jp/`. Access to JHPS/KHPS microdata is governed by the Center's application and confidentiality procedures.

ries within the analytic panel of married first-birth women observed around childbirth, rather than nationally weighted population moments.

I use all available waves: KHPS 2004-2022 and JHPS 2009-2022. The combined dataset thus provides sufficient temporal coverage to observe women from several years before first birth through several years after, which is essential for the event study design.

Institutional context. Section 2 describes the institutional environment in detail. Three features are particularly relevant for the empirical design: (1) the regular/non-regular employment divide, which determines access to leave and job continuity; (2) firm-size variation in human resource capacity, which determines whether leave can be practically exercised; and (3) spousal tax and social-insurance thresholds, which shape women's labour supply incentives on re-entry. The variable definitions below operationalize these institutional features.

3.2 Sample construction

My unit of analysis is the married woman observed around the time of her first birth. I identify first births from the fertility modules of both surveys. Both KHPS and JHPS record household events (including births) that occurred between February of the previous year and January of the survey year, so the "birth year" in the data corresponds to the survey wave in which the birth was reported. I define event time k as the difference between the survey year and the birth year, so $t = 0$ is the wave in which a birth is first recorded, $t = -1$ is the prior wave (which may overlap with late pregnancy), and $t = +1$ is the first full post-birth wave.

To construct the event-study sample, I restrict the event window to $k \in [-5, +5]$. A woman enters the sample at whatever event time her first observed survey response corresponds to, and she remains until she drops out of the survey or the window ends. This produces an unbalanced panel in which sample sizes increase toward $t = -1$ (N = 656) and then decline at longer post-birth horizons due to panel attrition and the end of available waves. The sample sizes at the extremes - N = 186 at $t = -5$ and N = 454 at $t = +5$ - therefore reflect an unbalanced support and a differential observability across event time; I quantify these support/attrition patterns explicitly in Sections 4 and 8.

Pooling the two surveys yields 1,184 identified first births (622 from KHPS, 562 from JHPS). After restricting to women with at least one observation in the $[-5, +5]$ window and applying the flexible-window subsample criterion (which requires observations in both the pre- and post-birth periods), the core regression sample includes 662 women contributing 5,212 person-year observations. This is the sample I use for all event-study regressions, summary statistics, and mechanism analyses, unless otherwise noted.

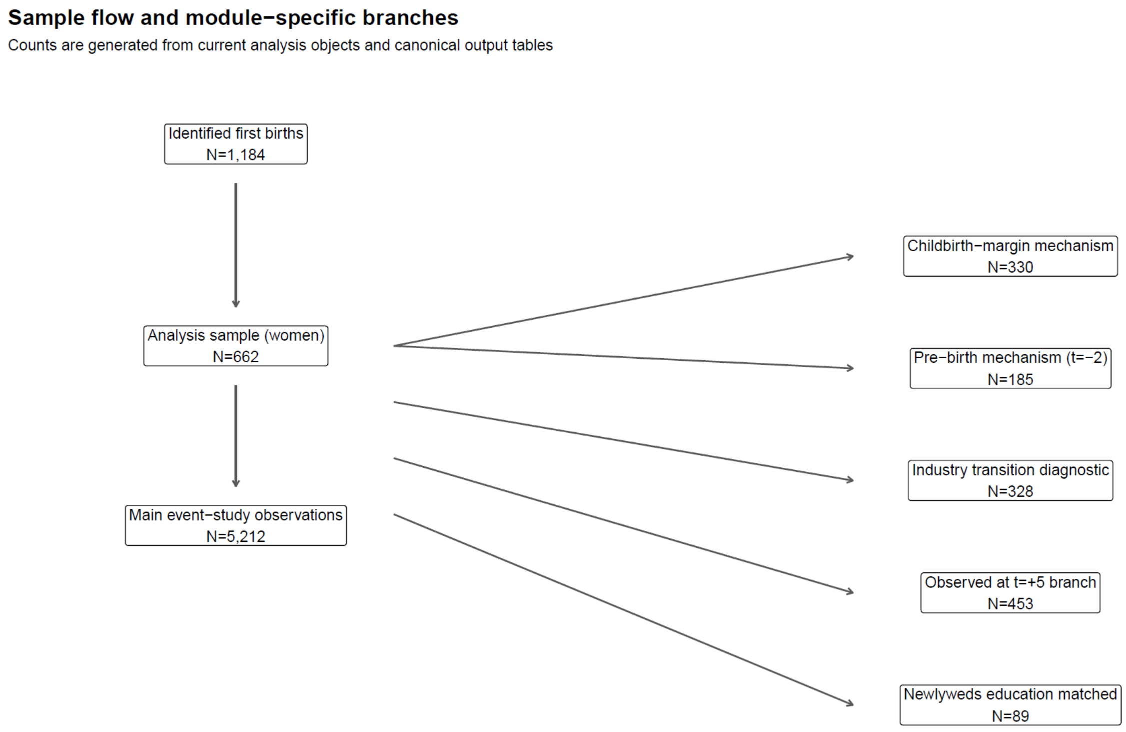

Figure 1: Sample flow from identified first births to module-specific analysis subsamples. The core event-study sample contains 662 women (5,212 person-year observations). Branches report the subsamples used for the childbirth-margin mechanism model, pre-birth mechanism model, industry transition diagnostic, long-horizon attrition/IPW checks, and newlyweds-education appendix diagnostic.

Table 2: Sample flow across analysis modules.

Step	Module	N
1	Identified first births	1,184
2	Analysis sample (event-study women)	662
3	Main event-study observations	5,212
3a	Childbirth-margin mechanism sample (employed at $t = -1$, observed at $t = 0$)	330
3b	Pre-birth mechanism complete-case sample ($t = -2$)	185
3c	Industry transition diagnostic sample	328
3d	Observed at $t = +5$ (attrition/IPW branch)	453
3e	Matched to newlyweds education	89

Notes: Rows 3a-3e are module-specific branches and are not sequentially nested within one another. They are reported to show which sample size applies to each analysis block.

3.3 Variable definitions

The choice of JHPS/KHPS reflects the paper's substantive focus. Relative to administrative earnings records, these household panels observe the job characteristics that matter for the argument here, including contract type, firm size, work hours, quit reasons, commute time, childcare time, and training-related variables. The tradeoff is that labour-force status is measured less cleanly than in administrative or dedicated labour-force surveys. In particular, the questionnaires do not provide a single harmonized unemployment indicator that can be used consistently across waves and surveys. I therefore define employment continuity using reported weekly work hours and interpret the main outcome as employment versus non-employment, rather than as a finer distinction between employment, unemployment, and non-participation.

Employment status. The primary outcome is a binary indicator equal to one if the woman reports any positive weekly work hours in the survey wave. It is coded as zero when reported weekly hours are zero, and as missing when the survey response does not permit employment status to be determined consistently. This outcome is best interpreted as employment versus non-employment. It does not distinguish unemployment from non-participation, because the harmonized JHPS/KHPS questionnaires do not measure that distinction consistently across all waves. I also decompose employment into full-time (35+ hours per week) and part-time (positive hours below 35) to track the composition of employment over the event window.

Hours and earnings. Weekly hours are measured as reported usual weekly work hours among employed women. Annual earnings are reported in units of 10,000 yen (man-yen) and refer to the respondent's own labour income. Both of these variables are conditional on employment: the intensive-margin event studies estimate changes among working women, but not the unconditional population. At $t = -1$, mean weekly hours among workers are 34.1 (SD 16.5) and mean annual earnings are 268.9 man-yen (SD 206.9).

Contract type. I classify employed women as regular (*seishain*) or non-regular (*hi-seiki*) based on the employment status question in each survey. Non-regular employment includes part-time (*pāto*), fixed-term, temporary, and dispatch contracts. The classification follows the survey's own coding and is harmonized across KHPS and JHPS using a crosswalk of variable numbers (see Appendix D.3 for variable mapping details). Contract type and working hours are distinct dimensions with substantial overlap: at $t = -1$, 36 percent of non-regular workers work full-time hours (35+ per week) and 12 percent of regular workers work part-time hours. Throughout this paper, "full-time" and "part-time" refer to hours worked (above or below 35 per week), while "regular" and "non-regular" refer to the employment contract type. At $t = -1$, 30.9 percent of women are employed full-time and 19.4 percent part-time (by hours); among employed women, approximately 38 percent hold non-regular contracts.

Firm size. Employer size is measured using the survey's firm-size categories. I define "small firm" as fewer than 30 employees, "medium" as 30-99, and "large" as 100 or more, with government employees classified separately. The firm-size by contract-type crosstab at $t = -2$ shows 198 women with complete job-characteristic data, distributed across the categories shown in Table 5. The regression sample is smaller (N = 185) because the mechanism models use complete-case observations for the binary covariates included in Table 2.

Quitter classification. I classify women as "quitters" using the survey quit-status mapping at $t = -2$ (quitters for any reason versus stayers), and I treat proxy transition-based definitions as robustness checks. Because the survey's quit-reason module records job-exit events rather than a direct "currently unemployed" category at this margin, classification is based on reported exit events (and not purely on contemporaneous labour-force status). Among the 662 women in the regression sample, 29 (4.4 percent) are classified as quitters at $t = -2$ (flagged as having exited employment between $t = -2$ and $t = -1$), 262 (39.6 percent) as stayers (not flagged as quitters at $t = -2$; note that stayer status is defined by the absence of a quit event rather than by continuous employment, so some stayers may be non-employed at $t = -1$), and the remainder are unclassified in the quit-module mapping at this margin. Event-study support at $t = -2$ is larger (N = 453), so quitter/stayer labels are a subset of women observed at that event time. Quitters' mean pre-birth income is 148.4 man-yen versus stayers' 302.7 man-yen; quitters' employment rate is 55.2 percent versus stayers' 61.0 percent. The women who exit earliest are, on average, those in lower-paying positions.

Household structure. Because co-resident grandparents can relax childcare constraints and potentially moderate the employment penalty, I classify households at $t = -1$ into three types based on the household roster: nuclear (no co-residing parents of either spouse, 89.6 percent), patrilocal (co-residing with the husband's parents, 5.8 percent), and matrilocal (co-residing with the wife's parents, 4.3 percent). Patrilocal and matrilocal residence are identified by matching parent identifiers in the household roster to either the husband or the wife, with adjustments for whether the survey respondent is the wife herself (Type A) or her spouse is the respondent and the wife is recorded as the spouse (Type B).

3.4 Summary statistics

Notes: Quitter/stayer labels are defined using quit-status mapping at $t = -2$ (not by contemporaneous $t = -1$ employment status). The remaining women are unclassified in the quit-module mapping at this margin; event-study support at $t = -2$ is larger than the classified subset.

Table 3: Summary statistics for the analytic sample of married first-birth mothers at $t = -1$ (N = 662). Employment shares are reported for all mothers; weekly hours and annual earnings are reported among employed mothers only. Earnings are in man-yen (10,000 yen units).

	Full Sample (N = 662)	Quitters (N = 29)	Stayers (N = 262)
Employed (%)	50.3	55.2	61.0
Full-time (%)	30.9	31.0	36.7
Part-time (%)	19.4	24.1	24.3
Hours/week (mean)	34.1	32.5	33.4
SD	16.5	17.5	16.3
Income, man-yen (mean)	268.9	148.4	302.7
SD	206.9	95.2	242.2

Table 3 reports summary statistics at $t = -1$ for the full sample, quitters, and stayers. The full sample of 662 women has an employment rate of 50.3 percent, with 30.9 percent in full-time and 19.4 percent in part-time work. Among employed women, mean weekly hours are 34.1 (median 40.0), and mean annual earnings are 268.9 man-yen. The gap between quitters and stayers is considerable: quitters earn roughly half what stayers earn (148.4 versus 302.7 man-yen), and they are more likely to hold non-regular positions. These patterns motivate the mechanism analysis in Section 5, which asks whether the job characteristics that differ between quitters and stayers - contract type and firm size - predict exit in a multivariate framework.

3.5 Empirical Strategy

3.5.1 Event study specification

I estimate an event study of employment, hours, and earnings around first birth:

$$Y_{it} = \sum_{k \neq -1} \beta_k \cdot \mathbf{1}[\text{event\_time}_{it} = k] + \gamma_t + \varepsilon_{it}, \tag{1}$$

where Y_{it} is the outcome for woman i in year t, the indicators $\mathbf{1}[\text{event\_time}_{it} = k]$ capture each event year $k \in \{-5, \ldots, -2, 0, \ldots, +5\}$ with $t = -1$ as the omitted reference period, and γ_t are calendar-year fixed effects that absorb common macroeconomic shocks (e.g., labour market conditions, policy changes). The coefficients β_k trace the evolution of the outcome relative to the year before first birth.

I estimate three versions of this equation: (1) the binary employment indicator for the full sample, which captures the extensive margin; (2) weekly hours conditional on employment, which captures the intensive margin; and (3) annual earnings conditional on employment. Heteroskedasticity-robust standard errors are reported throughout.

Estimand. The primary estimand is a descriptive event-time profile for married first-birth women observed in the JHPS/KHPS analytic panel. Under this specification, each β_k combines within-person change and composition change from entry, exit, and differential observation across event time. The coefficients should therefore be read as sample-average trajectories around first birth, and not as causal treatment effects.

Data limitations. The JHPS/KHPS do not record several variables that would be needed for causal channel identification: employer-level leave eligibility or enforcement records, contract renewal decisions or HR accommodation behaviour, exact childbirth date relative to survey interview, fine-grained childcare availability or slot constraints at prefecture or municipality level, or a harmonized education measure for the full analytic sample.[3] These gaps limit the paper's ability to isolate specific channels and motivate the descriptive and predictive framing used throughout.

Validation diagnostics. Identification is evaluated with three complementary checks that are reported centrally in the paper: (i) a near-lead test at $t = -2$ in the full sample, (ii) a trimmed pre-period specification that removes far leads, and (iii) a balanced pre-support specification restricted to women observed across all pre-birth event times. I also report a JHPS/KHPS split check to verify that the key timing moments are not driven by pooling alone. Due to timing noise around the survey birth-report window, I choose to treat the $[-2, +2]$ window as the primary interpretation window for the birth-year break and present $[-5, +5]$ estimates as extended descriptive context.

Specification choice. The main specification uses calendar-year fixed effects without individual fixed effects. The year-FE coefficients β_k are interpreted as descriptive sample-average trajectories relative to $t = -1$, and not within-person causal effects. In this short, unbalanced event window, adding individual fixed effects makes the design matrix near-collinear and standard errors uninformative (e.g., at $t = 0$: estimate -0.171, SE $>$ 3,100). I therefore report individual-FE results in the appendix for transparency and keep year-FE descriptive estimates in the main text.

3.5.2 Full-time/part-time decomposition

To understand whether the employment recovery masks a compositional shift, I supplement the event-study estimates with transition matrices that track changes in hours-based employment status. In this part of the analysis, full-time and part-time are defined by hours worked rather than by contract status; regular and non-regular remain separate

[3] Education can only be recovered for a restricted subset through the supplementary Newlyweds module, which is not part of the main questionnaire in every wave and therefore does not cover the full analytic sample. Only 13.4% of the analytic sample can be matched to this module; Appendix D.4 reports the coverage diagnostic and matching details.

contract-type classifications. I classify women who are employed full-time at $t = -1$ and observe their status at $t = +1$, $t = +3$, and $t = +5$. I report both the full-sample transitions and the transitions restricted to women with no second birth by the horizon date. The transition matrices are descriptive tabulations. They do not condition on covariates or adjust for attrition, and they should be read as documenting the composition of post-birth employment rather than estimating causal effects of childbirth on job type.

I also examine subsequent trajectories among women who are part-time at $t = +1$. Specifically, I track the shares who move to full-time work, remain part-time, or exit employment by $t = +5$. This complements the event-study estimates by showing whether part-time re-entry functions as a stepping stone back to full-time work or instead persists as a longer-run downgrade.

3.5.3 Risk-stratification regressions

To identify which job characteristics predict pre-birth exit, I estimate logistic regressions of the form:

$$\Pr(\text{quitter}_i = 1) = \Lambda(\mathbf{X}_i'\boldsymbol{\beta}), \tag{2}$$

where $\Lambda(\cdot)$ is the logistic CDF and $\mathbf{X}_i$ includes non-regular employment status, small firm size, commute time, and the husband's overwork indicator. I estimate five nested models (M1-M5) to show how each predictor contributes independently.

The dependent variable is quitter status on the pre-birth margin at $t = -2$; see Section 3.3 for the coding details. In practice, this indicates whether a woman is flagged as having exited employment between the $t = -2$ and $t = -1$ survey waves. The regression sample is restricted to the 185 women in the complete-case mechanism sample with non-missing values on the included predictors; among them, 13 exit employment (7.0 percent). This sample is smaller than the full event-study support at $t = -2$ because the mechanism regressions require complete covariate information. Because the event count is low, I treat these estimates as supporting evidence on early risk stratification rather than as a central result. The full nested specifications are reported in Appendix D.8.

The low event count - 13 exits against 4 or 5 predictors - raises legitimate concerns about separation and small-sample bias. I address these in two ways. First, I apply Firth's penalized likelihood estimator (Firth, 1993), which reduces finite-sample bias in rare-event logit settings. Second, I construct bootstrap confidence intervals from 2,000 resampled datasets, which provide nonparametric inference that does not rely on asymptotic normality. I report both corrections alongside the standard estimates in Appendix D.8.

3.5.4 Interpretation and timing

Births are reported for events occurring between February of the previous year and January of the survey year. Accordingly, $t = 0$ corresponds to the survey year in which a birth was reported, and $t = -1$ is the prior survey year. The event-study estimates therefore capture discrete changes around the reported birth window rather than exact conception or delivery dates.

The estimates are descriptive event-time trajectories relative to $t = -1$. They combine within-person change with composition change across event time and should not be interpreted as causal treatment effects. The pre-period coefficients at far leads ($t = -5$ to $t = -3$) are negative and jointly significant ($p < 0.001$), reflecting rising employment as young women enter the labour market rather than pre-birth anticipation. The critical pre-trend test is at $t = -2$: the coefficient is -0.012 (SE 0.031, $p = 0.70$), confirming that employment is stable in the year before first birth. This near-lead result is stable in the trimmed pre-period specification ($p = 0.72$) and remains economically small in the balanced pre-support sample ($p = 0.078$). Outcomes are measured at survey timing, so $t = 0$ is the birth-report wave, not a full post-birth calendar year. I therefore interpret the results as descriptive changes in level and composition around birth, especially over the $[-2, +2]$ window.

3.5.5 Second-birth censoring

A substantial share of women in the sample have a second birth within the event window: 19.2 percent by $t = +2$, 33.8 percent by $t = +3$, and 44.1 percent by $t = +5$. Subsequent births complicate the interpretation of medium-run coefficients because the "post-first-birth" trajectory includes responses to the second birth. I address this in two ways. First, I report the main event study on the full sample and note that late-period coefficients (particularly $t = +3$ through $t = +5$) reflect a mixture of first- and second-birth responses. Second, I estimate a specification that excludes observations after a second birth; this yields a $t = +5$ coefficient of -0.003 (versus -0.028 in the full sample), which suggests that part of the muted recovery reflects subsequent fertility. The transition matrices are also reported separately for the "no second birth by horizon" subsample.

3.6 Interpretation, identification, and scope

This section clarifies the estimand and the scope of interpretation. The main event-study specification, which includes calendar-year fixed effects, reports descriptive trajectories in an unbalanced panel rather than within-person causal treatment effects. Individual-fixed-effects versions are shown in the appendix as sensitivity checks. The near lead at $t = -2$ and the birth-year drop at $t = 0$ are stable in sign and magnitude across variants.

The event-time measurement follows the survey reporting window (February-January). This implies that $t = -1$ may include late-pregnancy or early-childcare overlap for some women. I therefore interpret the key timing contrast as a near-lead check at $t = -2$ and a discrete break at $t = 0$, rather than as a clean pre-treatment period at every pre-birth event time.

The mechanism regressions target pre-birth exit risk. The regressors are job characteristics measured while employed at $t = -2$, and the dependent variable indicates whether a woman is flagged as having exited employment between $t = -2$ and $t = -1$. These models are designed to explain who exits early in the pre-birth margin; they are not designed to identify childcare-margin effects at $t = 0$.

The $t = -2$ and $t = 0$ margins therefore reflect different risk sets. The former is more marriage-related, while the latter is predominantly childcare-related. Subgroup coefficients at birth should not be expected to mechanically mirror the pre-birth mechanism logits.

The intensive-margin results for hours and earnings are conditional on employment and therefore combine within-person adjustment with composition effects in who remains employed. I interpret these coefficients as evidence on the composition and quality of post-birth work rather than as pure within-person treatment effects.

More broadly, these margins are not independent in Japan's labour market structure. Contract type, hours, and employment continuity are tightly bundled features of the same employment tier. I therefore treat non-regular status as a marker of labour-market position and interpret post-birth hours and earnings as downstream expressions of that position, rather than as separate channels that can be cleanly isolated in this design.

Finally, the subgroup coefficients in the heterogeneity tables are estimated in separate subgroup event studies. They are useful for ranking exposure across groups, but they are not intended as a decomposition identity and need not aggregate arithmetically to the pooled coefficient.

4 Results: aggregate break and pre-birth exit composition

The child-penalty literature often interprets pre-birth employment declines as anticipatory adjustments: women who expect to bear children reduce their labour supply in the years before birth (Kleven et al., 2019b; Kuziemko et al., 2018). In this sample, the aggregate near lead at $t = -2$ is flat; what differs pre-birth is the composition of exits among the minority who leave before birth.

4.1 Aggregate employment trajectory

I focus first on the primary outcome: the binary employment indicator, which equals one if a woman reports positive weekly hours. The employment rate at $t = -1$ is 50.3 percent (N = 656). Against this baseline, Figure 2 reports the event-study estimates.

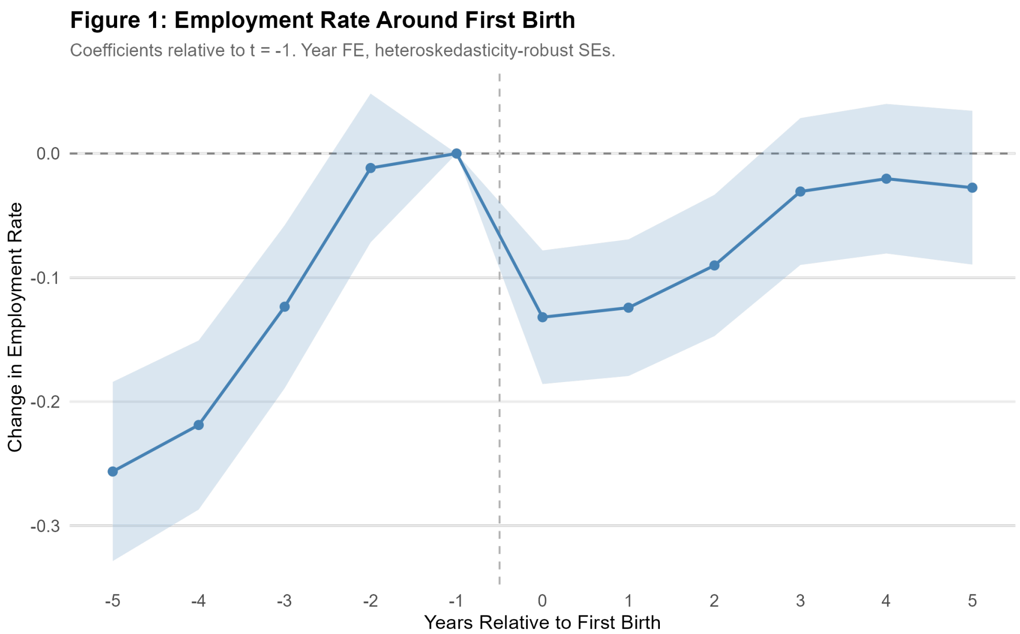

Figure 2: Event-study estimates of maternal employment around first birth (year fixed effects; heteroskedasticity-robust standard errors). Each point is the coefficient $\hat{\beta}_k$ relative to $t = -1$; bars are 95% confidence intervals. The baseline employment rate at $t = -1$ is 50.3%. Because $t = -1$ may overlap late pregnancy for a subset of women under the survey timing window, the primary interpretation window is $[-2, +2]$. Support is unbalanced, with N ranging from 186 ($t = -5$) to 656 ($t = -1$). Inference is unchanged under woman-level clustered standard errors (Appendix D.12).

The trajectory shows a clear pattern: the near lead at $t = -2$ is close to zero (-0.012, $p = 0.70$), followed by a sharp drop of 13.2 percentage points at $t = 0$ ($p < 0.001$). To put this in perspective, the coefficient implies that the employment rate falls from approximately 50 to 37 percent at the birth year - roughly 87 fewer women working among the 662 in the sample. Employment partially recovers in subsequent years, with coefficients of -0.124 at $t = +1$, -0.090 at $t = +2$, and -0.031 at $t = +3$. By $t = +5$ the coefficient is -0.028 (SE 0.032), statistically indistinguishable from zero in the pooled specification; this late-horizon point is panel-sensitive and should be interpreted with caution. Because $t = -1$ can overlap late pregnancy for a subset of respondents, the baseline may already embed partial pre-birth adjustment; if so, the estimated $t = 0$ drop is conservative relative to a fully pre-pregnancy reference.

The extended pre-birth trajectory ($t = -5$ to $t = -3$) shows large negative coefficients (-0.256, -0.219, -0.124), which might appear to indicate a pre-birth decline. In fact, the pattern reflects rising employment as women enter the labour market. The lower employment rates at earlier event times reflect both the life-cycle trajectory of women in their twenties and compositional selection - fewer women are observable at $t = -5$ (N = 186) than at $t = -1$ (N = 656), and those observable early are a selected subsample. The key identification result is that by $t = -2$, when the sample stabilizes at N = 453, employment is flat relative to $t = -1$. The support profile is materially unbalanced (N = 186 at $t = -5$, 662 at $t = 0$, 454 at $t = +5$), and observability at $t = +5$ is related to baseline employment. The IPW sensitivity that targets observability at $t = +5$ leaves the near-lead and at-birth break intact, while making the late-horizon coefficient more negative and marginally significant ($t = +5$: -0.057, $p = 0.076$). I therefore treat late-horizon recovery as composition-sensitive.

The recovery, however, masks a profound compositional shift. At baseline ($t = -1$), full-time workers comprise 30.9 percent and part-time workers 19.4 percent of the sample. After birth, the composition shifts sharply: among women who held full-time positions at $t = -1$, only 45.7 percent remain full-time at $t = +1$, while 33.7 percent exit employment and 20.6 percent shift to part-time. By $t = +5$, only 42.2 percent of those initially full-time remain so. Section 6 documents this compositional transformation in more detail.

4.2 Quit reasons: marriage versus childcare

Table 4 provides the key descriptive evidence on why women exit. Among exits at $t = -2$ (N = 29), 41.4 percent cite marriage as the reason, while only 6.9 percent cite childcare. The pattern reverses sharply at $t = 0$ (N = 113 exits): 85.0 percent cite childcare, and only 2.7 percent cite marriage. Missing reason codes are prevalent (51.7 percent at $t = -2$, 12.4 percent at $t = 0$), so these figures should be interpreted as suggestive rather than definitive.

This pattern is difficult to square with a pure anticipation interpretation. If women were leaving at $t = -2$ primarily in preparation for childbirth, childcare should appear more prominently among the stated reasons. Instead, the women who leave two years before first birth most often cite marriage - an event that typically precedes first birth by one to two years and that carries its own normative expectations about women's labour-force participation in Japan (Ochiai, 1997; Yu, 2009).

The distinction is important analytically because it suggests that the pre-birth exit margin and the birth-year break reflect different institutional pressures and may therefore respond to different policy margins. Childcare expansion alone is unlikely to prevent marriage-related exits, just as changing marriage norms alone would not address childcare-related exits at birth. A single pooled "child penalty" estimate can therefore obscure meaningful heterogeneity in timing and mechanism.

Table 4: Quit reasons by event time among women classified as exits in the quit-reason module. Percentages are within event-time exit groups, with missing reason responses shown explicitly. At $t = -2$, marriage is the leading reported reason (41.4%); at $t = 0$, childcare is dominant (85.0%). Missingness is high at $t = -2$ (51.7%) and lower at $t = 0$ (12.4%).

	$t = -2$ (N = 29)		$t = 0$ (N = 113)	
Reason	N	%	N	%
Marriage	12	41.4	3	2.7
Childcare	2	6.9	96	85.0
Other/Missing	15	51.7	14	12.4

4.3 Reentry penalty and sequencing

Although employment rates recover after the birth year, the composition of work changes in ways that imply a reentry penalty. Conditional on working, the number of hours fall by roughly 10 hours per week by $t = +5$ (coefficient -10.1, $p < 0.001$), and earnings remain substantially below pre-birth levels even when employment returns (coefficient -96.2 man-yen at $t = +5$, $p < 0.001$). This pattern is consistent with reentry into non-regular or lower-hour jobs rather than a return to pre-birth career tracks. In the Japanese dual labour market, job status and firm protection are relatively sticky; exit around birth can move mothers onto a lower-quality track with weaker wage progression, which shows up as persistent earnings shortfalls even as employment rates do rebound. Importantly, the long-run trajectory is also shaped by subsequent births: censoring at second birth attenuates the late-period coefficients, suggesting that part of the muted recovery reflects additional fertility rather than a single, permanent shock.

With the timing pattern established, I turn in the next section to the question of who exits and what structural features of the labour market predict departure at each margin - first as sequencing context for pre-birth exits, then as the centrepiece risk-stratification analysis at the childbirth margin.

5 Risk stratification by labour-market tier

The centrepiece risk-stratification result is at the childbirth margin ($t = 0$), where the sample is largest (N = 330; 124 exits) and inference is sharpest. Before presenting that analysis, I examine the pre-birth margin ($t = -2$) for sequencing context: in this lower-power sample (N = 185; 13 events), the aggregate profile is flat but the composition of early exits is informative - marriage is the most common stated reason, while at $t = 0$ childcare dominates. The pre-birth models are retained as supporting evidence for the risk-stratification gradient, not as the primary mechanism identification. The empirical question is straightforward: does exit risk cluster by labour-market position? In the data, it does at both margins. Risk varies sharply by contract tier and firm size, which is consistent with segmentation in employment continuity.

5.1 Descriptive patterns: job type and firm size

I begin with bivariate comparisons from the firm-size by job-type crosstab. Among employed women at $t = -2$ who hold non-regular contracts, 12.2 percent exit employment (11 of 90). Among regular employees, 1.9 percent do (2 of 108). The interaction of job type and firm size is also informative. Among non-regular workers at small firms (fewer than 30 employees), the exit rate reaches 20.0 percent (6 of 30). Among regular workers at large firms (100+ employees), the rate is 0.0 percent (0 of 63). This contrast suggests that the pre-birth exit risk is concentrated in the least protected labour-market cells. In some jobs, childbirth can plausibly be absorbed through leave, temporary absence, and return pathways; in others, those pathways appear much weaker, so childbirth is more likely to coincide with employment exit.

Table 5: Pre-birth exit rates at $t = -2$ by labour-market cell (contract type × firm size), based on women with complete job characteristics (N = 198). Non-regular workers in small firms show the highest exit rate (20.0%), while regular workers in large firms show 0.0% in this sample.

Firm Size	Contract Type	N	N quit	% quit
Small (<30)	Non-Regular	30	6	20.0
Small (<30)	Regular	15	1	6.7
Medium (30–99)	Non-Regular	6	1	16.7
Medium (30–99)	Regular	19	1	5.3
Large (100+)	Non-Regular	48	3	6.2
Large (100+)	Regular	63	0	0.0
Government	Non-Regular	6	1	16.7
Government	Regular	11	0	0.0

These bivariate contrasts could still reflect confounding across correlated job and household characteristics. I therefore estimate nested logistic specifications to assess whether contract tier and firm size continue to predict pre-birth exit risk once additional controls are added. Because this margin contains only 13 exit events, I treat the multivariate results as lower-power supporting evidence rather than as a central result; the full nested specifications are reported in Appendix D.8.

The pattern is directionally stable across those specifications. Non-regular status remains strongly associated with higher pre-birth exit risk, and small-firm employment is also positively associated with exit, while commute time and the husband's overwork do not add much predictive power once the woman's labour-market position is included. Given the low-event sample, the exact magnitudes should be read cautiously. The main takeaway is the ranking: early exit risk is concentrated in labour-market segments with weaker continuity around family formation.

5.2 Risk stratification at the childbirth margin ($t = 0$): absolute risks

Before presenting the childbirth-margin model, Table 6 compares pre-birth characteristics of regular and non-regular workers at $t = -1$. The two groups differ substantially: regular workers average 41.3 hours per week and 390 man-yen in annual earnings, while non-regular workers average 28.0 hours and 132 man-yen. Exit rates at $t = 0$ are 18.3 percent for regular and 61.6 percent for non-regular workers. These differences confirm that the regular/non-regular distinction captures a substantive divide in labour-market position before childbirth, not merely a labelling difference.

Table 6: Pre-birth characteristics by contract type at $t = -1$ (employed women)

	Regular	Non-Regular
N	169	125
Mean weekly hours	41.3	28.0
Full-time (%)	88.2	36.0
Mean annual earnings (man-yen)	390.2	132.1
Median annual earnings (man-yen)	363.5	101.0
Exit rate at $t = 0$ (%)	18.3	61.6

Notes: Sample restricted to women employed at $t = -1$ with non-missing contract-type classification. Regular = codes 1–3 (seishain, executive, family business). Non-Regular = codes 4–7 (part-time, arbeit, dispatched, contract). Full-time defined as $\geq$35 hours/week. Exit rate = share not employed at $t = 0$ among those observed at both $t = -1$ and $t = 0$.

Because the main exit spike occurs at childbirth, the primary risk-stratification model targets non-employment at $t = 0$ among women employed at $t = -1$ (N = 330; 124 exits, 37.6 percent). In this childbirth-margin sample, non-regular status at $t = -1$ is strongly associated with exit (OR = 7.50, $p < 0.001$), while the small-firm coefficient is positive but imprecisely estimated (OR = 1.55, $p = 0.126$). Small-sample corrections leave the qualitative ranking unchanged: the non-regular effect remains strong, while the small-firm effect is directionally positive but less precise.

Within this same risk set ($N = 330$), the $t = -1$ composition is 75 full-time regular women (22.7%), 22 full-time non-regular (6.7%), 17 part-time regular (5.2%), and 40 part-time non-regular (12.1%); contract type is missing for 176 women (53.3%). Here, full-time and part-time refer to hours worked, while regular and non-regular refer to contract status. This high missingness is not random in observables: missing contract type is concentrated among lower-hours and lower-income women, that is, profiles closer to marginal or non-regular attachment (Appendix D.9).

The absolute-risk table reports these values for transparency. Raw exit rates are 60.8 percent for non-regular workers versus 16.7 percent for regular workers. By profile, model-implied probabilities range from 14.6 percent for regular workers in large firms or government employment to 66.5 percent for non-regular workers in small firms, closely tracking the raw cell means.

The table-level rates use the 282 women with non-missing contract type and firm size at $t = -1$; by contrast, the childbirth-margin logit is estimated on all 330 women using missing indicators. Taken together, these results indicate that contract tier is the strongest observed predictor of childbirth-margin exit in this sample.

Table 7: Absolute-risk view of childbirth-margin exit: raw rates and model-implied probabilities at $t = 0$.

Profile / group	N	Exits	Exit rate
By contract type: Non-Regular	120	73	0.608
By contract type: Regular	162	27	0.167
By firm size: Large/Gov	175	53	0.303
By firm size: Small	107	47	0.439
By contract x firm: Non-Regular x Large/Gov	66	38	0.576
By contract x firm: Non-Regular x Small	54	35	0.648
By contract x firm: Regular x Large/Gov	109	15	0.138
By contract x firm: Regular x Small	53	12	0.226
Profile			Predicted Pr(exit at $t = 0$)
Regular x Large/Gov			0.146
Regular x Small			0.209
Non-Regular x Large/Gov			0.562
Non-Regular x Small			0.665

A missing-data sensitivity check shows that the non-regular gradient is highly stable across covariate treatments: OR = 7.50 (complete-case), 7.48 (missing-indicator), and 7.39 (high- information sample), all with $p < 0.001$. This indicates that the core childbirth-margin stratification is not a missing-data artefact. By contrast, small-firm coefficients remain positive but less stable across specifications, so I interpret firm-size effects as suggestive rather than definitive. A balance diagnostic (Appendix D.9, Table 38) confirms that contract-type missingness is concentrated among lower-hours and lower-income women, consistent with ambiguous classification in marginal employment, but the non-regular OR is invariant to missingness treatment. Under this pattern, complete-case estimates are more likely to under-represent the highest-risk non-regular segment than to mechanically inflate the OR.

5.3 Small-sample robustness

The pre-birth mechanism estimates are based on only 13 exit events in the complete-case sample, so precision is limited and the results should be treated as supporting evidence. To assess whether the core pattern is driven by small-sample bias, I apply Firth's penalized likelihood estimator and construct bootstrap confidence intervals from 2,000 resampled datasets. These corrections are reported for the parsimonious two-predictor specification

(Model 2); with only 13 events, the saturated model is too parameter-rich for reliable penalized estimation.

The corrected estimates preserve the qualitative ranking. The Firth estimates are attenuated relative to the standard logit, as expected, but both predictors remain statistically significant: non-regular employment (OR = 5.32, $p = 0.009$) and small-firm employment (OR = 3.84, $p = 0.021$). The bootstrap intervals likewise exclude 1 for both predictors: non-regular $[1.68, 19.66]$($p = 0.008$) and small firm $[1.10, 18.34]$ ($p = 0.038$). The intervals are wide, which reflects the low event count, but the direction of the relationship is stable. I therefore treat the pre-birth mechanism evidence as suggestive and supportive rather than as a central pillar of the argument. The stronger risk-stratification result remains the childbirth-margin model at $t = 0$, which is estimated on a much larger risk set and is robust to missing-data treatment.

Table 8: Small-sample robustness for the parsimonious pre-birth mechanism model (Model 2), with exit at $t = -2$ as the outcome: standard logit, Firth penalized likelihood, and nonparametric bootstrap (2,000 resamples).

	OR	p	95% CI	Method
Non-regular employment				
Standard logit	6.41	0.019	–	MLE
Firth	5.32	0.009	–	Penalized ML
Bootstrap	–	0.008	[1.68, 19.66]	2,000 resamples
Small firm				
Standard logit	4.14	0.025	–	MLE
Firth	3.84	0.021	–	Penalized ML
Bootstrap	–	0.038	[1.10, 18.34]	2,000 resamples

5.4 Ruling out selection

A natural concern is that women in non-regular jobs may differ in ways that independently predict exit (cf. Budig and England, 2001; Correll et al., 2007). I address this concern with descriptive comparisons and leave-coding sensitivity checks.

First, I compare pre-birth labour-force attachment between quitters (N = 29) and stayers (N = 262). Quitters have an employment rate of 55.2 percent at $t = -1$, compared with 61.0 percent for stayers ($p > 0.3$). On this margin, women who exit do not appear less attached to the labour force than those who remain.

Second, I compare pre-birth income. Quitters earn 148.4 man-yen versus 302.7 man-yen for stayers. This income measure is observed at the $t = -1$ survey wave, which follows the quit interval, so it may already reflect post-exit adjustment or reduced hours. Even so, the pattern is informative. In a simple labour-supply framework, lower earnings could reflect either voluntary sorting into exit or structural vulnerability. Two features of the data favour the latter interpretation. First, the multivariate results point to contract type and firm size rather than earnings as the relevant predictors. Second, the quit-reason evidence at $t = 0$ shows that 85 percent of exits cite childcare rather than voluntary preference. Taken together, these patterns suggest that low income marks jobs that are more vulnerable to the motherhood transition, rather than revealing a purely preference-based exit margin.

A supplementary diagnostic compares the 13 complete-case quitters to the 16 quitters excluded because of missing covariates. Both groups are concentrated in small firms (69.2 percent among included cases and 100 percent among excluded cases), while contract type is missing for excluded cases by construction. This limits precision but does not suggest that the complete-case mechanism sample is selecting a qualitatively different group of exits.

I also examine childcare-leave take-up directly. In the partner-linked analytic sample, only 1 of 661 partners reports taking childcare leave (0.15 percent). For mothers, leave take-up is higher but still limited: in the broader leave-analysis sample, 250 of 1,183 women (21.1 percent) ever report taking childcare leave.[4] This percentage should not be interpreted as unconstrained choice. The leave item is observed only when women remain employed and report temporary absence from work; women who exit employment are coded as non-employed and therefore do not enter the leave-response risk set. Observed leave take-up thus reflects both policy availability and whether an employment relationship survives through childbirth. For many women, formal leave provisions are no longer operational once the employment relationship has ended.

To assess potential misclassification at childbirth, I run a sensitivity check that reclassifies women coded as non-working at $t = 0$ as employed if they report leave take-up near birth. This affects 22 women. The $t = 0$ event-study coefficient attenuates from -0.1319 (SE 0.0275) to -0.0988 (SE 0.0276), a +3.31 percentage-point shift, while remaining strongly negative in both cases ($p < 0.001$). This suggests that leave-related coding contributes to measured exit levels but does not overturn the birth-year break (Appendix D.9, Table 39).

Taken together, these checks suggest that labour-market position is a stronger observed predictor of exit risk than the household variables measured here. They do not eliminate the possibility of unobserved selection into labour-market tiers, but they indicate that the main gradient is not reducible to the observable differences examined here.

5.5 Observed exit probabilities by labour-market cell

The odds ratios in the preceding subsections are informative about relative risk, but they do not convey the absolute magnitudes as directly as the raw cell rates. Table 5 therefore

[4] The event-study analytic sample has 662 mothers; the partner-linked father-leave denominator is 661; and the leave-analysis denominator (1,183) comes from the broader first-birth-linked leave module used for leave take-up tabulations.

reports unconditional observed exit rates by contract-type × firm-size cell. These are descriptive rates, not model-based predictions.

The observed rates form a steep gradient. Regular workers at large firms have a 0.0 percent exit rate (0 of 63), meaning that no exits are observed in that cell in this sample. The corresponding rates are 5.3 percent for regular workers at medium firms (1 of 19) and 6.7 percent for regular workers at small firms (1 of 15), which is consistent with weaker absence- management capacity outside large firms. Among non-regular workers, the rate is 6.2 percent at large firms (3 of 48), 16.7 percent at medium firms (1 of 6), and 20.0 percent at small firms (6 of 30). The non-regular small-firm cell is therefore the most exposed labour-market position in the sample, with one in five women exiting employment before first birth.

The range from 0.0 to 20.0 percent is substantial. Looking only at contract type and firm size, the observed exit rate spans from none in the regular-large-firm cell to one in five in the non-regular-small-firm cell. In concrete terms, if 100 non-regular workers at small firms faced the same exit rate as this sample, about 20 would leave their jobs before the birth year, compared with no observed exits among regular workers at large firms. In this sample, pre-birth extensive-margin exit risk is therefore highly concentrated in a narrow set of labour-market positions.

5.6 The time-budget constraint

The risk models indicate which jobs are most exposed to exit. A separate question is why women in those jobs leave. In the logistic models, the husband's overwork indicator does not add much explanatory power once the wife's job type is included (Model 4: OR = 1.21, $p = 0.749$). I do not interpret this null as strong evidence against household constraints. In this sample, husbands' hours are uniformly high, which limits the variable's ability to explain cross-sectional differences in exit risk once labour-market position is already accounted for.

The descriptive comparisons at $t = -1$ are consistent with that interpretation. Quitters' husbands work an average of 49.9 hours per week versus 46.9 hours for stayers ($p = 0.30$, N = 271), and the difference is not statistically precise. Commute time, if anything, runs in the opposite direction: quitters' mean one-way commute is 21.4 minutes versus 30.2 minutes for stayers ($p = 0.11$, N = 172). The exiting group is disproportionately drawn from non-regular local jobs with short commutes, which is consistent with commute time capturing job quality rather than pure household burden.

The post-birth time-use data make the household constraint more visible. At $t = +1$, mothers average 82.7 hours per week of childcare, while fathers contribute 13.2 hours (Figure 14). The wife-to-husband ratio is 6.3-to-1. Husbands in the sample work a mean of 47 hours per week, and 52 percent work 50 or more hours. Once commuting time is added, the typical husband has limited remaining time available to absorb a substantial share of infant care.

For mothers, the constraint is more binding still. At 82.7 hours per week of childcare - nearly 12 hours per day, seven days per week - combining infant care with employment is difficult without access to alternative care, employment continuity, or reduced-hours work. In the sample, 89.6 percent of households are nuclear, childcare leave is taken by only 21.1 percent of mothers, and work-from-home is available at only 10.1 percent of employers. For a non-regular worker at a small firm with no co-residing grandparents and no work-from-home option, the logistics of combining employment with infant care are consistent with a binding time-budget constraint. This interpretation is not separately identified from the data - the evidence is descriptive rather than a direct causal test - but it offers a plausible account of why exit is concentrated in the least protected labour-market segments.

I therefore interpret the motherhood penalty in this sample as plausibly constrained by labour-market structure rather than as purely preference-based. The women who exit are not necessarily choosing leisure over work; in these descriptive data, they appear to face a time-budget constraint in which infant-care requirements can exceed the hours that remain available once the husband's work schedule and limited employer flexibility are taken into account.

5.7 Heterogeneity at the birth transition

I next examine heterogeneity in the employment penalty across labour-market positions at $t = 0$. Figures 3–5 report the descriptive employment change, normalized to $t = -1$, by subgroup. The gradient is steep. Non-regular workers at large firms experience the largest penalty (-12.1 percentage points, N = 33), followed by all non-regular workers (-9.8 pp, N = 70). Regular workers - whether at large or small firms - experience penalties of only -1.6 to -1.9 pp. Small-firm workers as a group show essentially no aggregate penalty (+0.4 pp, N = 115), though this masks heterogeneity by contract type within small firms (-4.5 pp for non-regular, N = 35; -1.9 pp for regular, N = 54; the remaining 26 women have missing contract-type data). Because these coefficients come from separate subgroup event studies, they do not aggregate arithmetically to the pooled small-firm coefficient.

For these heterogeneity event studies, subgroup definitions are based on pre-birth job information at $t = -2$ rather than $t = -1$. Some women in each subgroup are no longer employed by $t = -1$, so subgroup baselines are below 100 percent employment.

The fact that the largest penalty falls on non-regular workers at large firms - rather than non-regular workers at small firms - is consistent with a composition effect. Large firms retain more women in employment up to the birth transition, while small-firm non-regular workers appear more likely to have already exited or shifted onto lower-hour trajectories before birth. The pre-birth mechanism evidence in Appendix D.8 is consistent with this interpretation: small-firm non-regular workers are disproportionately selected out before the birth year, so the heterogeneity at $t = 0$ is estimated on a different and more selected risk set than the pre-birth logit models. The main pattern is that contract type matters more than firm size. The penalty is -9.8 pp for all non-regular workers and -1.6 pp for all regular workers, a difference of roughly five to one. This heterogeneity corroborates the childbirth-margin logistic results in Section 5.2 and provides descriptive evidence that the dual labour market channels the motherhood penalty.

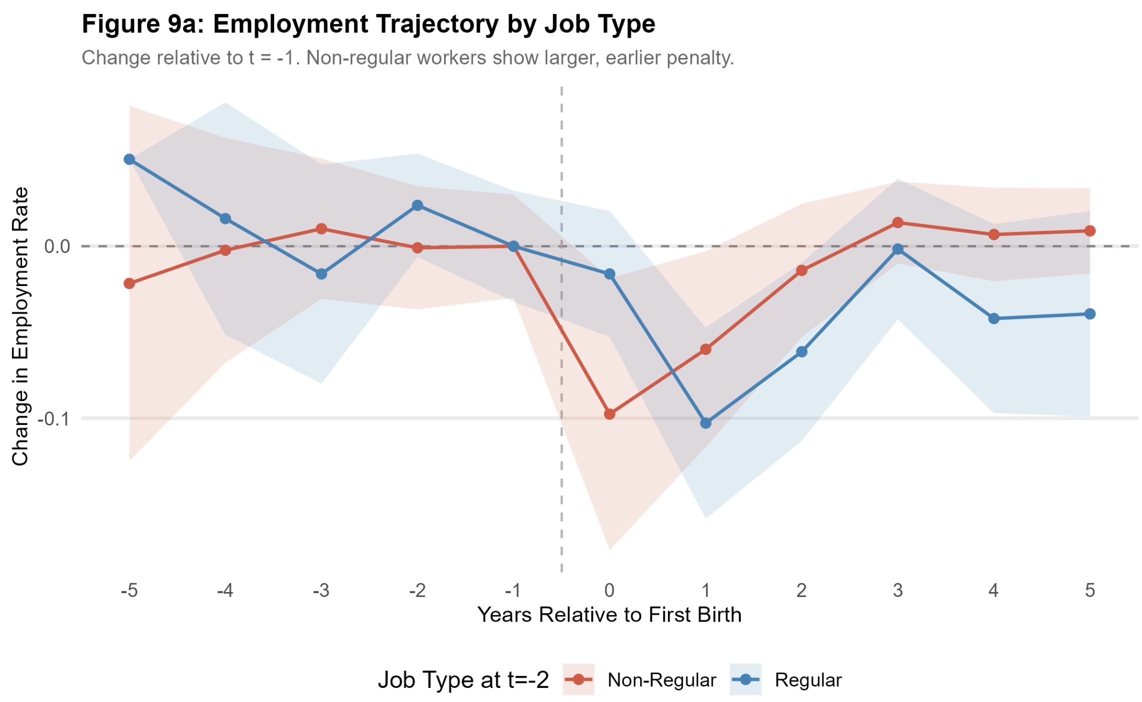

Figure 3: Employment event-study heterogeneity by contract type (year fixed effects; robust standard errors). Regular and non-regular refer to contract status, not to hours-based full-time or part-time categories. Coefficients are percentage-point changes relative to $t = -1$. At $t = 0$, non-regular workers show a -9.8 pp decline versus -1.6 pp for regular workers.

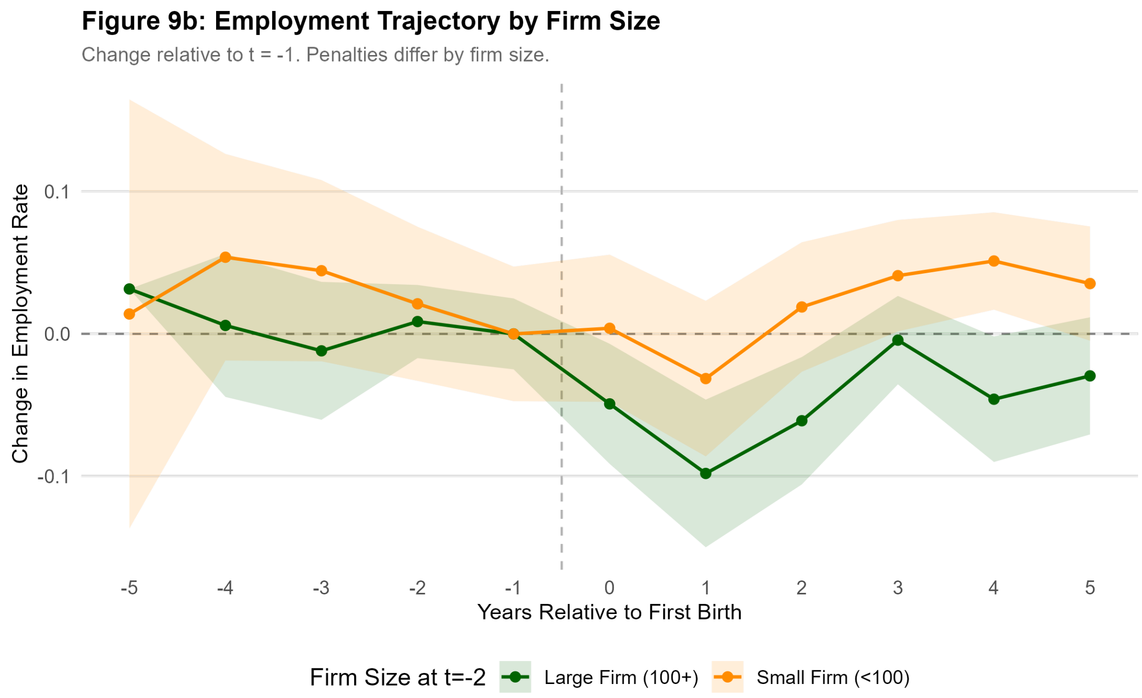

Figure 4: Employment event-study heterogeneity by firm size (year fixed effects; robust standard errors). Coefficients are percentage-point changes relative to $t = -1$. At $t = 0$, large-firm workers show a -4.9 pp decline versus $+0.4$ pp for small firms.

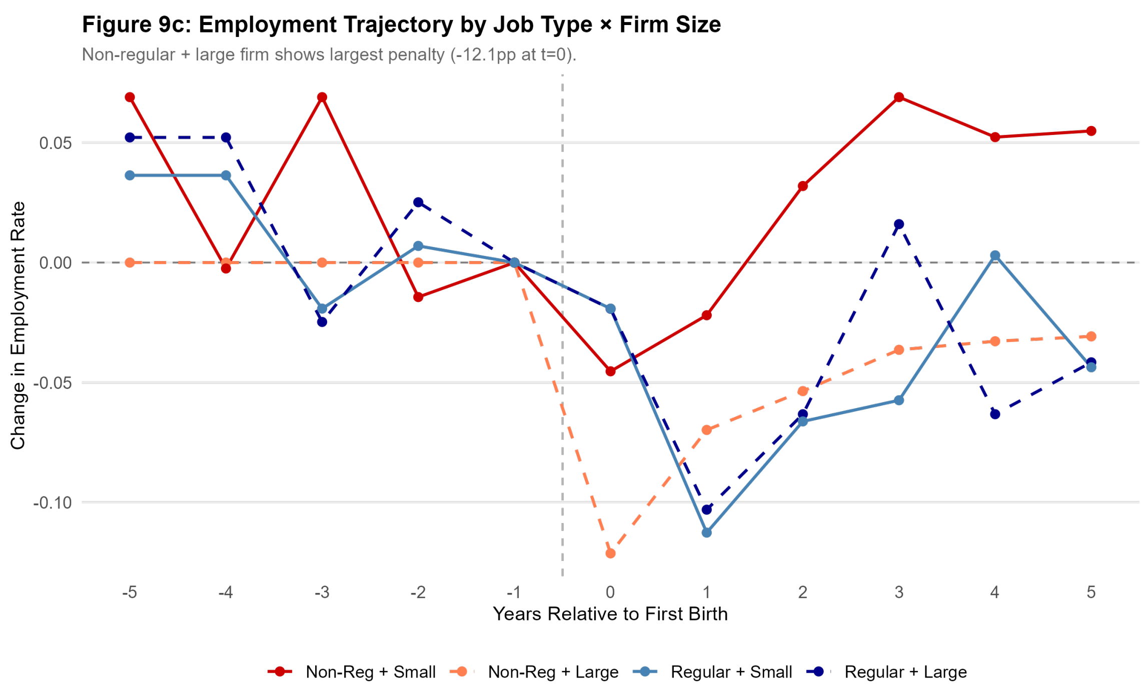

Figure 5: Employment event-study heterogeneity by contract type × firm size (year fixed effects; robust standard errors). Coefficients are percentage-point changes relative to $t = -1$. The largest at-birth decline appears for non-regular workers in large firms (-12.1 pp at $t = 0$).

5.8 Geographic non-variation

The heterogeneity results above show that the penalty varies by contract type and firm size. A natural question is whether it also varies by geography - whether the penalty is concentrated in metropolitan areas with intense corporate work cultures, or in rural areas with stronger traditional norms. I split the sample by metropolitan status (Kanto and Kinki regions, which contain Tokyo and Osaka, versus all other regions) and estimate separate event studies with the same year-FE specification. [5]

The result is a null. At $t = 0$, the employment penalty is -13.6 percentage points in metropolitan areas (N = 355 women, 2,793 obs, $p < 0.001$) and -12.9 percentage points in non-metropolitan areas (N = 307 women, 2,419 obs, $p = 0.002$). The difference of 0.7 percentage points is economically negligible. Pre-trends are clean in both subsamples: the $t = -2$ coefficient is $+0.031$ ($p = 0.46$) in metropolitan areas and -0.063 ($p = 0.16$) in non-metropolitan areas. Recovery paths are also similar, with both groups reaching coefficients near zero by $t = +3$.

A secondary split by city size (21 designated major cities versus other areas) shows a suggestive difference - the at-birth penalty is -9.9 pp in major cities versus -14.8 pp elsewhere - but this comparison is complicated by differential pre-trends (the major-city $t = -2$ coefficient is $+0.097$, $p = 0.07$) and is not interpretable as a clean contrast.

The metropolitan split provides limited evidence of large geographic divergence in the estimated penalty, but this should be interpreted cautiously. A Kanto/Kinki versus other-regions comparison is coarse and cannot capture within-region heterogeneity in wages, childcare access, commuting constraints, or local labour demand. At the same time, the geographic null is itself informative: if local mechanisms differ (e.g., marriage-related exits more salient in some areas, labour-market segmentation more salient in others), similar aggregate penalties can still emerge. The results are therefore consistent with the dual labour market operating nationwide, including beyond Tokyo and Osaka, while leaving open geographic heterogeneity in the composition of underlying channels.

Regional disparities in labour-market conditions are substantial in Japan: prefectural indicators show persistent dispersion in female employment and earnings levels (Statistics Bureau of Japan, 2025; Ministry of Health, Labour and Welfare, 2025). Childcare constraints also vary geographically, as reflected in differences in childcare capacity and waiting-child indicators across jurisdictions (Children and Families Agency, Government of Japan, 2024c,a). However, because childcare fees are locally administered and means-tested, there is no single harmonised prefecture-level fee series directly comparable to the labour outcomes used here (OECD, 2025b). Appendix Table 23 reports the corresponding all-prefecture dispersion statistics for childcare infrastructure and utilisation and provides

[5] In the JHPS/KHPS researcher-use files, prefecture identifiers are not provided; available geographic variables are region and city size. I therefore use a metropolitan split (Kanto/Kinki vs. other regions) as the finest comparable geographic partition available in this study.

geographic context for the metropolitan split in the main analysis.

5.9 Household structure and education heterogeneity

To assess whether the main pattern is driven by household composition rather than labour market position, I examine subgroup trajectories by co-resident grandparent status and by patrilocal versus matrilocal household arrangements. I also report heterogeneity by education. These plots are descriptive and complement the main event-study estimates.

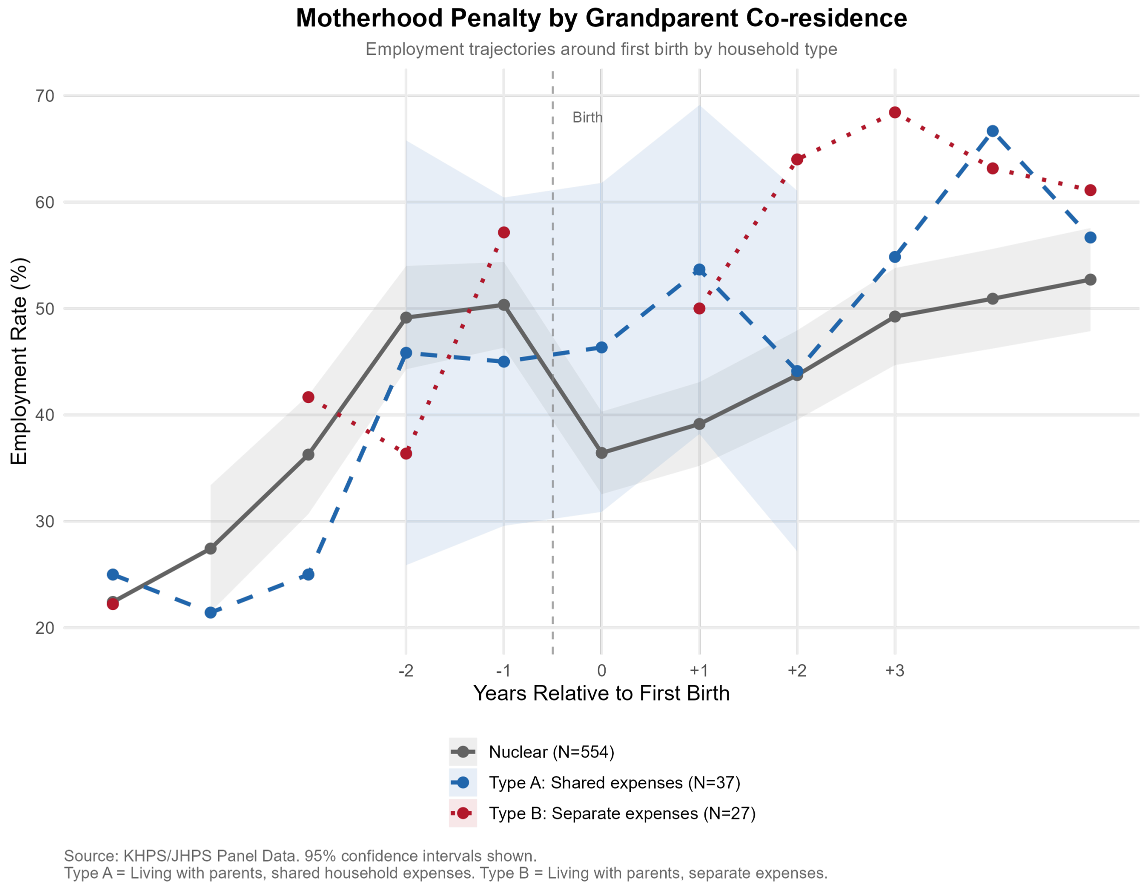

Figure 6: Additional heterogeneity diagnostic by co-resident grandparent status. The event-time coefficients are relative to $t = -1$ in the same year-FE framework. Both groups show a birth-year employment decline, with magnitude differences but no sign reversal.

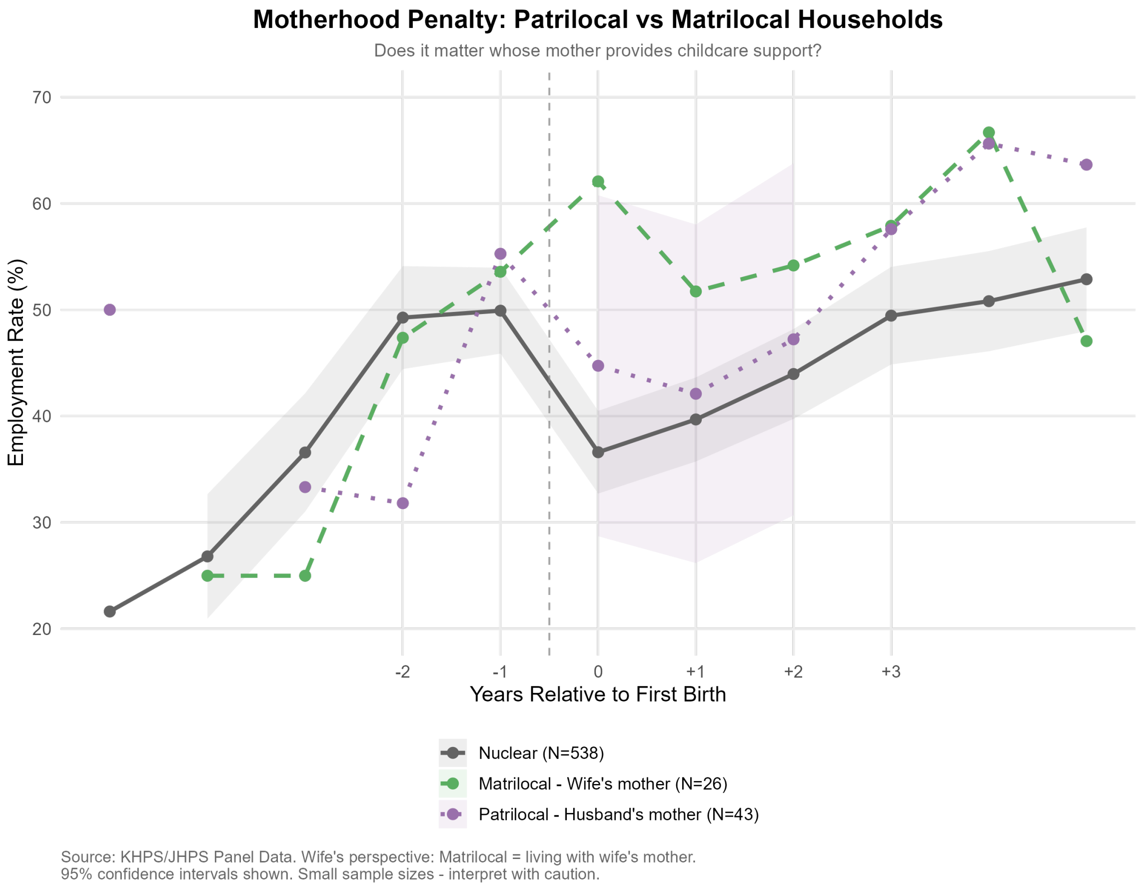

Figure 7: Additional heterogeneity diagnostic by patrilocal versus matrilocal household structure. The event-time coefficients are relative to $t = -1$ in the same year-FE framework. Both household structures exhibit the same qualitative birth-related employment decline.

6 The penalty's transformation: from exit to downgrade

The preceding sections focused on who leaves employment and why. But the motherhood penalty does not end with the exit decision. Women who remain employed - or who return to work after a period of absence - face a second penalty that operates through reduced hours, lower-quality re-entry, and earnings losses. This section documents the intensive margin of the penalty and shows that it comes to dominate the total penalty over time. Among employed women at $t = -1$, mean weekly hours are 34.1 (SD 16.5) and mean annual earnings are 268.9 man-yen (SD 206.9). These are the baselines against which the intensive-margin coefficients should be read.

6.1 Employment recovery and its limits

Figure 2 shows that employment partially recovers after the sharp decline at $t = 0$. The event-study coefficient narrows from -0.132 at $t = 0$ to -0.031 at $t = +3$ and -0.028 at $t = +5$. Women do return to work, but the aggregate series is misleading because the composition of employment changes substantially.

This timing is also consistent with Japan's childcare-system transitions. The years closest to birth coincide with the most constrained care period, especially ages 0-2, whereas care options broaden from age 3 onward, when enrolment pathways and subsidy coverage are typically more favourable (Children and Families Agency, Government of Japan, 2024b,a; OECD, 2025b). The observed recovery profile therefore aligns with a staged institutional environment rather than a single post-birth adjustment.

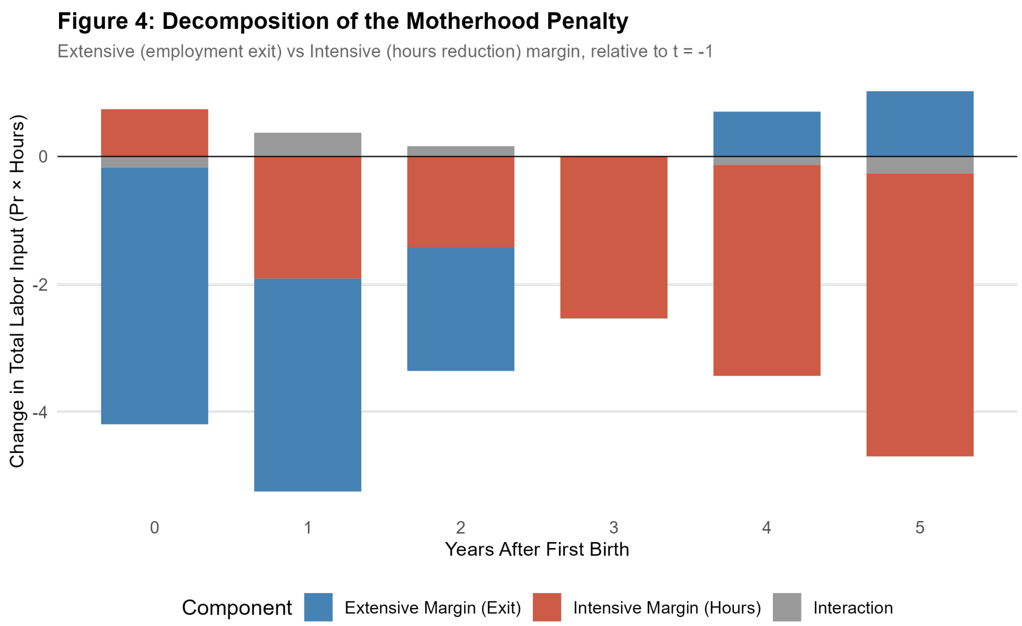

Figure 8: Decomposition of employment status into full-time and part-time shares around first birth. Full-time is defined as 35+ hours/week and part-time as employed below 35 hours/week; these are hours categories, not contract-type categories. Relative to $t = -1$, the full-time share falls sharply at birth and remains depressed, while the part-time share rises and accounts for most aggregate employment recovery (baseline at $t = -1$: 30.9% full-time, 19.4% part-time).

6.2 The collapse of full-time employment

Before first birth, 30.9 percent of all women in the sample are employed full-time. Here, full-time and part-time refer to hours worked rather than to regular versus non-regular contract status. Using the transition matrix, I track the employment status of women who held full-time positions at $t = -1$. At $t = +1$, only 45.7 percent remain full-time; 33.7 percent have exited employment and 20.6 percent have shifted to part-time. At

$t = +3$, full-time persistence improves to 53.3 percent, but by $t = +5$ it falls back to 42.2 percent, likely reflecting the impact of second births. When observations after second birth are censored, full-time persistence at $t = +5$ rises to 53.8 percent, which indicates that subsequent fertility explains part, but not all, of the late-horizon non-recovery.

Among women who shift to part-time at $t = +1$, relatively few transition back to full-time work. By $t = +5$, only 23.3 percent of women who were part-time at $t = +1$ have moved to full-time, while 18.6 percent have exited entirely. This pattern is consistent with reduced-hours re-entry operating less as a temporary bridge back to full-time work and more as a persistent downgrade in job quality.

6.3 Hours and earnings conditional on working

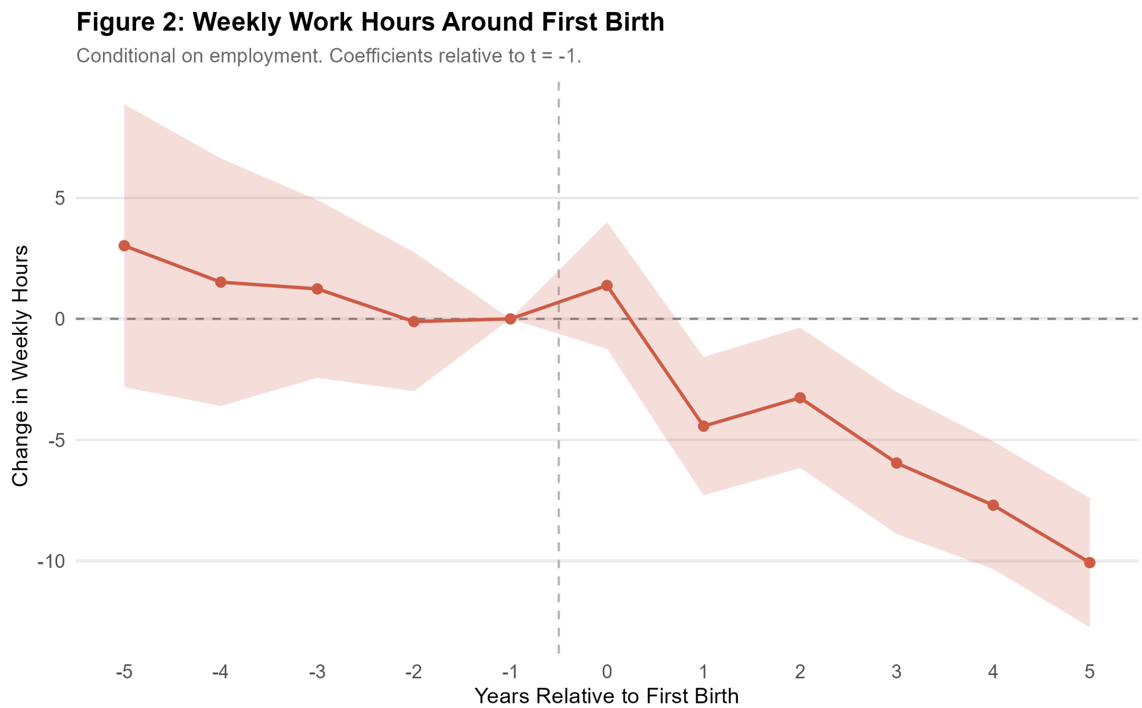

Figure 9: Intensive-margin event study for weekly hours among employed mothers only (year fixed effects; robust standard errors). Coefficients are hour changes relative to $t = -1$; by $t = +5$, weekly hours are lower by 10.1 from a baseline mean of 34.1. Bars show 95% confidence intervals.

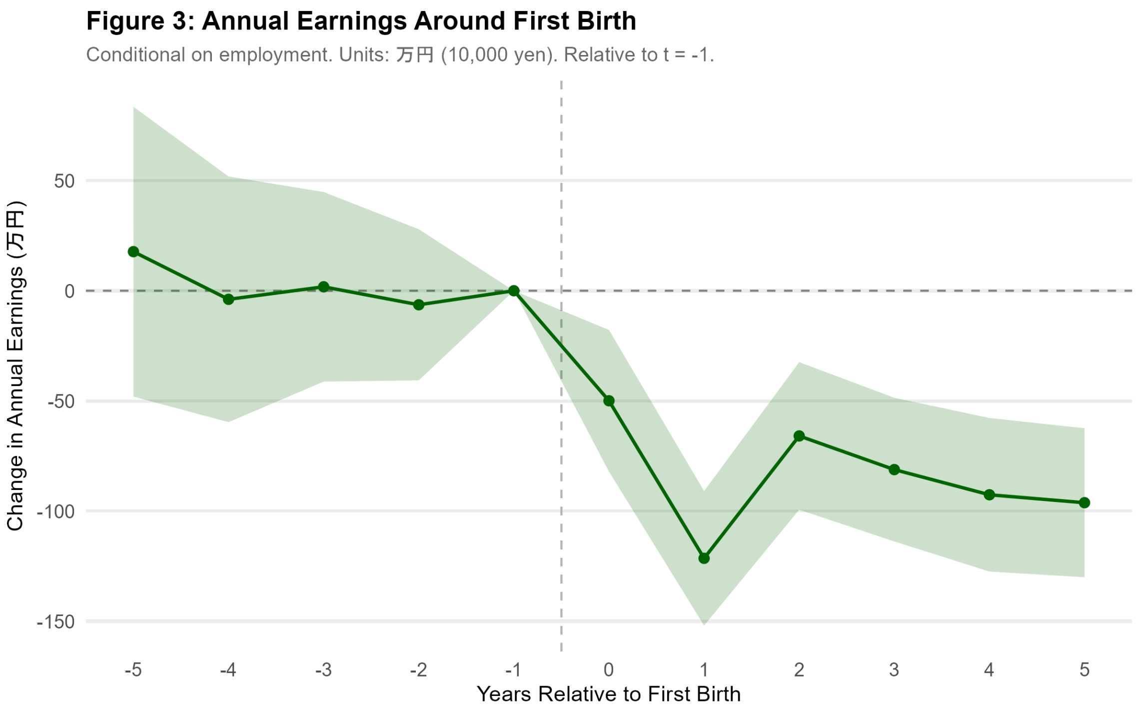

Figure 10: Intensive-margin event study for annual earnings among employed mothers only (year fixed effects; robust standard errors). Coefficients are changes in man-yen (10,000 yen units) relative to $t = -1$; by $t = +5$, earnings are lower by 96.2 man-yen from a baseline mean of 268.9. Bars show 95% confidence intervals.

The shift from full-time to part-time is reflected in the intensive-margin outcomes. Conditional on being employed, the event-study coefficient for weekly hours at $t = +5$ is -10.1 hours ($p < 0.001$), a 30 percent decline relative to the baseline mean of 34.1. In annual terms, this corresponds to roughly 525 fewer hours of paid work per employed mother per year. The decline is sharpest between $t = -1$ and $t = +1$ (coefficient -4.4 hours) but continues through $t = +5$, which indicates ongoing adjustment rather than a one-time shock.

Conditional earnings follow a parallel trajectory. The event-study coefficient at $t = 0$ is -49.9 man-yen, deepening to -121.4 at $t = +1$ and remaining at -96.2 at $t = +5$ ($p < 0.001$ at all horizons). Relative to the baseline mean of 268.9 man-yen, the $t = +5$ coefficient represents a 36 percent decline, or about 962,000 yen in foregone annual earnings per employed mother. The earnings loss is larger in proportional terms than the hours loss (36 versus 30 percent), which is consistent with post-birth employment shifting toward lower-paying and lower-quality jobs rather than simply fewer hours at the same wage schedule.

To assess whether these intensive-margin patterns are only a by-product of conditioning on employment, I re-estimate hours and earnings on the full sample by assigning zero outcomes to non- employed observations. The post-birth declines remain large and statistically strong. Relative to $t = -1$, unconditional hours are lower by 3.9 hours at $t = 0$, 6.1 hours at $t = +1$, and 6.0 hours at $t = +5$ (all $p < 0.001$ with woman-clustered SEs). Unconditional earnings are lower by 55.9 man-yen at $t = 0$, 86.8 man-yen at $t = +1$, and 52.2 man-yen at $t = +5$ (all $p < 0.001$). The near lead at $t = -2$ remains close to zero for both outcomes. This supports the interpretation that the post-birth deterioration is not solely an artefact of conditioning on employment (Appendix D.7, Table 29).

Directionally, the most plausible selection in this setting is positive selection among post-birth employed women: those with stronger attachment and better pre-birth jobs are more likely to remain employed. If so, the conditional intensive-margin declines likely understate the full-sample deterioration associated with childbirth. I therefore interpret the conditional coefficients as conservative in magnitude and use the unconditional outcomes as a complementary check.

Table 9 reports the earnings event-study coefficients alongside the number of women observed and employed at each event time, which makes the changing sample composition transparent.

Table 9: Annual earnings event study with sample composition

t	N	Employed	Emp.%	Mean earn.	Coef.	SE	p
-5	186	42	22.6	254.7	4.5	33.6	0.894
-4	241	64	26.6	249.5	-9.6	28.2	0.735
-3	316	113	35.8	262.7	4.6	21.9	0.835
-2	453	219	48.3	256.8	-4.2	17.5	0.812
-1	656	330	50.3	264.7	0.0	–	–
0	662	255	38.5	217.8	-54.0	16.3	< 0.001
1	654	265	40.5	140.0	-138.9	15.3	< 0.001
2	594	265	44.6	210.6	-68.9	16.9	< 0.001
3	509	256	50.3	196.8	-81.8	16.5	< 0.001
4	487	255	52.4	191.6	-92.5	17.6	< 0.001
5	454	242	53.3	190.6	-94.8	17.1	< 0.001

Notes: Earnings are in man-yen (10,000 yen). Coefficients come from the year-FE event study, conditional on employment and relative to $t = -1$, with heteroskedasticity-robust standard errors. N is the total number of women observed at each event time; Employed is the number with positive hours; Mean earn. is mean annual earnings among employed women.

The absence of recovery is important. In countries where the child penalty is driven primarily by career interruptions, earnings often recover as women rebuild human capital over time (Kleven et al., 2019b; Goldin, 2014; Bertrand et al., 2010; Blau and Kahn, 2017). In Japan, the lack of recovery suggests that the interruption is not merely temporary. Women appear to re-enter a different tier of the labour market rather than returning to their previous career tracks.

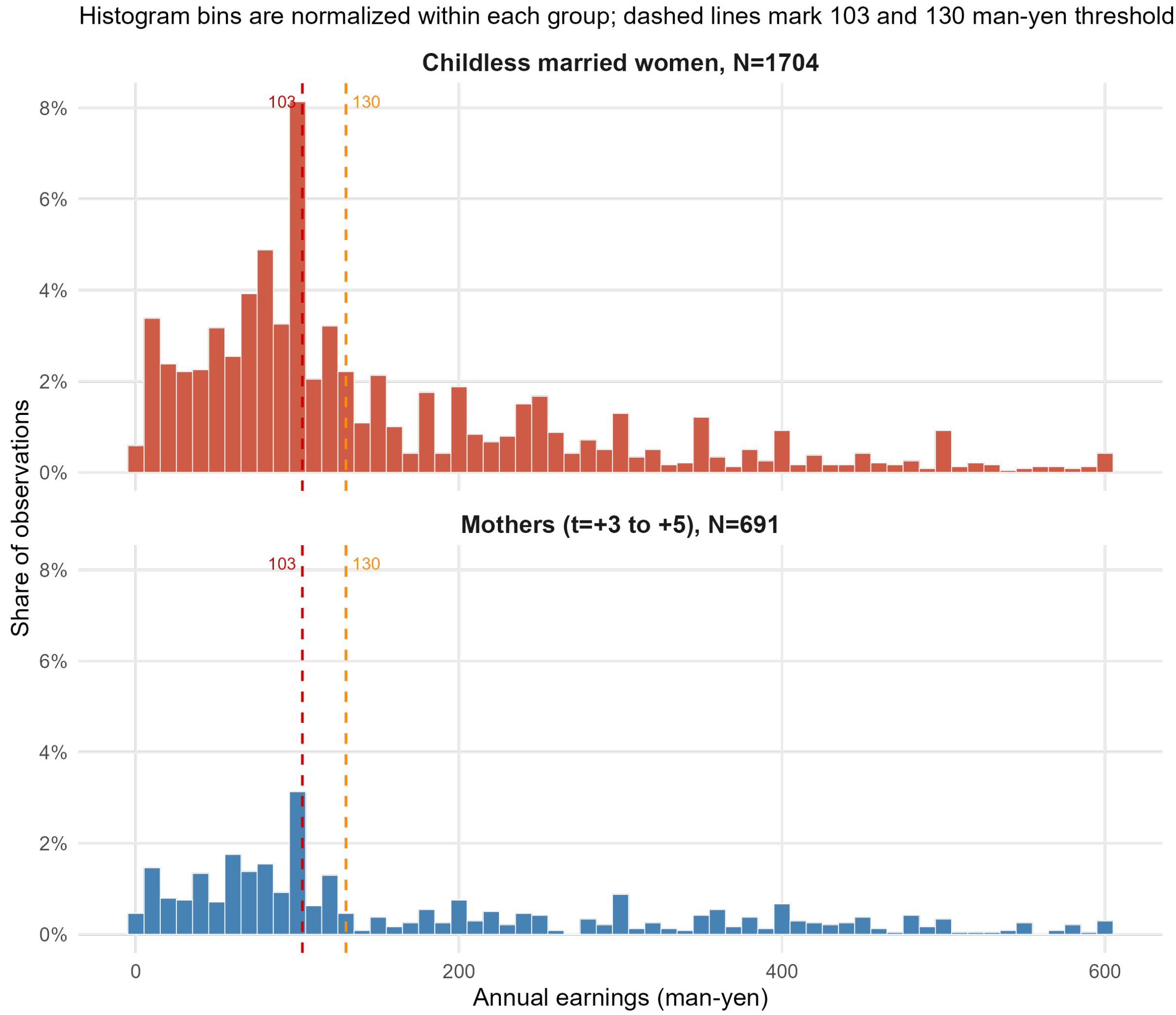

Figure 11: Normalized earnings distributions for employed mothers at $t = +3$ to $t = +5$ and employed childless married women. Dashed lines mark the 103 and 130 man-yen thresholds. The concentration near these cutoffs appears in both groups, which suggests that threshold bunching is not unique to mothers and likely reflects a broader secondary-earner pattern among married women.

Figure 11 compares normalized earnings distributions for employed mothers in the post-birth return period ($t = +3$ to $t = +5$) and employed childless married women. In the pooled mothers' sample, 47.8% of observations lie at or below 103 man-yen and 55.9% at or below 130 man-yen (median 115 man-yen). Among childless married employed women, the corresponding shares are very similar or slightly higher (49.7% and 60.0%, median 105 man-yen). I interpret this comparison cautiously: the childless group is a descriptive benchmark, not a causal counterfactual. Still, the similarity of the two distributions suggests that concentration around these thresholds is not unique to mothers and likely reflects a broader structural feature of earnings choices among married secondary earners. Motherhood appears to increase exposure to reduced-hours jobs in which these pre-existing thresholds are more likely to bind.

6.4 The shifting composition of the penalty

The preceding results - the collapse of full-time employment, the persistent hours decline, and earnings losses that exceed the hours losses - reveal a penalty that changes form over time. At $t = +1$, the penalty is primarily extensive: the event-study employment coefficient is -0.124, which means that much of the overall hours loss reflects women who are not working at all. By $t = +5$, the employment coefficient has narrowed to -0.028 and is statistically indistinguishable from zero, but the hours coefficient has deepened to -10.1 and the earnings coefficient remains at -96.2 man-yen. The penalty is still present, but it has shifted from the extensive margin to the intensive margin. The apparent recovery in employment should therefore not be read as full labour-market recovery, especially because the long horizon is panel-sensitive while the intensive margin remains persistently negative.

Figure 13 makes this shift visible using an accounting decomposition of average hours. Writing mean hours as the product of the employment rate and mean hours conditional on employment, the figure separates the change relative to $t = -1$ into an extensive component (employment decline), an intensive component (lower hours among employed women), and an interaction term. At $t = +1$, when the employment penalty is deepest, the extensive component explains most of the hours loss. By $t = +5$, the intensive component dominates: more of the remaining hours loss comes from women who are working fewer hours than before birth.

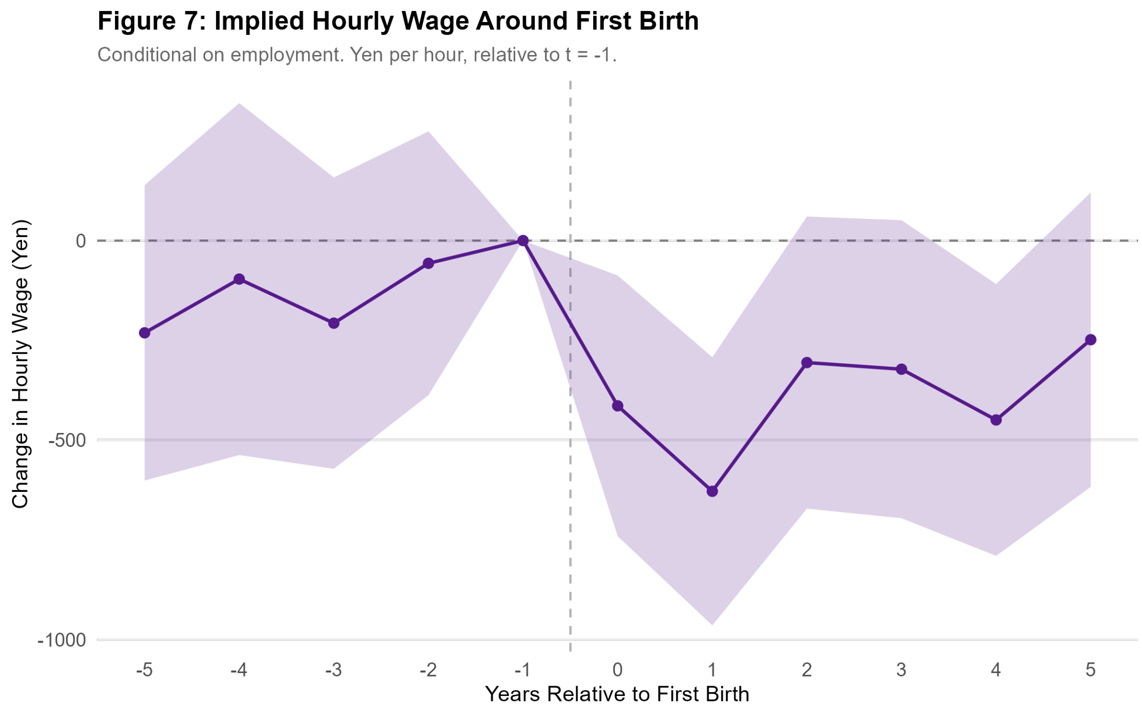

Figure 12: Implied hourly wage among employed mothers, computed as annual earnings divided by annualized hours (weekly hours $\times 52$). Mean implied wage declines from 1,857 yen/hour at $t = -1$ to 1,726 yen/hour at $t = +5$ (-7.1%), which is consistent with lower-quality post-birth job matches.

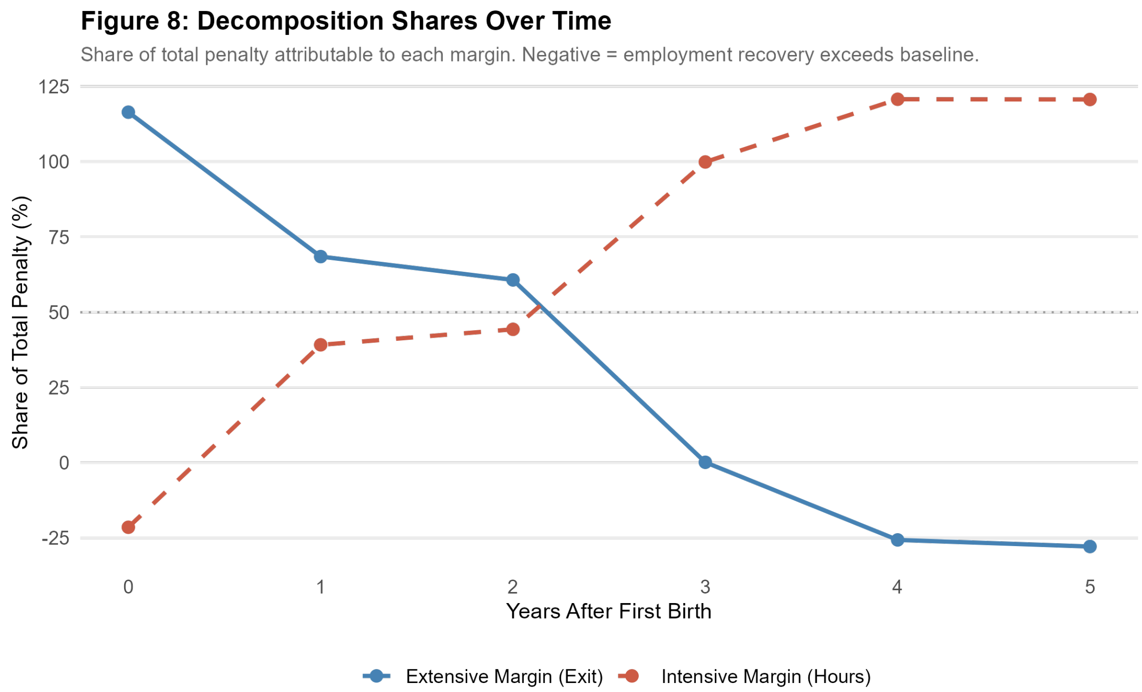

Figure 13: Accounting decomposition of the change in average hours into extensive (employment) and intensive (hours within employment) margins by horizon. Shares are computed from the identity linking mean hours to the employment rate and conditional hours, and can fall outside $[0, 100]$ when one component offsets the other. The extensive margin dominates immediately after birth, while the intensive margin dominates later horizons.

The decomposition and transition evidence are the core visual diagnostics in this section: they show that the penalty moves from immediate exit to persistent job-quality downgrade. The implied hourly wage in Figure 12 provides an additional descriptive indicator of post-birth job quality. The mean implied wage declines from 1,857 yen per hour at $t = -1$ to 1,726 yen per hour at $t = +5$, a drop of 7.1 percent. This estimate is approximate because it divides annual earnings by annualized hours, and both variables are measured with error; in a ratio measure of this kind, the direction of any resulting bias depends on the underlying error structure and cannot be signed without additional assumptions. The decline should therefore be read as suggestive rather than definitive. Even so, its direction is consistent with the broader pattern in the data: post-birth employment appears to shift not only toward fewer hours, but also toward lower-paying and lower-quality job matches.

6.5 The return hazard

Among women who exit employment at or around birth, how quickly do they return? The cumulative return hazard provides a direct descriptive answer. Among women who are not working at $t = +1$, 30.6 percent have returned to employment by $t = +2$. After that, the pace of return slows: cumulative re-entry reaches 45.4 percent by $t = +3$ and 56.5

percent by $t = +5$. In the observed data, no additional women return between $t = +4$ and $t = +5$, which suggests that re-entry becomes less common once women have remained out of employment for several years.

When the sample is restricted to women with no second birth by the horizon, the pattern is nearly identical: cumulative return reaches 55.9 percent by $t = +5$. This similarity suggests that the stalling at longer horizons is not explained solely by subsequent fertility and is consistent with frictions in re-entry. After several years out of employment, the remaining non-returners may face persistent barriers, including loss of firm-specific capital, weaker labour-market networks, and stigma associated with long employment gaps.

6.6 Workplace systems and the limits of accommodation

The data include a module on workplace systems at the employer level. Among employed mothers in the sample, 60.3 percent report that their employer offers reduced working hours, and 65.2 percent report half-day or hourly leave options. These figures indicate that some degree of flexibility is nominally available. However, only 10.1 percent report work-from-home availability, and 49.1 percent report rehiring after childcare leave. A further 62.3 percent report that their employer offers a pathway from non-regular to regular employment.

These provisions should not be interpreted as evidence that effective accommodation is widely available in practice. The transition evidence shows that relatively few women move from part-time hours back to full-time work, even when they remain attached to the labour market. More broadly, the results in this chapter suggest a gap between formal availability and realised continuity. Workplace systems may exist on paper, but their practical value depends on contract status, employer capacity, and whether the employment relationship survives through childbirth.

6.7 The non-regular re-entry trap

Taken together, the findings in this section are consistent with a self-reinforcing re-entry trap. Women in non-regular employment are the most likely to exit around motherhood, and post-birth employment shifts strongly toward reduced-hours work among those who remain employed or return. Among employed mothers, hours, earnings, and implied hourly wages remain below their pre-birth levels throughout the observed window. At the same time, transitions back to full-time work appear limited. The overall pattern is therefore consistent with a durable shift toward lower-quality post-birth employment in a segmented labour market.

Because this design is not an individual-level causal re-entry model, I do not claim that every exiting woman re-enters through the same pathway. The claim is narrower:

in the aggregate, post- birth re-entry appears to occur disproportionately through lower-quality positions, and these positions do not seem to function as temporary bridges back to pre-birth employment trajectories.

This interpretation connects the findings to Goldin (2014)'s argument that the "last chapter" of gender convergence in earnings depends on the structure of jobs rather than the characteristics of workers. In Japan, the relevant structure is the institutional divide between regular and non-regular employment rather than occupational temporal flexibility alone (Goldin, 2014). The mechanism differs, but the implication is similar: closing the gender gap requires changes in job structure rather than in worker characteristics.

6.8 Skills training and the Training Grants system

The employment penalty documented above may understate the full cost of childbirth-related labour-market withdrawal if mothers also reduce investment in job-related skills. Career interruptions can depreciate accumulated skills and generate wage losses that persist beyond the interruption itself (Mincer and Ofek, 1982; Mincer and Polachek, 1974). In Japan, where non-regular re-entry positions offer limited employer-provided training (OECD, 2021, 2022), self-directed skill investment may be an important margin for maintaining employability during and after a career break.

Both the KHPS and JHPS ask respondents and spouses whether they have taken voluntary actions - such as attending school, lectures, or self-study - to improve work-related skills, and whether they have used the government's Training and Education Benefits System (*kyōiku kunren kyūfu seido*). This system, administered through Hello Work (*haroōwaku*) offices, subsidises 20-70 percent of approved training-course costs for workers enrolled in employment insurance (Ministry of Health, Labour and Welfare, 2024). A 2018 reform extended the eligibility window from 4 to 20 years for individuals who left employment due to pregnancy, childbirth, or childcare.

Table 10 reports the share of mothers reporting any skills-training activity by event time. Before birth, about one-third of mothers report active skill investment in a given year (31–39 percent across $t = -5$ to $t = -1$). This rate drops sharply to 8.6 percent in the birth-report wave ($t = 0$) and recovers only partially, reaching 13–16 percent by $t = +3$ through $t = +5$. The pattern broadly tracks the employment penalty and is consistent with skill investment remaining tied to labour-market attachment rather than functioning as a bridge back to work.

Use of the Training and Education Benefits System is negligible in this sample: at each event time, fewer than 3 percent of mothers with valid responses report current or recent use. The more informative pattern lies in why usage is so low. Table 11 decomposes awareness and self-reported eligibility at selected event times. Roughly half of mothers have never heard of the programme, and this share changes little around childbirth (47 percent at $t = -2$, 53 percent at $t = 0$, and 45 percent at $t = +5$). Among aware non-users, most report being either ineligible or uncertain about their eligibility. Before birth ($t = -1$), 24 percent of aware non-users report that they are eligible, 38 percent report that they are not, and 38 percent do not know. By $t = +5$, the ineligible share rises to 46 percent and the eligible share falls to 19 percent, which is consistent with the loss of employment-insurance connection after labour-market exit. The 2018 reform extended the eligibility window to 20 years for mothers who left employment due to pregnancy or childcare (Ministry of Health, Labour and Welfare, 2024), but most observations in this sample predate that reform, and whether it has improved access in practice remains an open question.

The pattern therefore points to a three-layer access barrier: invisibility (roughly half of mothers are unaware of the programme), ineligibility (38–46 percent of aware non-users report not qualifying), and uncertainty (roughly 35–40 percent do not know whether they qualify). This interpretation is consistent with OECD assessments that Japan's adult-learning system relies heavily on employer-provided training, offers limited flexibility, and underserves non-regular workers and career returners (OECD, 2021).

These patterns imply a double penalty: childbirth reduces both labour supply and investment in job-related skills. The persistent training gap may contribute to the durable intensive-margin losses documented above if returning mothers re-enter employment without maintaining or upgrading their skills during the interruption (Mincer and Ofek, 1982; Adda et al., 2017). The near-zero use of the training-grants system, despite its explicit design as a maternal reskilling channel, suggests that formal eligibility alone is insufficient when time constraints, information gaps, loss of employment-insurance connection, and weak rewards to additional qualifications in non-regular jobs remain in place (Asai, 2015; OECD, 2022).

Table 10: Skills Training Participation and Training Grants Awareness by Event Time

Event time	Q2.1: Skills Training (%) N_{valid}	Any (1+2)	Current (1)	Q2.4: Training Grants (%) N_{valid}	Used (1+2)	Aware (3)
−5	45	26.7	15.6	37	2.7	54.1
−4	70	28.6	8.6	63	0.0	46.0
−3	134	38.8	13.4	107	2.8	52.3
−2	253	35.2	14.2	185	2.2	51.4
−1	475	31.2	13.1	309	2.3	50.8
0	568	8.6	2.6	391	1.0	45.8
+1	594	10.6	4.4	387	1.3	49.6
+2	580	13.1	5.3	363	0.6	49.6
+3	493	15.6	4.3	292	0.3	53.4
+4	472	13.6	6.1	234	1.3	52.1
+5	437	12.8	6.2	177	0.0	54.8

Notes: The sample is the main analytic panel of 662 mothers around first birth. Q2.1 asks whether the wife took voluntary actions to improve job-related skills in the past year (1 = presently taking action; 2 = took action; 3 = did not). "Any" reports codes 1+2 as a share of valid responses (codes 1–3). Q2.4 asks about use of the Training and Education Benefits System (*kyōiku kunren kyūfu seido*): 1 = used and received benefits; 2 = using and will receive; 3 = aware but not used; 4 = unaware. "Used" reports codes 1+2, and "Aware" reports code 3; both are calculated as shares of valid responses (codes 1–4). Coverage varies by event time because of unbalanced panel support and item non-response. The 2018 reform extended eligibility to 20 years for individuals who left employment because of pregnancy, childbirth, or childcare (Ministry of Health, Labour and Welfare, 2024).

Table 11: Training Grants Access Barriers: Awareness, Eligibility, and Usage by Event Time

	All respondents (%)			Among aware non-users (%)		
Event time	Used	Aware, not used	Unaware	Eligible	Not eligible	Don't know
−2	2.2	51.4	46.5	–	–	–
−1	2.3	50.8	46.9	24.2	37.6	37.6
0	1.0	45.8	53.2	26.8	38.5	34.6
+1	1.3	49.6	49.1	19.8	35.4	44.3
+3	0.3	53.4	46.2	17.9	42.3	39.1
+5	0.0	54.8	45.2	18.6	46.4	35.1

Notes: The sample is the main analytic panel of 662 mothers around first birth. The left panel reports shares of all valid Q2.4 responses: "Used" = codes 1+2 (used and received or are receiving benefits); "Aware, not used" = code 3; "Unaware" = code 4. The right panel conditions on aware non-users (code 3) and reports self-assessed eligibility: "Eligible" = reports qualifying; "Not eligible" = reports not qualifying; "Don't know" = uncertain about eligibility. Eligibility for the Training and Education Benefits System (*kyōiku kunren kyūfu seido*) requires enrolment in employment insurance (*koyō hoken*). Women who exit employment at childbirth lose this connection, which is consistent with the rising "Not eligible" share at later event times. The 2018 reform extended the eligibility window to 20 years for individuals who left employment because of pregnancy, childbirth, or childcare (Ministry of Health, Labour and Welfare, 2024). The eligibility breakdown at $t = -2$ is omitted because of small cell sizes. Coverage varies by event time because of unbalanced panel support and item non-response.

7 Limited paternal adjustment

Parenthood creates a household shock with two potential margins of adjustment. Either partner can absorb childcare time through lower market work, and the realised allocation reflects wages, job security, workplace flexibility, and social norms within the household. Evidence from Scandinavian settings shows modest but measurable paternal adjustment, and stronger paternal adjustment is associated with smaller maternal penalties (Kleven et al., 2019b; Andresen and Nix, 2022). In Japan, I find no detectable systematic paternal adjustment in this sample window, so post-birth labour-supply responses remain concentrated among mothers.

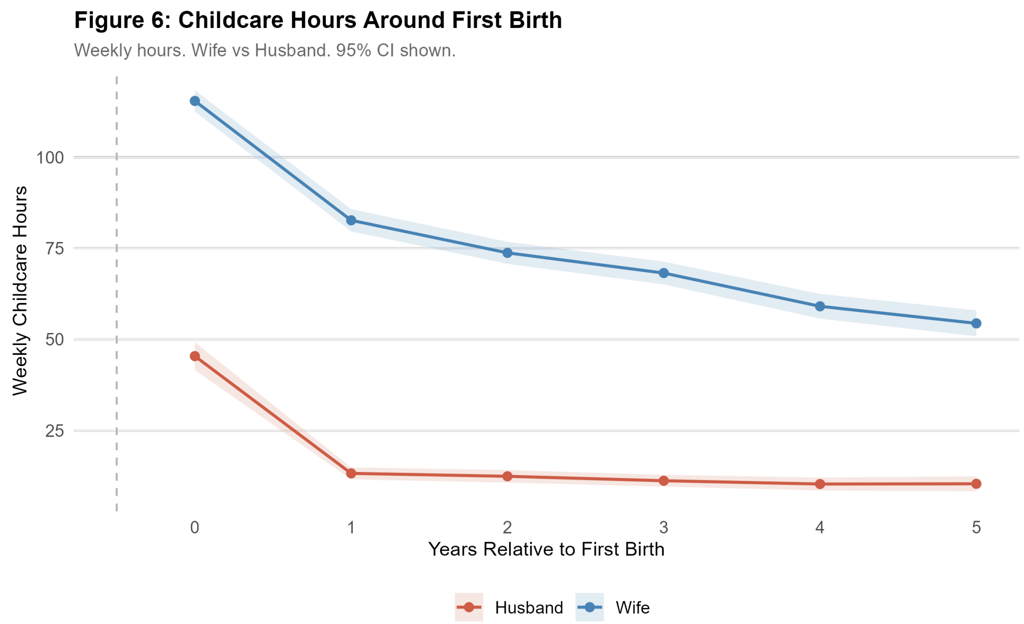

Figure 14: Mean weekly childcare hours for wives and husbands around first birth. Series are restricted to $t \geq 0$ because pre-birth survey timing can overlap pregnancy and early postpartum reporting. At $t = +1$, the wife-to-husband ratio is 6.3-to-1 (82.7 vs. 13.2 hours/week); by $t = +5$, it remains 5.3-to-1 (54.4 vs. 10.4).

7.1 Childcare time asymmetry

Interpretation of pre-birth childcare responses requires care because of survey timing. Births recorded in wave t can occur between February of the previous year and January of the survey year, so some observations coded as $t = -1$ already overlap early childcare. Empirically, 43% of women at $t = -1$ report daily childcare. I therefore interpret childcare time primarily for $t \geq 0$, where event-time classification is cleaner.

From $t = 0$ onward, the asymmetry is large and persistent. At $t = 0$, mothers report 115.5 weekly childcare hours versus 45.4 for fathers (about 2.5-to-1). By $t = +1$, the ratio widens to 6.3-to-1 (82.7 vs. 13.2), and by $t = +5$ it remains 5.3-to-1 (54.4 vs. 10.4). The key point for labour-market interpretation is persistence rather than a one-off birth-year shock: the caregiving burden remains heavily concentrated on mothers throughout the observed post-birth window.

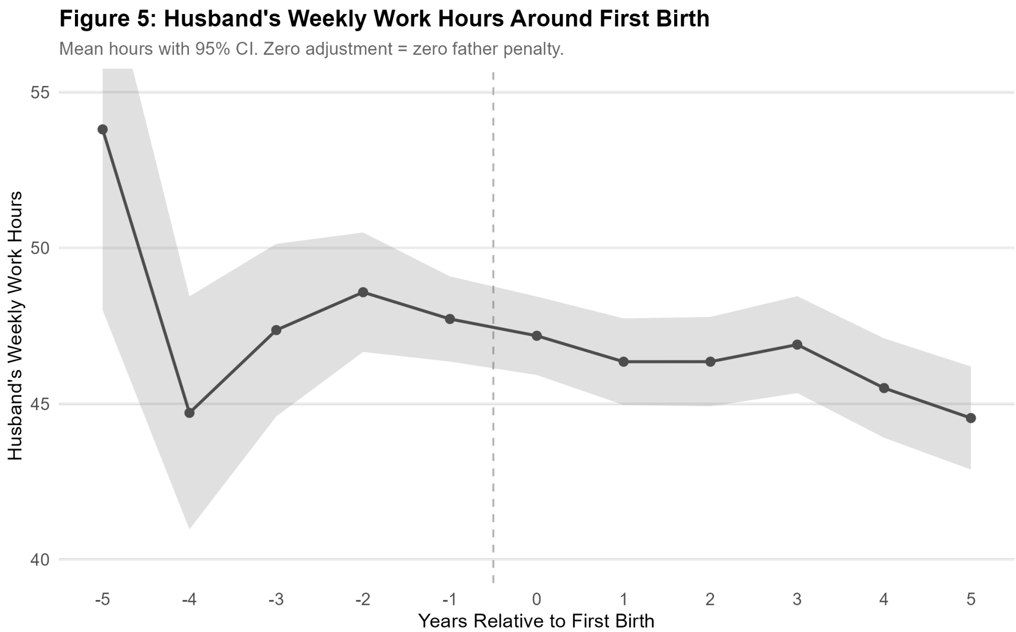

Figure 15: Father-hours benchmark around first birth. Displayed points are mean husband weekly hours with 95% confidence intervals by event time. The profile is broadly flat, indicating no detectable systematic paternal labour-supply adjustment around childbirth in this sample.

7.2 Leave take-up

In the broader leave-analysis sample (N = 1,183), only 3 husbands (0.3%) report taking childcare leave, compared with 250 wives (21.1%). Paternity leave, while formally available since 2010 and expanded in subsequent reforms, was taken by a vanishingly small number of fathers during the observation period. National statistics indicate that uptake has increased in recent years, but average leave duration among fathers remains short (OECD, 2024a; Ministry of Health, Labour and Welfare, 2019). Recent Japan-specific evidence also suggests that higher paternal leave norms are associated with better maternal employment continuity (Nakayama and Ishikawa, 2025). This broader trend occurs largely after my sample window and does not alter the core within-sample result of near-zero paternal adjustment.

7.3 Interpretation

The near-zero father-hours profile in this sample is consistent with the institutional context described in Section 2. Japanese corporate culture places strong demands on male employees' time and availability, which leaves limited room for paternal caregiving. The result is a household in which the arrival of a child creates an asymmetric shock. The mother faces much stronger pressure to absorb the childcare adjustment, and, as the earlier sections show, the extent of that adjustment depends heavily on the structure of her job. In these data, fathers do not show comparable reductions in market work around childbirth.

This finding helps interpret the earlier results. The pattern is consistent with severe temporal constraints that limit the scope for within-household reallocation of caregiving. When husbands work very long hours, the family's ability to redistribute childcare is correspondingly reduced, and more of the burden falls on the mother in this sample. In the regression models, husbands' overwork is imprecisely estimated and not statistically significant once labour-market tier is included. I therefore treat the father-side evidence as descriptive context rather than as a standalone mechanism estimate.

8 Robustness

I examine robustness and scope along several dimensions: a childless comparison group, support and pre-trend diagnostics, second-birth dynamics, sensitivity of the intensive-margin interpretation, and small-sample inference for the mechanism regressions.

Table 12: Evidence-to-claim map: what each finding identifies and what it does not.

Claim	**Evidence**	**What it identifies**	**What it does not identify**	**Level**
Birth-year employment break	$t = 0$ coeff. $= -0.132$ ($p < 0.001$); near lead null; stable across cohort splits, survey sources, IPW, balanced panel, childless placebo	Sharp timing of employment decline at first birth in this panel	Causal effect of childbirth vs. anticipatory sorting; composition vs. within-person change	High
Recovery masks downgrade	FT share drops from 100% to 45.7% at $t = +1$; hours -10.1/wk; earnings -96.2 man-yen at $t = +5$; unconditional outcomes confirm	Persistent shift toward reduced-hours, lower-earnings re-entry	Whether downgrade is voluntary (preference) or involuntary (constrained); selection into return	High
Non-regular tier predicts exit	Childbirth margin: OR $= 7.50$ ($p < 0.001$), $N = 330$, 124 exits; absolute risk 13.8% vs. 64.8%; stable across missingness specs	Strong predictive gradient by contract type at $t = 0$	Whether non-regular status causes exit or proxies for unobserved worker characteristics	High
Pre-birth exit composition	$t = -2$ logit: OR $= 6.41$ ($p = 0.019$); $N_{\text{events}} = 13$; Firth/bootstrap confirm sign	Direction of pre-birth predictive association	Precise magnitude (wide CIs); causal sorting vs. selection	Moderate
Institutional interpretation	Earnings bunching at thresholds (47.8% mothers $\leq$103 man-yen; 49.7% childless married $\leq$103 man-yen) - suggests a broader secondary-earner phenomenon; training grants $<$3% usage with three-layer access barrier (invisibility, ineligibility, uncertainty); father hours flat; leave take-up 0.3%	Descriptive facts; threshold bunching is a pre-existing constraint on married women's earnings; training grant barriers are structurally linked to employment exit	Whether segmentation causes the penalty or merely correlates with it	Interpretive

Notes: "High" = robust across major sensitivity checks. "Moderate" = direction is stable but precision is limited by small event counts. "Interpretive" = institutionally grounded but not separately identified by the panel design. The claim hierarchy mirrors the ordering in the introduction.

8.1 Childless Comparison Group

A within-sample childless comparison is infeasible because the harmonized classified panels are constructed from mothers only. I therefore build a descriptive benchmark from the raw panels: married women aged 20–45 with no observed births and at least five years of panel observation. This yields N = 4,111 unique women and 50,222 person-year observations. Their average employment rate is 53.2 percent - close to the mothers' rate of 50.3 percent at $t = -1$ - with mean weekly hours of 28.5 among workers. However, the composition differs considerably: 71.9 percent of employed childless women hold non-regular positions, and 59.9 percent work at small firms. Employment rates are stable in the 0.51–0.57 range across the 2004–2022 period, with no visible trend.

This benchmark serves two purposes. First, it confirms that the employment decline observed among mothers at $t = 0$ is specific to the birth event rather than to a cohort- or period-level trend: childless married women show no comparable drop. Second, the higher non-regular share among childless women (71.9 versus approximately 38 percent among employed mothers at $t = -1$) suggests that selection into motherhood is not random with respect to job quality: women with regular employment are more likely to appear in the mother sample, presumably because stable employment supports family formation. If so, the analytic mother sample may be positively selected on job quality relative to the broader population of married women.

I emphasize that this comparison is descriptive rather than causal. The childless women differ from mothers in age, cohort, and unobserved characteristics that preclude a difference-in-differences interpretation. See Appendix A for full details.

Figure 16 summarizes this support profile directly and motivates the emphasis on the near-birth window in interpretation.

8.2 Attrition and support composition

Because the panel is unbalanced, compositional change across event time is a first-order concern. I therefore report two diagnostics in Appendix D: (i) support counts by event time, and (ii) baseline ($t = -1$) differences between women observed through $t = +5$ and those not observed through $t = +5$. Support thins substantially in the tails (from 186 women at $t = -5$ to 662 at $t = 0$, then 454 at $t = +5$). Observation at $t = +5$ is also weakly selective on baseline employment: women observed at $t = +5$ have lower $t = -1$ employment (47.7%) than those not observed (56.2%, $p = 0.045$), while baseline hours and income differences are not statistically significant. These diagnostics do not identify causal attrition effects, but they bound long-horizon interpretation and motivate caution for $t = +5$ coefficients.

I also re-estimate the employment event study on a balanced panel restricted to women observed continuously from $t = -2$ to $t = +1$ (N = 662 women; 2,648 observations). In

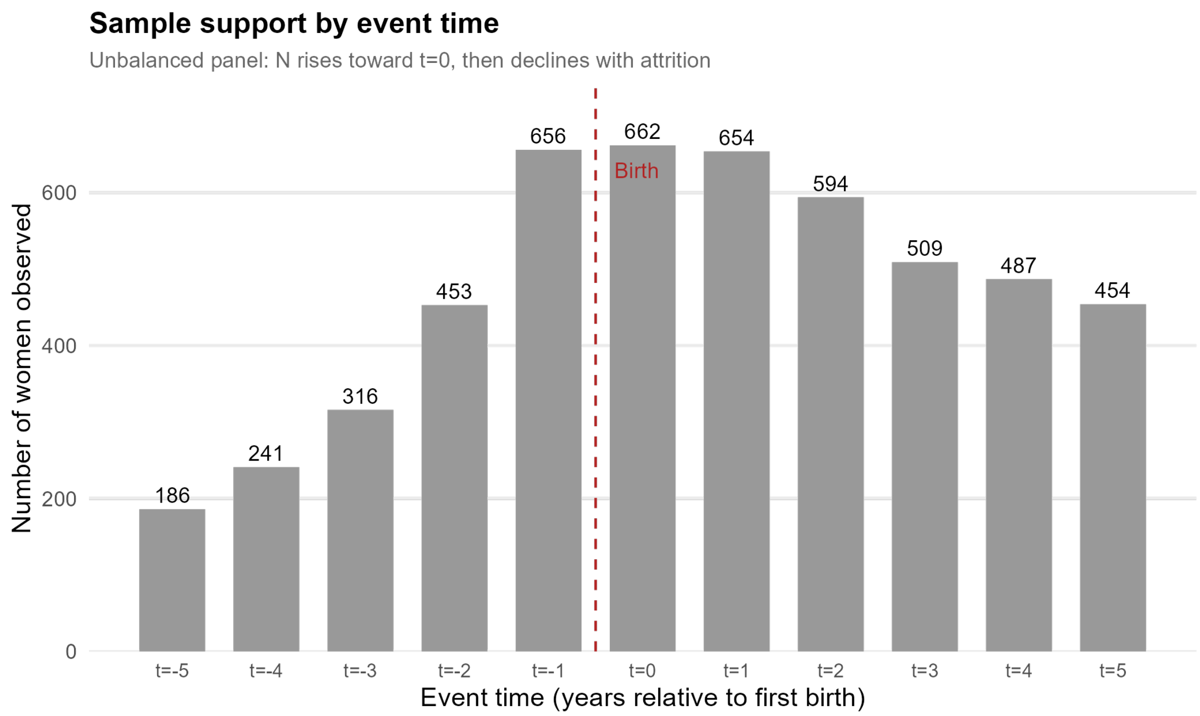

Figure 16: Sample support by event time in the unbalanced panel. The number of observed women rises toward birth and declines at longer horizons, so far leads and lags should be interpreted with greater caution.

this sample, the key coefficients are unchanged: $t = -2$ is -0.0085 ($p = 0.781$), $t = 0$ is -0.1313 ($p < 0.001$), and $t = +1$ is -0.1292 ($p < 0.001$). This check supports the interpretation that the birth-year break is not driven by short-horizon support imbalance (Appendix D.6, Table 27).

Additional cleaner-sample checks reach the same conclusion. In an age-restricted sample (age ≥ 23), the childbirth-year drop remains large ($t = 0 = -0.123$, $p < 0.001$) and the near lead remains close to zero ($t = -2 = -0.007$, $p = 0.836$). In the balanced short-window sample ($t = -2$ to $t = +1$), the break is also preserved ($t = 0 = -0.128$, $p < 0.001$). By contrast, stricter attachment-based restrictions produce larger magnitudes, consistent with selection into higher-attachment subgroups rather than a cleaner neutral baseline estimand (Appendix D.6, Table 28).

8.3 Pre-trend diagnostics

The key pre-trend diagnostic in this design is whether employment is stable close to birth. In the unbalanced panel, the far leads ($t = -5, -4, -3$) show large negative coefficients that are jointly significant ($p < 0.001$). I do not interpret this pattern as evidence of a conventional pre-trend. Rather, it reflects two features of the sample: first, women enter the panel at different event times, so the far-lead estimates are based on a selected subsample observable early; and second, employment rises over the life cycle as women in their twenties enter the labour market.

The critical diagnostic is therefore the near lead at $t = -2$, which asks whether employment is stable in the year immediately before birth. In the baseline specification, the coefficient is -0.012 ($p = 0.70$, N = 5,212). I conduct three robustness checks. In a trimmed pre-period specification that drops $t = -5, -4, -3$, the $t = -2$ coefficient remains null ($p = 0.72$, N = 4,469). In a balanced pre-support sample restricted to women observed at every pre-birth event time, the $t = -2$ coefficient remains economically small and only marginally significant ($p = 0.078$, N = 1,814), which is consistent with reduced precision under the stricter support restriction. Adding dataset fixed effects with a JHPS/KHPS indicator leaves the near lead unchanged ($p = 0.71$, N = 5,212).

Across all four specifications, the near lead remains economically close to zero and is statistically null in three of the four cases. Together, these checks support the interpretation that employment is stable in the year immediately before first birth. The sharp break at $t = 0$ (coefficient -0.132, $p < 0.001$) against this flat near lead supports the view that the birth-year decline is not simply a continuation of the earlier trajectory.

8.4 Implied hourly wage

As discussed in Section 6, the implied hourly wage declines from 1,857 yen per hour at $t = -1$ to 1,726 yen per hour at $t = +5$, a drop of 7.1 percent. This measure is computed as annual earnings divided by annualized hours (weekly hours $\times$ 52) among employed women. The decline is consistent with post-birth employment shifting toward lower-paying job matches rather than reflecting fewer hours alone. At the same time, this measure should be treated as approximate: measurement error in both earnings and hours introduces noise, and the direction of bias in a ratio estimator cannot be signed without further assumptions about the error structure. If labour-force selection is positive on earning potential, the observed conditional decline may understate the deterioration associated with childbirth in the full sample.

8.5 Second-birth dynamics

Birth spacing in the sample is relatively short. Among the 662 mothers, 255 (38.5 percent) have a recorded second birth within the panel window. The share with a second birth rises to 5.0 percent by $t = +1$, 19.2 percent by $t = +2$, 33.8 percent by $t = +3$, and 44.1 percent by $t = +5$. The 44.1% value is conditional on women observed at $t = +5$ (N = 454), whereas 38.5% is the full analytic-panel share (255/662); these are different conditioning sets. By $t = +3$, one-third of the observed sample has experienced a second birth within the panel window, which complicates the interpretation of medium-run coefficients after first birth.

I address this in two ways. First, the specification that excludes post-second-birth observations yields a $t = +5$ employment coefficient of -0.003 (N = 4,495), compared with -0.028 in the full sample (N = 5,212). This attenuation suggests that part of the residual employment gap at $t = +5$ reflects subsequent fertility.

Second, a stratified event study separates women by whether they eventually have a second birth within the panel window. Women without a second birth have a $t = 0$ employment coefficient of -0.124, recover to $+0.002$ by $t = +3$, and reach $+0.013$ by $t = +5$, which is consistent with recovery to pre-birth employment levels within three years. Women with a second birth show a deeper initial decline at $t = 0$ (-0.151), partial recovery by $t = +3$ (-0.072), and a persistent deficit at $t = +5$ (-0.068), with no full recovery within five years.

The gap between the two groups is present even before birth ($t = -2$: $+0.042$ for no second birth versus -0.089 for second birth), which indicates that women who go on to have a second child are already on a different employment trajectory at baseline. The stratification therefore shows that the medium-run penalty in the pooled sample combines near-full recovery for one-child mothers with more persistent depression among those who have additional children.

8.6 Small-sample inference

The mechanism regressions in Section 5 are estimated on 13 exit events, so the results should be interpreted as supporting evidence rather than as a central result. I report three layers of inference to assess sensitivity to this constraint.

Standard maximum likelihood yields odds ratios of 6.41 (non-regular, $p = 0.019$) and 4.14 (small firm, $p = 0.025$). Firth's penalized likelihood estimator (Firth, 1993), which reduces finite-sample bias in rare-event logit settings, produces attenuated but still statistically significant estimates: OR = 5.32 ($p = 0.009$) and OR = 3.84 ($p = 0.021$). In addition, bootstrap inference from 2,000 nonparametric resamples yields 95 percent confidence intervals that exclude 1 for both predictors: $[1.68, 19.66]$ ($p = 0.008$) for non-regular and $[1.10, 18.34]$ ($p = 0.038$) for small firm.

The consistency across methods - MLE, penalized ML, and bootstrap - suggests that the sign of the association is not being driven solely by small-sample bias. At the same time, the confidence intervals are wide, so the exact magnitudes remain uncertain. The most defensible conclusion is therefore directional: non-regular workers and small-firm workers are more likely to exit on the pre-birth margin in this sample.

I further test single-observation sensitivity using leave-one-out jackknife and influence diagnostics for the two-predictor mechanism model. The non-regular odds ratio remains above one in all 185 jackknife replications (OR range: 5.69–12.82; p range: 0.016–0.030). The small-firm odds ratio also remains above one in all replications (OR range: 3.54–5.52), with significance retained in 98.9 percent of runs (p range: 0.014–0.052). Influence statistics (Cook's distance, leverage, and DFBETAs) do not indicate that a single observation overturns the core pattern. Appendix D reports these diagnostics.

In plain terms, removing any single respondent does not materially change the mechanism result: the estimated association remains positive and of similar order of magnitude for both predictors.

See Appendix A for the childless benchmark details and placebo event study, and Appendix B for second-birth spacing and stratified event-study tables.

8.7 Summary of Findings

I documented four principal findings. First, the aggregate break in employment occurs at childbirth ($t = 0$), while the pre-birth exit margin is compositionally distinct: 41.4 percent of pre-birth leavers cite marriage, and at $t = 0$ 85.0 percent cite childcare. In this sample, the near lead at $t = -2$ is flat in net terms, so the pre-birth result is best interpreted as the composition of gross exits rather than as a second aggregate break.

Second, the dual labour market is the strongest observed predictor of exit. At the childbirth margin, non-regular employment status strongly predicts non-employment (OR = 7.50, $p < 0.001$), while small-firm effects are positive but imprecisely estimated in the same model. In the lower-power pre-birth model, both non-regular status (OR = 6.41, $p = 0.019$) and small-firm size (OR = 4.14, $p = 0.025$) are positively associated with exit. The household-level variables available in these data, such as the husband's hours and the wife's commute time, add little explanatory power once the woman's labour-market position is taken into account.

Third, the penalty transforms over time. Although aggregate employment recovers by $t = +5$, the composition shifts persistently: among women who were full-time at $t = -1$, only 42.2 percent remain full-time at $t = +5$. Conditional hours fall by about 10 hours and conditional earnings by about 96 man-yen, with no clear sign of recovery within five years. These post-birth patterns are consistent with re-entry through lower-quality job tracks, although this design does not separately identify individual re-entry paths for all exiters.

Fourth, I find no detectable systematic paternal adjustment in this sample window. Childcare-hour ratios stabilize near 5-to-1 by $t = +5$, and in the broader leave-analysis sample only 3 husbands (0.3%, N=1,183) report taking childcare leave.

9 Discussion and conclusion

9.1 Policy implications

The documented risk-stratification pattern points to several policy implications. These are implications consistent with the descriptive evidence, not estimates of the causal effects of any specific reform.

First, strengthen effective employment protections for non-regular workers. At the childbirth margin, the non-regular odds ratio is 7.50, the largest coefficient in the main risk- stratification results. Only 21.1 percent of mothers ever report taking childcare leave; for many women, the leave system is not operational because the employment relationship does not survive through childbirth. Japan's amendments to the Part-Time Workers Act and the Labour Contract Act are therefore directionally well targeted, especially where they protect continuity and limit contract non-renewal around childbirth (Ministry of Health, Labour and Welfare, 2019; OECD, 2024a; Yamaguchi, 2019).

Second, target implementation constraints in small firms. The firm-size gradient in exit rates reflects the practical difficulty of accommodating parental leave in organizations with limited staffing. In the childbirth-margin model, the conditional small-firm coefficient is positive but imprecise (OR = 1.55, $p = 0.126$), so this implication is grounded more in descriptive patterns and institutional constraints than in a precise conditional estimate. My workplace-systems data show that while 60.3 percent of mothers' employers nominally offer reduced-hours arrangements, only 10.1 percent offer work-from-home. Subsidies for replacement hiring during leave periods and simplified administrative procedures for small employers could help reduce these implementation frictions.

Third, reduce the overwork constraint on fathers. The near-zero father adjustment observed in this sample is consistent with a work culture that makes paternal caregiving structurally difficult. Greater scope for fathers to reduce hours or take leave may be an important complement to any policy aimed at improving maternal employment continuity.

Fourth, improve post-birth mobility from non-regular to regular tracks. High non-regular exit risk at childbirth and near-zero use of training grants - driven by a combination of unawareness, loss of employment-insurance eligibility after exit, and uncertainty about qualification - point to a bottleneck at re-entry. Policies that strengthen conversion pathways for qualified non- regular workers, improve information about eligibility, and tie re-entry support more closely to actual job placement may be more effective than formal training eligibility alone.

In interpreting this re-entry pattern, I do not treat institutional segmentation and fiscal thresholds as competing explanations. They are likely complementary. Threshold bunching is a broad secondary-earner phenomenon - childless married women bunch at least as much as returning mothers - so the thresholds constrain married women's earnings independently of parenthood. What motherhood appears to add is a transition into the part-time earnings range where these pre-existing constraints are more likely to bind. Reduced-hours re-entry can therefore reflect both demand-side constraints, such as limited access to regular-track jobs, and supply-side responses to tax and social-insurance thresholds. The policy implication is correspondingly joint: improve transition opportunities into regular employment while reducing sharp earnings cliffs for second earners.

9.2 External validity and broader implications

I consider the implications of these findings beyond the Japanese context with some caution. The core stratification pattern - that labour-market segmentation is associated with a larger motherhood penalty among workers with weaker contractual protections - is not unique to Japan. Dual labour markets exist throughout Southern Europe, South Korea, and parts of Latin America, wherever fixed-term contracts and firm-size-dependent regulations create tiers of employment protection (see Adda et al., 2017; Kleven et al., 2019a, 2024, for cross-country evidence on career costs). The specific institutional details differ - Japan's spousal tax thresholds, for example, have no direct analogue in Spain's temporary-contract system - but the structural logic is similar: when a large share of women hold jobs that cannot accommodate career interruption, leave expansions and childcare subsidies may underperform their potential.

These parallels are especially visible in other dual labour markets. In Spain, where approximately one quarter of the workforce holds fixed-term contracts, mothers become increasingly likely to hold temporary contracts after childbirth, and the child penalty widens from 11 percent in the first year to 28 percent after a decade (de Quinto et al., 2021). In Italy, matched employer-employee data reveal a 57 log-point earnings penalty driven partly by mothers sorting into lower-paying firms and shifting to part-time contracts - a pattern that is directionally consistent with the post-birth downgrading documented here (Casarico and Lattanzio, 2023). In Korea, the M-shaped female participation curve, the OECD's largest gender wage gap, and a non-regular share exceeding Japan's produce a similar configuration in which broad policy expansion coexists with persistent motherhood-related career discontinuity (OECD, 2024b; Schauer, 2018). The common thread across these settings is that statutory protections designed for regular-track careers do not fully reach workers in precarious tiers, so aggregate policy coverage can overstate effective coverage for the women most at risk of exit.

The findings also speak to Goldin (2014)'s argument that the "last chapter" of gender convergence depends on the structure of jobs. Goldin emphasizes temporal flexibility within occupations; my results point to a different dimension of job structure - the regular/non-regular divide - but the broader implication is similar. In these data, the observed constraints align more closely with institutional job features than with preference-based accounts. Recent within-firm evidence from Japan is also consistent with this interpretation: Okuyama et al. (2025) show that the initial motherhood penalty within firms is driven by overtime and bonus reductions, but over time it transforms into a promotion penalty through the job-rank hierarchy. That pattern is directionally similar to my finding that the penalty shifts from an immediate extensive-margin break to a more persistent intensive-margin penalty.

9.3 Limitations

Several limitations should be noted. First, the mechanism regressions are estimated on 13 exit events in the complete-case sample, which limits precision and the ability to estimate richer interaction structures. Replication with a larger sample would strengthen the mechanism identification.

Second, I cannot rule out all forms of unobserved heterogeneity. While the selection tests address several leading alternative explanations, it remains possible that non-regular workers differ from regular workers in unobserved dimensions that independently predict exit. A research design with exogenous variation in contract type would provide stronger causal identification.

Relatedly, the pre-birth "marriage-stage" exits should be interpreted as marriage-related reporting patterns rather than a clean marriage effect. In Japan, marriage and expected near-term childbirth are tightly linked, so stage-one exits may bundle marriage transitions with anticipatory family planning. The paper therefore uses this stage as descriptive sequencing evidence, not as an isolated causal channel.

Third, my observation window of five years after first birth may not capture the full long-run trajectory. The earnings penalty shows no sign of recovery by $t = +5$, but recovery may occur at longer horizons as children enter school (cf. Angelov et al., 2016; Adda et al., 2017, who track penalties over 15-20 years).

Fourth, direct childcare-utilisation measurement is extremely limited in these panels. Although JHPS/KHPS include childcare and early-education items, many are not available consistently across waves, which prevents construction of a stable woman-year childcare panel aligned to event time. As a result, I cannot estimate dynamic effects of specific care arrangements (e.g., public/private nursery at ages 0-2 versus kindergarten pathways at ages 3-5) within the same longitudinal framework, nor can I measure whether individual women were denied childcare placement. The analysis therefore infers childcare pressure indirectly from exit reasons, household structure, and labour-market position. Since approximately 90 percent of mothers in the sample live in nuclear households without co-resident grandparents, childcare access may be a more binding constraint than my analysis suggests.

Cross-country evidence suggests that family care structures are related to child penalties. In China, parenthood is associated with negative labour-market effects for mothers, but those penalties are smaller where grandparental childcare is available (Meng et al., 2023; Deng et al., 2023). Related work also shows that grandchild care can reduce older adults' labour supply, which implies an intergenerational time-allocation tradeoff (Du et al., 2019). This is relevant for retirement-policy debates: simulations for China indicate that raising older women's retirement age can increase their labour supply while reducing grandmother-provided childcare and, in turn, lowering labour supply among younger low-skilled mothers (Gao, 2025). By contrast, evidence for Korea points to a large and persistent motherhood penalty and weaker post-birth career continuity, despite broad family-policy expansion (OECD, 2024b, 2025a). In that sense, Korea appears closer to the Japan pattern documented here, though institutional mechanisms and cultural aspects differ. My finding that labour-market structure dominates household characteristics does not rule out the possibility that childcare access is a necessary complementary condition.

Recent administrative evidence from Fukai and Kondo (2025) uses municipal tax records to document that mothers' salary income falls by up to 80 percent immediately after first birth and remains at roughly 50 percent of the pre-birth level four years later. This is a substantially larger intensive-margin penalty than the one estimated here, which is consistent with their longer earnings panel and more precise measurement. In a related study, Fukai and Kondo (2024) link childcare-admission records to parental tax data within specific municipalities and find that access to accredited childcare centres raises maternal employment by 18-40 percentage points for mothers of children under two. These studies exploit data that the JHPS/KHPS cannot match: individual-level administrative earnings trajectories and municipal childcare slot assignments. The present paper is complementary: the household panel provides contract type, firm size, and workplace-level mechanism variables that administrative tax records do not contain, at the cost of smaller samples and survey-based measurement. The ideal design would link employer-employee records (with contract type and firm characteristics) to municipal childcare-admission data and individual earnings histories. Japan is making progress on administrative data access through evidence-based policy-making initiatives, but a nationwide linked employer-employee dataset comparable to Denmark's IDA or the US LEHD does not yet exist (Kawaguchi, 2023). Until such linkage is feasible, the transparent descriptive approach used here fills a gap by documenting which labour-market positions predict exit, even if it cannot isolate specific causal channels with the precision that administrative data would allow.

Fifth, the pooled period (2004-2022) spans major policy and macroeconomic regime changes, including childcare expansions and the COVID period. The reported estimates should therefore be read as period-averaged patterns unless otherwise stratified (see Appendix D.10 for cohort-split and no-COVID sensitivity).

This caution also applies across source panels. In the JHPS/KHPS split check, the near lead ($t = -2$) and at-birth drop ($t = 0$) are highly stable across datasets, but the long horizon ($t = +5$) differs in sign and significance. I therefore treat the birth-year break as the most robust cross-panel result and interpret long-run recovery as panel-sensitive.

Sixth, the analytic sample is restricted to married women around first birth. This scope matches the dominant setting of childbirth in Japan, but it narrows external validity. The conclusions apply to married first-birth trajectories in this panel context and should not be mechanically generalized to unmarried mothers.

Seventh, selection into motherhood within the panel may itself be stratified by labour-market position. In the childless benchmark, non-regular employment is more common than in the mothers' analytic sample. This pattern is consistent with regular-track attachment operating as a precondition for observed family formation in-panel, which could compress observed penalties relative to a broader at-risk population.

Eighth, a substantial body of research on Japanese employment and family dynamics exists in Japanese-language sources that I do not survey here. The English-language literature cited in this paper most likely do not fully represent the state of knowledge in Japanese labour economics.

What would falsify the interpretation? Three results would substantially weaken the paper's core claims. First, if the matched childless placebo showed a sharp pseudo-birth break comparable to the maternal sample, the event-time attribution to childbirth would weaken - but the placebo coefficient at $t = 0$ is $+0.016$ ($p = 0.608$), and the joint far-lead test does not reject ($p = 0.749$). Second, if the non-regular contract-tier gradient disappeared when missingness was handled differently, the predictive stratification claim would weaken - but the odds ratio is stable across complete-case (7.50), missing-indicator (7.48), and high-information (7.39) specifications. Third, if the at-birth employment break varied substantially across early and late cohorts or across survey sources, the generality of the result would be in doubt - but the $t = 0$ coefficient is virtually identical across early (-0.131), late (-0.135), no-COVID (-0.145), KHPS-only, and JHPS-only samples. None of these diagnostics overturns the paper's core interpretation.

9.4 Concluding Remarks

The evidence in this paper indicates that Japan's motherhood employment penalty is closely associated with labour-market segmentation, not only with childcare availability. The women who exit are concentrated in more precarious job positions, and post-birth employment increasingly takes lower-quality forms: among those initially full-time, fewer than half remain so by $t = +5$, and conditional earnings are 96 man-yen lower with no clear recovery in the observed window. Father-side adjustment is not detectably large in this dataset (3 of 1,183 took childcare leave; wife-to-husband caregiving remains near 5-to-1). Taken together, these descriptive patterns are consistent with a segmented labour market in which continuity protections differ substantially by contract tier and employer type.

Achieving gender equality in Japanese employment will require more than expanding childcare or extending parental leave - policies that primarily benefit workers who already hold secure positions. It will require confronting the dual labour market itself: the divide between those whose jobs can accommodate parenthood and those whose jobs cannot.

References

Adda, J., Dustmann, C., and Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, 125(2):293–337.

Andresen, M. E. and Nix, E. (2022). What causes the child penalty? evidence from adopting and same-sex couples. *Journal of Labor Economics*, 40(4):971–1004.

Angelov, N., Johansson, P., and Lindahl, E. (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics*, 34(3):545–579.

Asai, Y. (2015). Parental leave reforms and the employment of new mothers: Quasi-experimental evidence from Japan. *Labour Economics*, 36:72–83.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–255.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Brinton, M. C. (1993). *Women and the economic miracle: gender and work in postwar Japan*. University of California Press.

Brinton, M. C. (2001). Married women's labor in east asian economies. In Brinton, M. C., editor, *Women's working lives in East Asia*, pages 1–37. Stanford University Press.

Budig, M. J. and England, P. (2001). The wage penalty for motherhood. *American Sociological Review*, 66(2):204–225.

Burdin, G., Kambayashi, R., and Kato, T. (2024). The impact of overtime limits on firms and workers: evidence from Japan's Work Style Reform. Technical Report 17583, Institute of Labor Economics (IZA), Bonn.

Casarico, A. and Lattanzio, S. (2023). Behind the child penalty: understanding what contributes to the labour market costs of motherhood. *Journal of Population Economics*, 36:1489–1511.

Children and Families Agency, Government of Japan (2024a). Municipal implementation plans for reducing childcare waiting lists. Official policy webpage. Accessed 2026-02-21.

Children and Families Agency, Government of Japan (2024b). Overview of free early childhood education and care (Mushoka). Official policy webpage. Accessed 2026-02-21.

Children and Families Agency, Government of Japan (2024c). Status of childcare provision and waiting children (FY2024 release). Official policy release webpage. Accessed 2026-02-21.

Correll, S. J., Benard, S., and Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology*, 112(5):1297–1338.

de Quinto, A., Hospido, L., and Sanz, C. (2021). The child penalty: evidence from Spain. *SERIEs*, 12(4):585–615.

Deng, Y., Zhou, Y., and Hu, D. (2023). Grandparental childcare and female labor market behaviors: Evidence from China. *Journal of Asian Economics*, 86:101614.

Du, F., Dong, X.-y., and Zhang, Y. (2019). Grandparent-provided childcare and labor force participation of mothers with preschool children in urban China. *China Population and Development Studies*, 2:347–368.

Dumauli, M. T. (2019). Motherhood wage penalty in Japan: What causes mothers to earn less in regular jobs? *Business and Economic Horizons*, 15(3):375–392.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.

Fukai, T. and Kondo, A. (2024). Access to formal childcare for toddlers and parental employment and earnings. Technical Report 387, ESRI Discussion Paper Series.

Fukai, T. and Kondo, A. (2025). Parental earnings trajectories around childbirth in Japan: evidence from local tax records. Technical Report 25-E-012, RIETI Discussion Paper Series.

Gao, H. (2025). Social security and female labor supply in China. *Journal of Economic Behavior & Organization*, 238:107203.

Genda, Y. (2005). *A nagging sense of job insecurity: The new reality facing Japanese youth.* International House of Japan, Tokyo.

Gietner, B. (2024). Japanese education statistics dataset (2005–2022): Prefecture-year panel aggregated from e-stat. GitHub repository and documentation. Compiled from e-Stat (Statistics of Japan); accessed 2026-02-21.

Goldin, C. (2014). A grand gender convergence: its last chapter. *American Economic Review*, 104(4):1091–1119.

Houseman, S. N. and Osawa, M. (2003). *Nonstandard work in developed economies: Causes and consequences*. W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

Kambayashi, R. and Kato, T. (2017). Long-term employment and job security over the past 25 years: a comparative study of Japan and the United States. *ILR Review*, 70(2):359–394.

Kawaguchi, D. (2023). Administrative data opens doors to frontier research. RIETI Column.

Kikuchi, S. (2026). Who bears the burden? heterogeneous labor market penalties of child and eldercare. *Journal of the Japanese and International Economies*. Forthcoming.

Kleven, H., Landais, C., and Leite-Mariante, G. (2024). The child penalty atlas. *Review of Economic Studies*, 92(5):3174–3207.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2019a). Child penalties across countries: evidence and explanations. *AEA Papers and Proceedings*, 109:122–126.

Kleven, H., Landais, C., and Søgaard, J. E. (2019b). Children and gender inequality: evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4):181–209.

Kuroda, S. and Yamamoto, I. (2013). Do peers affect determination of work hours? Evidence based on unique employee data from global Japanese firms in Europe. *Journal of Labor Research*, 34:359–388.

Kuziemko, I., Pan, J., Shen, J., and Washington, E. (2018). The mommy effect: Do women anticipate the employment effects of motherhood? Technical Report 24740, NBER Working Paper.

Meng, L., Zhang, Y., and Zou, B. (2023). The motherhood penalty in China: Magnitudes, trends, and the role of grandparenting. *Journal of Comparative Economics*, 51(1):105–132.

Mincer, J. and Ofek, H. (1982). Interrupted work careers: depreciation and restoration of human capital. *Journal of Human Resources*, 17(1):3–24.

Mincer, J. and Polachek, S. (1974). Family investments in human capital: earnings of women. *Journal of Political Economy*, 82(2):S76–S108.

Ministry of Health, Labour and Welfare (2019). Law concerning the welfare of workers who take care of children or other family members including child care and family care leave (English text). MHLW. Accessed 2026-02-14.

Ministry of Health, Labour and Welfare (2024). Kyōiku kunren kyūfu seido (Education and training benefits system): general overview and Q&A. Official MHLW policy webpage. Accessed 2026-02-23.

Ministry of Health, Labour and Welfare (2025). Basic survey on wage structure. Official statistics portal. Accessed 2026-02-21.

Nagase, N. (2012). Japanese female labor supply and the tax/social insurance system. *Japanese Journal of Labour Studies*, 54(12):12–25.

Nakayama, M. and Ishikawa, Y. (2025). Effects of paternity leave take-up rate in fathers' industry of work on mothers' employment and health. *Japan Labor Issues*, 9(51):13–29.

Nemoto, K. (2016). *Too few women at the top: the persistence of inequality in Japan*. Cornell University Press.

Nishitateno, S. and Shikata, M. (2017). Has improved daycare accessibility increased Japan's maternal employment rate? municipal evidence from 2000–2010. *Journal of the Japanese and International Economies*, 44:67–77.

North, S. (2009). Negotiating what's 'natural': persistent domestic gender role inequality in Japan. *Social Science Japan Journal*, 12(1):23–44.

Ochiai, E. (1997). *The Japanese family system in transition: A sociological analysis of family change in postwar Japan*. LTCB International Library Foundation, Tokyo.

OECD (2021). Creating responsive adult learning opportunities in Japan. Technical report, OECD Publishing, Paris.

OECD (2022). The new workplace in Japan. Technical report, OECD Publishing, Paris.

OECD (2024a). OECD economic surveys: Japan 2024. Technical report, OECD Publishing, Paris.

OECD (2024b). OECD economic surveys: Korea 2024. Technical report, OECD Publishing, Paris.

OECD (2025a). Korea's unborn future: Understanding low-fertility trends. Technical report, OECD Publishing, Paris.

OECD (2025b). Net childcare costs (indicator). OECD Data indicator page. Accessed 2026-02-21.

Okuyama, Y., Murooka, T., and Yamaguchi, S. (2025). Unpacking the child penalty using personnel data: how promotion practices widen the gender pay gap. Technical Report 17673, IZA Discussion Papers.

Piotrowski, M., Kalleberg, A., and Rindfuss, R. R. (2015). Contingent work rising: implications for the timing of marriage in Japan. *Journal of Marriage and Family*, 77(5):1039–1056.

Raymo, J. M. and Iwasawa, M. (2005). Marriage market mismatches in Japan: an alternative view of the relationship between women's education and marriage. *American Sociological Review*, 70(5):801–822.

Schauer, J. (2018). Labor market duality in Korea. Working Paper WP/18/126, International Monetary Fund.

Statistics Bureau of Japan (2025). Statistical observations of prefectures. Official statistics portal. Accessed 2026-02-21.

Waldfogel, J. (1998). Understanding the "family gap" in pay for women with children. *Journal of Economic Perspectives*, 12(1):137–156.

Yamaguchi, K. (2019). *Gender inequalities in the Japanese workplace and employment*. Springer, Singapore.

Yu, W.-h. (2009). *Gendered trajectories: Women, work, and social change in Japan and Taiwan*. Stanford University Press.

Appendix A. Childless benchmark and placebo test

Because the harmonized analysis panels are constructed from mothers only, a childless comparison group is not available within the main sample. I therefore build a descriptive benchmark from the raw panels (married women with no observed births and at least five years of observation). This group has N_ids = 4,111 and N_obs = 50,222, with mean employment 53.2 percent, mean weekly hours 28.5, 71.9 percent non-regular employment, and 59.9 percent small-firm employment. Yearly employment rates are stable in the 0.51-0.57 range from 2004-2022. I emphasize that this is descriptive context rather than a causal counterfactual. Tables A.C1-A.C3 report the childless summary statistics and annual employment rates.

Table 13: Descriptive benchmark sample of married childless women (context only; not a causal control group).

N (observations)	N (women)	Employment rate	Mean weekly hours
50,222	4,111	0.532	28.54

Notes: Employment rate is the proportion employed. Mean weekly hours are averaged across all observations.

Table 14: Descriptive benchmark for employed observations in the married childless sample (context only).

N (observations)	N (women)	Mean weekly hours	Median weekly hours
26,704	3,088	28.54	26.00

Notes: Statistics are conditional on employment and are reported for descriptive context, not identification.

I also implement a placebo event study by assigning each childless woman a pseudo-birth year matched to mothers on dataset, age bin, and cohort bin, then estimating the same event-study specification with year fixed effects. Pseudo-event years are drawn from the mothers' first-birth-year donor pools with replacement and constrained to years in which each childless woman is actually observed, so the placebo does not create artificial missingness by construction. Match quality is strong (91.7% exact matches, 5.9% dataset-only matches, and 2.4% dataset-age matches). The placebo profile does not show a pseudo-birth discontinuity: at $t = 0$ the coefficient is +0.016 (SE 0.031, 95% CI $[-0.045, 0.076]$, $p = 0.608$), and the near lead at $t = -2$ is -0.003 ($p = 0.937$). Coefficients are small and statistically insignificant across the window, including far leads (e.g., $t = -4$: $p = 0.373$). A joint far-lead placebo test for $t \in \{-5, -4, -3\}$ also does not reject ($p = 0.749$). This placebo is therefore interpreted as a composition diagnostic rather than a strict falsification test: pseudo-assignment is still non-exogenous to panel support, so small non-zero coefficients can arise from lifecycle and selection drift. I therefore treat the childless benchmark as descriptive context rather than an identification strategy. Table A.C4 reports the placebo event-study coefficients.

Note on individual-FE event-study figures. For transparency, I also generate versions of the main event-study figures with individual and year fixed effects. In a short event window with event-time dummies, this specification is weakly identified and produces large standard errors and smooth ramp (triangular) patterns. I therefore do not

Table 15: Annual employment rate in the descriptive childless benchmark sample (married women with no observed births). Rates are proportions, not percentages, and are shown to document the absence of a birth-related break in this benchmark group.

Year	N	Emp. rate (proportion)
2004	1,642	0.516
2005	1,642	0.543
2006	1,641	0.548
2007	2,313	0.548
2008	2,312	0.567
2009	3,625	0.533
2010	3,535	0.531
2011	3,467	0.527
2012	3,724	0.531
2013	3,611	0.518
2014	3,357	0.525
2015	3,153	0.531
2016	2,953	0.525
2017	2,684	0.534
2018	2,456	0.537
2019	2,269	0.539
2020	2,104	0.532
2021	1,916	0.523
2022	1,818	0.509

Table 16: Matched childless pseudo-event placebo: employment event-study coefficients (year fixed effects).

Event time	Estimate	SE	p-value
$t = -5$	-0.037	0.058	0.524
$t = -4$	-0.039	0.044	0.373
$t = -3$	-0.031	0.039	0.427
$t = -2$	-0.003	0.035	0.937
$t = +0$	0.016	0.031	0.608
$t = +1$	0.038	0.031	0.221
$t = +2$	0.062	0.032	0.052
$t = +3$	0.060	0.032	0.063
$t = +4$	0.056	0.034	0.099
$t = +5$	0.049	0.036	0.173

Joint Wald test for placebo far leads ($t = -5, -4, -3$): p=0.749.

emphasize those figures; the main text relies on the year-FE specification and descriptive trajectories.

The second-birth dynamics reinforce this interpretation. In the main sample, the share with a second birth rises to 19.2 percent by $t = +2$, 33.8 percent by $t = +3$, and 44.1 percent by $t = +5$ (conditional on women observed at $t = +5$). Censoring post-second-birth observations or excluding women with a second birth within three years makes the long-run coefficients less negative and close to zero by $t = +5$. Stratifying the event study shows that the "second-birth" group exhibits persistently more negative coefficients (e.g., $t = 0$: -0.151; $t = +5$: -0.068), while the "no second birth" group recovers to near zero by $t = +3$ to $+5$. These patterns suggest that the muted rebound in the full sample partly reflects subsequent births rather than permanent non-recovery.

Appendix B. Second-birth dynamics

This appendix reports second-birth spacing, the share of women with a second birth by event time, and stratified event-study coefficients for mothers who do and do not have a second birth within the panel window.

The purpose is interpretive rather than causal. In this setting, subsequent fertility is an endogenous life-course process that overlaps with post-first-birth recovery time. The stratified coefficients therefore describe how trajectories differ across fertility paths, not how a second birth causally shifts outcomes holding all else fixed.

The key implication for the main text is scope: pooled long-horizon coefficients mix first-birth recovery with additional-child transitions. The second-birth tables and figure make that mixture explicit and show that part of the medium-run persistence in the pooled sample reflects subsequent births rather than only non-recovery from first birth.

Table 17: Observed spacing between first and second birth in the analytic sample. Counts are the number of mothers with a recorded second birth at each spacing (years).

Years to Second	Count
1	33
2	83
3	72
4	35
5	16
6	10
7	3
8	2
13	1

Table 18: Share of mothers with a recorded second birth by event time relative to first birth. Share is a proportion (0-1), with N denoting the number of mothers observed at each event time.

Event time	N	Share (proportion)
0	662	0.000
1	654	0.050
2	594	0.192
3	509	0.338
4	487	0.407
5	454	0.441

Table 19: Employment event-study coefficients stratified by whether mothers have a second birth within the panel window. Coefficients are relative to $t = -1$ (year fixed effects); values are descriptive and reflect heterogeneous trajectories rather than causal second-birth effects.

Group	Event Time	Estimate	Std. Error	CI Lower	CI Upper
No second birth	-5	-0.125	0.058	-0.239	-0.012
Second birth	-5	-0.423	0.046	-0.512	-0.333
No second birth	-4	-0.102	0.051	-0.203	-0.002
Second birth	-4	-0.361	0.047	-0.453	-0.269
No second birth	-3	-0.051	0.046	-0.142	0.040
Second birth	-3	-0.219	0.049	-0.316	-0.123
No second birth	-2	0.042	0.040	-0.037	0.121
Second birth	-2	-0.089	0.047	-0.182	0.003
No second birth	0	-0.124	0.035	-0.193	-0.054
Second birth	0	-0.151	0.044	-0.238	-0.065
No second birth	1	-0.115	0.037	-0.187	-0.043
Second birth	1	-0.149	0.045	-0.236	-0.061
No second birth	2	-0.070	0.038	-0.145	0.005
Second birth	2	-0.125	0.045	-0.214	-0.036
No second birth	3	0.002	0.040	-0.077	0.081
Second birth	3	-0.072	0.046	-0.163	0.019
No second birth	4	0.029	0.041	-0.050	0.109
Second birth	4	-0.072	0.048	-0.166	0.021
No second birth	5	0.013	0.042	-0.070	0.095
Second birth	5	-0.068	0.049	-0.164	0.029

Figure 17 is constructed from an at-risk subgroup rather than from all mothers: women who were full-time at $t = -1$ and are not full-time at $t = +1$. For each event year $t \in \{2, 3, 4, 5\}$, I compute first returns to full-time among women still at risk, then compute the cumulative first-return proportion. The resulting series is a descriptive cumulative first-return profile, not a Kaplan-Meier survival estimate.

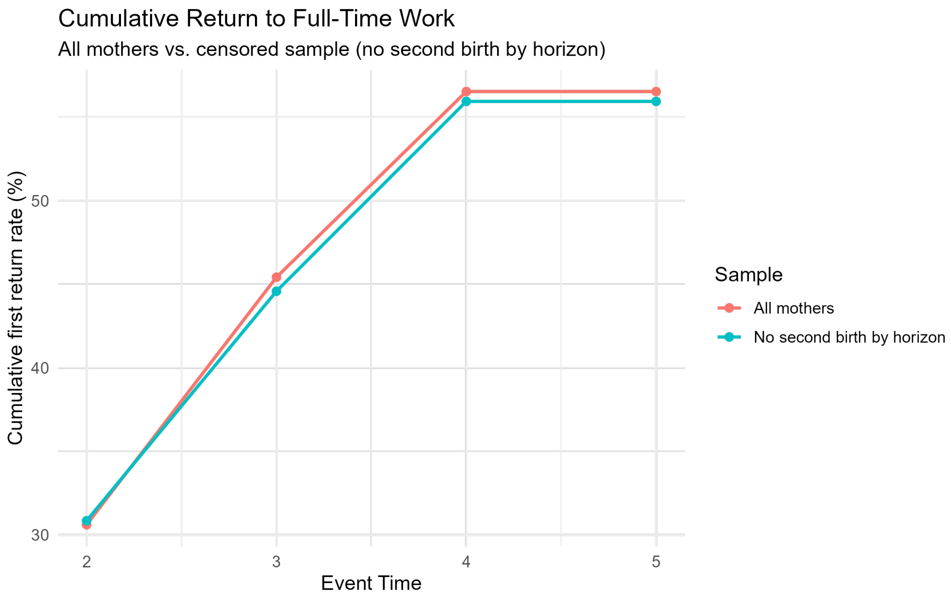

Figure 17: Appendix Figure B1: Cumulative return to full-time employment after childbirth. This curve is estimated on the at-risk subgroup only (mothers full-time at $t = -1$ and not full-time at $t = +1$), not on all mothers. Points show the cumulative first-return proportion over $t = 2$ to $t = 5$, so the series is descriptive rather than a Kaplan-Meier survival estimate.

Appendix C. Father placebo event study

This appendix reports the event-study coefficients for husbands' weekly hours around first birth. The estimates are small and statistically indistinguishable from zero across the event window, consistent with no systematic adjustment in fathers' labour supply.

This is a placebo-style benchmark, not a structural model of paternal labour supply. Its value is comparative: against large maternal employment and hours shifts, the husband's near-flat trajectory helps document asymmetric household adjustment around childbirth in this sample.

At the same time, a near-zero father-hours pattern does not identify an exclusion restriction for maternal outcomes. It should be read as descriptive supporting evidence for the gendered division-of-labour mechanism developed in the main text.

Table 20: Event-study coefficients for husband weekly hours around first birth (year fixed effects). Estimates are relative to $t = -1$ and reported with robust standard errors and 95% confidence intervals; the profile is broadly flat, indicating no systematic father-hours adjustment.

Event time	Estimate (hours/week)	SE	95% CI lower	95% CI upper
-5	5.468	3.0004	-0.4128	11.3488
-4	-3.5484	2.0471	-7.5606	0.4639
-3	-0.978	1.5621	-4.0396	2.0837
-2	0.7091	1.1979	-1.6388	3.057
-1	0	0	0	0
0	0.4497	0.9717	-1.4548	2.3543
1	0.2265	1.0344	-1.8009	2.2538
2	0.0181	1.0522	-2.0443	2.0804
3	0.7421	1.0791	-1.373	2.8572
4	-0.1797	1.1171	-2.3691	2.0098
5	-0.4993	1.1313	-2.7167	1.718

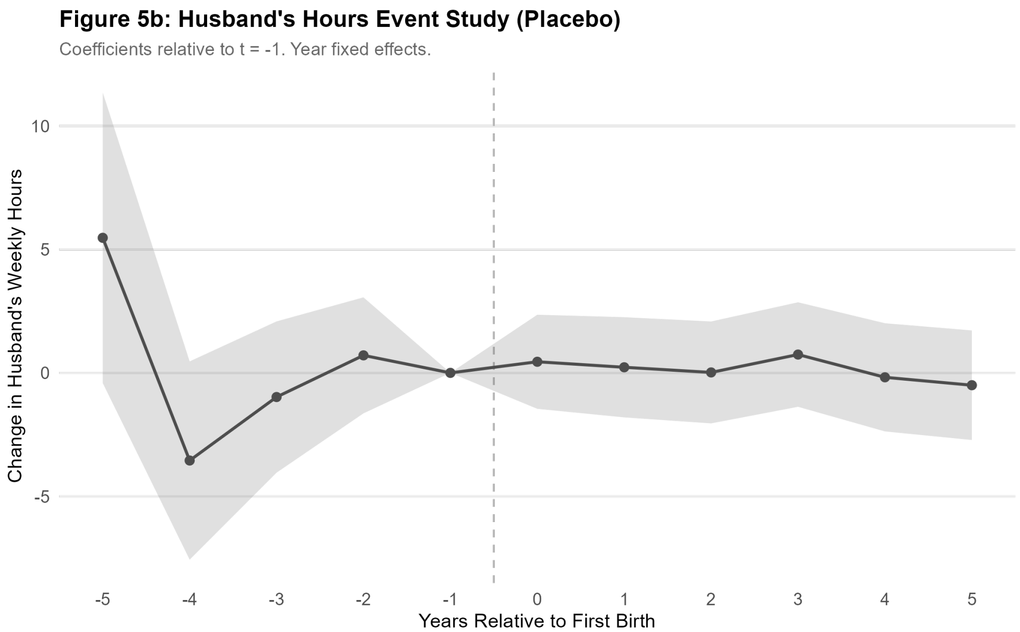

Figure 18: Appendix Figure C1: Husband weekly-hours event study (year fixed effects; robust standard errors). Coefficients are relative to $t = -1$. Estimates are small and imprecise, with no stable post-birth pattern consistent with a father labour-supply penalty.

Appendix D. Identification, harmonisation, and additional sensitivity checks

This appendix consolidates diagnostics that support interpretation of the main estimates and clarifies which remaining checks are planned for the next revision cycle.

D.1 Event-time identification and individual-FE collinearity

The core design of this study uses event-time indicators around first birth with calendar-year fixed effects. In this setup, adding individual fixed effects substantially weakens identification because event time, calendar time, and birth cohort are mechanically linked in a short and unbalanced panel. Empirically, this appears as extreme variance inflation in the individual-plus-year FE version: point estimates remain directionally similar to the year-FE specification, but standard errors become uninformative. This is why in the main text I report year-FE trajectories and treat individual-FE figures as transparency checks instead of the primary inferential design.

The interpretation is therefore explicit: the reported coefficients are descriptive sample-average trajectories relative to $t = -1$, not causal within-person treatment effects. The key timing result remains that the near lead at $t = -2$ is close to zero, while the break at $t = 0$ is sharp and negative.

D.2 Pre-trend diagnostics in an unbalanced panel

Far leads ($t = -5, -4, -3$) are negative and jointly significant. In this application, that pattern is interpreted as composition in an unbalanced panel rather than anticipatory response to childbirth. Women observed at far pre-birth horizons are a selected subgroup with different life-cycle employment profiles, and support thins markedly in those tails. The identification-relevant diagnostic is the near lead at $t = -2$, which remains small across specifications.

This interpretation is reinforced by robustness exercises that trim far leads, restrict to balanced pre-support, and add a dataset indicator. Across these alternatives, the near lead remains economically small and statistically weak, while post-birth coefficients preserve their qualitative pattern.

D.3 JHPS/KHPS pooling and harmonisation

The pooled panel is built from the official harmonisation program and variable correspondence files. Core constructs used in the paper (employment, hours, earnings, contract type, and firm size) are mapped through that framework and checked against survey-specific coding before analysis. As a sensitivity check, the main employment specification includes a JHPS/KHPS dataset indicator; this does not materially change the near-lead or at-birth coefficients.

The practical implication is that pooling does not appear to drive the headline dynamics. Nonetheless, the estimates should be read as trajectories in the harmonised analytic sample of married first-birth women, not as nationally weighted population moments.

Table 21 clarifies where pooling is and is not consequential. The two identification-relevant moments are stable across surveys: the near lead at $t = -2$ is close to zero and statistically insignificant in both KHPS and JHPS, and the at-birth coefficient at $t = 0$ is nearly identical in magnitude (about -13.4 percentage points in each panel). The long horizon is less stable: by $t = +5$, KHPS is near zero while JHPS remains negative. This divergence is consistent with panel-sensitive long-run composition (different support, attrition profiles, and sample sizes at distant event times) rather than with disagreement about the birth-year break itself. Accordingly, the paper treats the sharp at-birth decline as the most robust cross-panel finding and interprets long-run recovery more cautiously.

Table 21: Survey-split stability check for key employment event-study coefficients (year fixed effects).

Sample	Event time	Estimate	SE	p-value	N obs	N women
Pooled	$t = -2$	-0.012	0.031	0.703	5,212	662
Pooled	$t = +0$	-0.132	0.027	0.000	5,212	662
Pooled	$t = +5$	-0.028	0.032	0.384	5,212	662
KHPS	$t = -2$	0.002	0.039	0.966	3,323	394
KHPS	$t = +0$	-0.134	0.036	0.000	3,323	394
KHPS	$t = +5$	0.010	0.040	0.796	3,323	394
JHPS	$t = -2$	-0.027	0.049	0.576	1,889	268
JHPS	$t = +0$	-0.136	0.044	0.002	1,889	268
JHPS	$t = +5$	-0.126	0.057	0.027	1,889	268

D.4 Education measurement and Newlyweds-module diagnostic

In this study I do not use education-attainment heterogeneity as a core split because JHPS/KHPS do not provide a single harmonised, respondent-level highest-attainment variable with stable coverage for the full mother event-study sample. Education-related items in the regular married-respondent instruments are not attainment stocks. In the household roster, schooling fields capture current enrolment status (in school vs not in school) and school type. In JHPS, roster schooling/work cells for "yourself" and "your spouse" are crossed out; in KHPS, roster schooling and work are combined in a status field. In both surveys, these roster items are unsuitable as highest-completed-education measures for wives. Likewise, respondent/spouse modules on current attendance, recent training actions, and time-use/education expenditures measure contemporaneous human-capital activity rather than educational attainment.

The only module with an attainment-style schooling item is the Newlyweds spouse questionnaire, administered when a panel respondent newly reports marriage in-panel. This instrument is answered by the respondent's spouse and is therefore not a standard annual respondent covariate. To provide transparency, I construct a bounded appendix diagnostic using newlyweds variable "School that R last attended", recoded into a college-plus indicator (junior college/specialized, university, graduate school).

Coverage is limited and selected in the analytic sample: of 662 mothers, 89 (13.4%) have matched non-missing newlyweds education. Within this matched subset, 60.7% are college-plus, 33.7% are non-college, and 5.6% are unclassified. Because inclusion depends on in-panel marriage timing and spouse-module availability, this diagnostic is descriptive only and is not used as identifying heterogeneity in the main design.

D.5 Prefecture-level childcare context (external data)

To contextualise the metropolitan split in the main text, I compile prefecture-level childcare indicators from external official statistics (all 47 prefectures). The purpose is descriptive: to document the degree of geographic heterogeneity in childcare environments over time. These series are not linked to individual JHPS/KHPS respondents and are therefore not used as identifying variation in the individual-level event-study design.

The indicators show substantial dispersion and structural transition. Day-nursery provision per 100,000 children aged 0-5 rises in median terms from 397.7 (2005) to

Table 22: Coverage of newlyweds-module education in the analysis sample

Metric	Value
Analysis sample (N)	662
Matched newlyweds education (N)	89
Matched share (%)	13.4
College-plus share in matched (%)	60.7
Non-college share in matched (%)	33.7

Notes: Newlyweds education uses spouse-module item "School that R last attended". "College-plus" includes junior college/specialized, university, and graduate school categories.

572.0 (2020), while cross-prefecture spread remains wide (2020: min 404.8, max 1042.8). Nursery-utilisation medians are above full capacity in 2005 and 2014 (103.8 and 102.0), then lower in 2020 (92.2), consistent with easing average pressure alongside persistent regional variation. Public provision shares decline markedly over the same period (public day-nursery institution share median: 56.2% in 2005 to 28.1% in 2020; public day-nursery children share median: 48.8% to 26.7%). For authorised child centres, available in the later period, the 2022 median is 212.2 per 100,000 children (min 26.5, max 667.4). Table 23 reports the summary statistics. The prefecture-level series are compiled from official e-Stat tables; the harmonised extraction code and compiled panel used for these appendix summaries are documented in Gietner (2024).

Fertility, childcare demand, and childcare supply can co-evolve at the prefecture level, so these series should not be read as exogenous supply shifters.

Table 23: Prefecture-level childcare environment dispersion (all 47 prefectures). Facility indicators are reported per 100,000 children aged 0-5. Share indicators refer to public-provider institution shares or enrolled-children shares. Source: official e-Stat tables; harmonised extraction documented in Gietner (2024).

Indicator	Year	Min	P25	Median	P75	Max
Day nurseries per 100k (0-5)	2005	172.9	284.9	397.7	535.5	760.0
Day nurseries per 100k (0-5)	2014	258.9	323.6	425.5	561.9	775.6
Day nurseries per 100k (0-5)	2020	404.8	501.6	572.0	740.6	1042.8
Nursery utilisation rate (%)	2005	90.3	98.8	103.8	106.7	112.1
Nursery utilisation rate (%)	2014	88.1	96.5	102.0	106.0	115.1
Nursery utilisation rate (%)	2020	74.5	88.7	92.2	95.0	100.2
Public day-nursery share (%)	2005	18.9	41.2	56.2	66.0	82.1
Public day-nursery share (%)	2014	5.6	26.2	40.9	55.0	78.6
Public day-nursery share (%)	2020	0.4	19.6	28.1	44.8	63.6
Public day-nursery children share (%)	2005	17.0	37.0	48.8	59.5	79.0
Public day-nursery children share (%)	2014	4.9	23.7	38.2	46.3	75.4
Public day-nursery children share (%)	2020	0.6	15.0	26.7	35.7	57.9
Authorized child centers per 100k (0-5)	2022	26.5	158.7	212.2	298.0	667.4

D.6 Attrition and support composition

I report attrition/support diagnostics to make composition shifts explicit. The event-time support is strongly unbalanced: 186 women at $t = -5$, 662 at $t = 0$, and 454 at $t = +5$. I also compare baseline characteristics at $t = -1$ by whether women are observed at $t = +5$. Baseline employment differs (47.7% for those observed at $t = +5$ versus 56.2% for those not observed; $p = 0.045$), while baseline hours and income do not differ significantly ($p = 0.509$ and $p = 0.311$). These diagnostics do not convert the design into a causal panel estimator, but they quantify how much long-horizon interpretation depends on changing support.

Table 24: Attrition/support diagnostics: baseline balance by observation at $t = +5$.

Group	N women	Employment rate	Mean weekly hours	Mean annual income
Observed at $t = +5$	453	0.477	33.64	252.30
Not observed at $t = +5$	203	0.562	34.92	276.57
Difference p-value		0.045	0.509	0.311

Notes: Employment is measured as a proportion at $t = -1$. Hours and income are baseline means at $t = -1$.

Table 25: IPW diagnostics for observability at $t = +5$.

Diagnostic	Value	Note
Weight mean	1.521	Baseline sample
Weight p95 / p99	2.117 / 2.212	Tail concentration
Weight max	2.292	Extreme leverage check
Trim share (low / high)	0.0% / 0.0%	Bounds: [0.05, 0.95]
Employment gap (obs − full), unweighted	-0.026	Baseline comparison
Employment gap (obs − full), IPW	-0.006	Weighted comparison

Table 26: IPW sensitivity for observability at $t = +5$: key employment event-study coefficients.

Specification	Event time	Estimate	SE	p-value
Unweighted	$t = -2$	-0.012	0.031	0.703
Unweighted	$t = +0$	-0.132	0.027	¡0.001
Unweighted	$t = +5$	-0.028	0.032	0.384
IPW (obs at t=+5)	$t = -2$	-0.024	0.031	0.436
IPW (obs at t=+5)	$t = +0$	-0.136	0.028	¡0.001
IPW (obs at t=+5)	$t = +5$	-0.057	0.032	0.076

The IPW model is a logit for $1\{\text{observed at } t = +5\}$ estimated on baseline covariates (employment status, weekly hours, annual income, survey indicator, and age/education fields when available in the analytic file). Balance diagnostics are reported in levels: employment is in proportions (percentage-point interpretation), hours in weekly hours, and income in annual man-yen units. In this run, estimated propensity scores already lie within the trimming bounds $[0.05, 0.95]$ (no trimming activated). IPW is constructed for observability at $t = +5$ because attrition pressure is strongest at the long horizon; it is used as a sensitivity check for late-horizon composition rather than as a full correction for all event times. Under IPW, the long-horizon coefficient shifts from -0.028 ($p = 0.384$) to -0.057 ($p = 0.076$, 95% CI $[-0.120, 0.006]$), while the near lead and birth-year break remain unchanged in sign and significance.

As an additional support check, I re-estimate the employment event study on a balanced short window, restricting to women observed continuously from $t = -2$ through $t = +1$. This removes short-horizon support imbalance by construction. The balanced-panel estimates are very close to the full-sample estimates in this window: the near lead remains near zero, and the childbirth-year break remains large and negative. This reinforces the interpretation that the core $t = 0$ result is not an artefact of short-window panel imbalance.

Table 27: Balanced-panel robustness: employment event-study coefficients for women observed continuously from $t = -2$ to $t = +1$.

	Full sample			Balanced panel		
Event time	Coef.	SE	p	Coef.	SE	p
$t = -2$	-0.008	0.031	0.781	-0.058	0.035	0.105
$t = +0$	-0.131	0.028	<0.001	-0.128	0.035	<0.001
$t = +1$	-0.129	0.029	<0.001	-0.121	0.035	<0.001
N obs		2,425			1,780	
N women		662			445	

Notes: Balanced panel restricts to women observed at all event times from $t = -2$ to $t = +1$. Specification uses year fixed effects with heteroskedasticity-robust standard errors. Reference period is $t = -1$.

I also run cleaner-sample sensitivity checks aimed at the far-lead composition concern. In an age-restricted sample (age ≥ 23), the childbirth break remains large ($t = 0 = -0.123$, $p < 0.001$) and the near lead remains close to zero ($t = -2 = -0.007$, $p = 0.836$), closely matching baseline estimates. The balanced short-window sample likewise preserves the core break ($t = 0 = -0.128$, $p < 0.001$). Stricter pre-window attachment restrictions generate larger coefficients, indicating subgroup selection toward higher-attachment women rather than a neutral support correction. Table 28 reports these estimates.

D.7 Intensive-margin selection and scope

Hours, earnings, and implied hourly wage are estimated conditional on employment. These intensive-margin coefficients therefore combine within-person adjustments with selection into who remains employed at each event time. In this study we should interpret them as evidence of post-birth labour market downgrading among employed mothers, while avoiding stronger claims that would require selection-corrected structural modelling.

I therefore report two additional diagnostics. First, I estimate unconditional hours and earnings event studies by coding non-employment as zero outcomes. The post-birth declines remain large and precisely estimated under both heteroskedasticity-robust and

Table 28: Cleaner-sample sensitivity: key employment event-study coefficients.

Specification	Event time	Estimate	SE	p-value	N obs	N women
Baseline (full analytic sample)	$t = -2$	-0.012	0.027	0.662	5212	662
Baseline (full analytic sample)	$t = 0$	-0.132	0.025	0.000	5212	662
Baseline (full analytic sample)	$t = +1$	-0.124	0.026	0.000	5212	662
Baseline (full analytic sample)	$t = +5$	-0.028	0.036	0.443	5212	662
Age-restricted (age ≥ 23)	$t = -2$	-0.007	0.032	0.836	3505	651
Age-restricted (age ≥ 23)	$t = 0$	-0.123	0.031	0.000	3505	651
Age-restricted (age ≥ 23)	$t = +1$	-0.115	0.041	0.005	3505	651
Age-restricted (age ≥ 23)	$t = +5$	-0.034	0.048	0.473	3505	651
Balanced short window ($t = -2$ to $t = +1$)	$t = -2$	-0.058	0.031	0.065	1780	445
Balanced short window ($t = -2$ to $t = +1$)	$t = 0$	-0.128	0.033	0.000	1780	445
Balanced short window ($t = -2$ to $t = +1$)	$t = +1$	-0.121	0.031	0.000	1780	445
Balanced short window ($t = -2$ to $t = +1$)	$t = +5$	–	–	–	1780	445
Pre-window attached (obs. $-3, -2, -1$; employed ≥ 2)	$t = -2$	0.063	0.027	0.018	1398	143
Pre-window attached (obs. $-3, -2, -1$; employed ≥ 2)	$t = 0$	-0.258	0.042	0.000	1398	143
Pre-window attached (obs. $-3, -2, -1$; employed ≥ 2)	$t = +1$	-0.278	0.052	0.000	1398	143
Pre-window attached (obs. $-3, -2, -1$; employed ≥ 2)	$t = +5$	-0.148	0.076	0.049	1398	143
Always-married pre-window	$t = -2$	0.137	0.026	0.000	3119	436
Always-married pre-window	$t = 0$	-0.186	0.029	0.000	3119	436
Always-married pre-window	$t = +1$	-0.187	0.029	0.000	3119	436
Always-married pre-window	$t = +5$	-0.101	0.040	0.012	3119	436

Notes: All specifications use year fixed effects and heteroskedasticity-robust SEs. Reference period is $t = -1$. "–" indicates that the balanced short-window design does not identify $t = +5$.

woman-clustered SEs, while the near lead remains close to zero. Second, I report Lee-style trimming sensitivity for the conditional intensive-margin outcomes. For post-birth years, the bounds remain negative for earnings from $t = +1$ onward and for hours at later horizons, though uncertainty is wider near birth. These bounds are most informative in post-birth windows; far pre-birth windows require large trimming shares and should be interpreted cautiously.

Table 29: Unconditional intensive-margin event studies (hours and earnings) relative to $t = -1$.

Outcome	Event time	Estimate	SE	p-value	SE type
Annual earnings (unconditional)	$t = -2$	-6.741	7.533	0.371	Cluster (woman)
Annual earnings (unconditional)	$t = -2$	-6.741	11.481	0.557	Hetero
Annual earnings (unconditional)	$t = +0$	-55.909	6.841	<0.001	Cluster (woman)
Annual earnings (unconditional)	$t = +0$	-55.909	9.971	<0.001	Hetero
Annual earnings (unconditional)	$t = +1$	-86.845	8.076	<0.001	Cluster (woman)
Annual earnings (unconditional)	$t = +1$	-86.845	9.197	<0.001	Hetero
Annual earnings (unconditional)	$t = +5$	-52.237	11.140	<0.001	Cluster (woman)
Annual earnings (unconditional)	$t = +5$	-52.237	11.305	<0.001	Hetero
Weekly hours (unconditional)	$t = -2$	-0.575	0.940	0.541	Cluster (woman)
Weekly hours (unconditional)	$t = -2$	-0.575	1.262	0.648	Hetero
Weekly hours (unconditional)	$t = +0$	-3.930	0.798	<0.001	Cluster (woman)
Weekly hours (unconditional)	$t = +0$	-3.930	1.117	<0.001	Hetero
Weekly hours (unconditional)	$t = +1$	-6.054	0.897	<0.001	Cluster (woman)
Weekly hours (unconditional)	$t = +1$	-6.054	1.112	<0.001	Hetero
Weekly hours (unconditional)	$t = +5$	-6.034	1.120	<0.001	Cluster (woman)
Weekly hours (unconditional)	$t = +5$	-6.034	1.159	<0.001	Hetero

Notes: Outcomes are defined on the full sample (not conditional on employment): hours are set to 0 when non-employed; earnings are set to 0 when non-employed. For employed observations with missing reported hours/earnings, outcomes remain missing. Models include calendar-year fixed effects.

Table 30: Lee-style trimming sensitivity for intensive-margin outcomes (hours and earnings measured among employed women), relative to $t = -1$.

Outcome	Event time	Observed diff.	Lower bound	Upper bound	Trim share	N trimmed
Annual earnings, man-yen	$t = +0$	-46.863	-112.561	34.332	0.234	74
Annual earnings, man-yen	$t = +1$	-124.667	-178.071	-54.856	0.195	61
Annual earnings, man-yen	$t = +2$	-54.096	-84.242	-8.214	0.113	35
Annual earnings, man-yen	$t = +3$	-67.812	-67.812	-67.812	0.000	0
Annual earnings, man-yen	$t = +4$	-73.095	-96.601	-65.894	0.039	9
Annual earnings, man-yen	$t = +5$	-74.104	-100.269	-63.188	0.056	13
Weekly hours	$t = +0$	1.471	-5.791	7.185	0.234	77
Weekly hours	$t = +1$	-3.791	-9.954	1.110	0.195	64
Weekly hours	$t = +2$	-2.814	-6.336	0.343	0.113	37
Weekly hours	$t = +3$	-5.035	-5.035	-5.035	0.000	0
Weekly hours	$t = +4$	-6.556	-7.682	-5.547	0.039	10
Weekly hours	$t = +5$	-8.805	-10.287	-7.536	0.056	13

Notes: Lee-style trimming sensitivity equalizes employment shares between each event-time cell and $t = -1$ by trimming the higher-employment cell. This is a selection- robustness diagnostic for conditional outcomes, not a causal treatment-control Lee bound.

D.8 Mechanism diagnostics: pre-birth margin ($t = -2$)

The table below reports the nested pre-birth mechanism regressions (M1–M5). Because this margin contains only 13 exit events, the estimates should be read as lower-power supporting evidence on early risk stratification rather than as a central result.

Table 31: Nested logistic regressions for pre-birth employment exit at $t = -2$

	M1	M2	M3	M4	M5
Non-regular contract	7.11^{**}	6.41^{**}	6.00^{**}	6.63^{**}	6.18^{**}
	(0.784)	(0.791)	(0.795)	(0.793)	(0.797)
Small firm		4.14^{**}	3.80^{**}	4.19^{**}	3.94^{**}
		(0.635)	(0.652)	(0.637)	(0.656)
Commute time			0.99		0.99
			(0.017)		(0.016)
Husband overwork				1.21	1.34
				(0.607)	(0.615)
N	185	185	169	176	160
Events	13	13	13	13	13

Notes: The outcome is a binary indicator equal to one if a woman exits employment between $t = -2$ and $t = -1$; the sample comprises women in the mechanism complete-case sample (13 exit events). Odds ratios are reported with log-odds standard errors in parentheses. M1 includes contract type only; M2 adds firm size; M3 adds one-way commute time (minutes); M4 adds a husband overwork indicator (≥ 50 hours/week); M5 includes all four predictors. N varies across models because of missing covariate values. ** $p < 0.05$. "Non-regular contract" refers to the survey contract-status classification and includes survey-coded part-time (*pāto*), arbeit, dispatch, and contract employment. It is distinct from the hours-based full-time/part-time classification used elsewhere in the paper. "Small firm" = employer with fewer than 30 employees. Data source: KHPS/JHPS 2004–2022, married first-birth women.

The table 32 reports leave-one-out and influence diagnostics for the baseline two-predictor mechanism model ('non-regular' and 'small firm'). The jackknife ranges show stable sign and magnitude for both predictors, and influence diagnostics indicate no single case that reverses the qualitative result.

Table 32: Leave-one-out jackknife sensitivity of mechanism estimates (Model 3).

Predictor	Baseline OR	Baseline p
Non-regular	6.41	0.019
Small firm	4.14	0.025

Predictor	Jackknife OR range	Jackknife p range	Share with $p < 0.05$
Non-regular	[5.69, 12.82]	[0.016, 0.030]	1.000
Small firm	[3.54, 5.52]	[0.014, 0.052]	0.989

Notes: Leave-one-out refits of Model 3; ranges show min–max across jackknife runs.

D.9 Mechanism check at childbirth margin ($t = 0$)

To align mechanism evidence with the childbirth margin, I estimate a parallel model where the outcome is non-employment at $t = 0$ among women employed at $t = -1$, with predictors measured at $t = -1$. This specification targets the childcare-dominant exit margin directly and is interpreted as predictive association evidence.

Table 33: Mechanism check at childbirth margin: predictors of non-employment at $t = 0$ among women employed at $t = -1$.

Predictor	OR (logit)	p (logit)
Non-regular (t=-1)	7.50	0.000
Small firm (t=-1)	1.55	0.126

Table 34: Absolute-risk view of childbirth-margin exit: raw rates and model-implied probabilities at $t = 0$.

Profile / group	N	Exits	Exit rate
By contract type: Non-Regular	120	73	0.608
By contract type: Regular	162	27	0.167
By firm size: Large/Gov	175	53	0.303
By firm size: Small	107	47	0.439
By contract x firm: Non-Regular x Large/Gov	66	38	0.576
By contract x firm: Non-Regular x Small	54	35	0.648
By contract x firm: Regular x Large/Gov	109	15	0.138
By contract x firm: Regular x Small	53	12	0.226
Profile			Predicted Pr(exit at $t = 0$)
Regular x Large/Gov			0.146
Regular x Small			0.209
Non-Regular x Large/Gov			0.562
Non-Regular x Small			0.665

To increase visibility on covariate missingness, I additionally report a missingness profile for key mechanism covariates (contract type and firm size), a direct missingness-exit association check, and side-by-side coefficient sensitivity across complete-case, missing-indicator, and high-information samples.

Table 35: Missingness profile for mechanism covariates in the childbirth-margin risk set (employed at $t = -1$, observed at $t = 0$).

Covariate missingness	N	Share of risk set	Exit rate at $t = 0$
Contract type missing	36	0.109	0.444
Firm size missing	16	0.048	0.750
Either missing	48	0.145	0.500

Notes: The risk-set size is $N = 330$ women. Exit is defined as non-employment at $t = 0$ among women employed at $t = -1$.

Table 36: Association between covariate missingness and childbirth-margin exit.

Missingness indicator	OR	p-value
Contract type missing	1.243	0.554
Firm size missing	5.254	0.005

Notes: Logit model with exit at $t = 0$ as the dependent variable. Odds ratios above 1 indicate higher exit odds for observations with missing covariates.

Table 37: Childbirth-margin mechanism sensitivity to missing-data handling.

Specification	N	Exits	OR Non-regular	p	OR Small firm	p
Complete-case	282	100	7.505	0.000	1.548	0.126
Missing-indicator	330	124	7.476	0.000	1.289	0.345
High-information	237	83	7.393	0.000	1.860	0.047

Notes: The high-information sample requires non-missing contract type, firm size, commute indicator, and husband's overwork indicator at $t = -1$. The missing-indicator specification includes missing-value dummies for contract type and firm size.

These diagnostics clarify the missingness pattern in the childbirth-margin risk set. Covariate missingness is non-trivial (14.5% missing on either contract type or firm size). Contract-type missingness is not strongly associated with exit (OR = 1.24, $p = 0.554$), while firm-size missingness is strongly associated with exit (OR = 5.25, $p = 0.005$; exit rate 75% among missing). The key non-regular coefficient remains essentially unchanged across complete-case, missing-indicator, and high-information specifications, while small-firm estimates vary in precision. I therefore treat the contract-tier gradient as robust and the firm-size gradient as supportive but sample-sensitive.

A natural follow-up concern is whether contract-type missingness is itself systematically related to baseline characteristics. If women with missing contract information differ on observable dimensions, the complete-case estimates could reflect selection into observability rather than a genuine labour-market gradient. I therefore report a balance diagnostic comparing baseline hours, income, and survey source between women with observed and missing contract type at $t = -1$ (Table 38), together with a logit for the probability of contract-type observability.

Table 38: Balance diagnostic: baseline characteristics by contract-type observability at $t = -1$ in the childbirth-margin risk set ($N = 330$).

Panel A: Means by observability group

Variable	Observed ($N = 294$)	Missing ($N = 36$)	p-value
Exit at t=0	0.367	0.444	0.389
Hours/week (at t=-1)	35.626	21.472	<0.001
Annual income (man-yen, at t=-1)	280.918	137.750	0.002
JHPS share	0.449	0.389	0.612

Panel B: Logit for 1{contract type observed}

Predictor	OR	SE (log-odds)	p-value
Intercept	1.234	0.379	0.579
Hours/week at $t = -1$	1.037	0.014	0.009
Annual income (man-yen) at $t = -1$	1.004	0.002	0.016
JHPS indicator	1.146	0.384	0.722

McFadden pseudo-$R^2 = 0.1364$ *Notes:* Risk set is women employed at $t = -1$ and observed at $t = 0$ ($N = 330$). Panel A reports means and two-sample t-test p-values (proportion test for JHPS share). Panel B reports logit odds ratios for the probability that contract type is non-missing at $t = -1$. Low pseudo-R^2 values indicate missingness is approximately conditionally random.

The balance diagnostic reveals that contract-type missingness is not random: women with missing contract type at $t = -1$ work significantly fewer hours (21.5 vs 35.6 hours/week, $p < 0.001$) and earn significantly less (137.8 vs 280.9 man-yen, $p = 0.002$) than women with observed contract type. The logit for observability confirms this (McFadden pseudo-$R^2 = 0.14$). However, this pattern is consistent with contract-type missingness being concentrated in marginal employment where formal classification is ambiguous or unreported. Because these women resemble the non-regular profile on observable dimensions, the complete-case non-regular gradient is likely conservative rather than inflated by selection into observability. Crucially, the non-regular OR is virtually identical across complete-case (7.50), missing-indicator (7.48), and high-information (7.39) specifications, confirming that the core stratification result is not driven by which women have observed contract type. The exit-rate difference between groups (44.4% vs 36.7%) is not statistically significant ($p = 0.389$), and survey source (JHPS vs KHPS) does not predict observability ($p = 0.722$).

A coding concern is that some women on leave may report zero market hours at $t = 0$ and therefore be classified as non-employed. To assess this, I run a leave-recoding sensitivity: women coded as non-working at $t = 0$ are reclassified as employed if they report leave take-up near birth. The adjusted estimates attenuate the childbirth coefficient but preserve the core result: the birth-year break remains large, negative, and statistically significant. I therefore treat leave-related coding as affecting level magnitude, not the qualitative timing pattern.

Table 39: Leave-as-employment sensitivity at childbirth ($t = 0$).

Specification	Coef. at $t = 0$	SE	p-value	Reclassified at $t = 0$
Original (hours > 0)	-0.1319	0.0275	<0.001	0
Leave-adjusted	-0.0988	0.0276	<0.001	22

Notes: Sensitivity recodes women as employed at $t = 0$ if they are coded non-working but report leave take-up near birth. The coefficient remains strongly negative after reclassification.

D.10 Policy-period cohort split

The observation window (2004-2022) spans substantial policy changes in Japan's work - family infrastructure, including the 2010 introduction of paternity leave, the 2012 revised Child Care and Family Care Leave Act, the acceleration of childcare expansion under the Abe government's "Womenomics" agenda from 2013, and the 2019 free early childhood education reform. If these reforms materially altered the motherhood employment penalty, pooling across the full period could obscure regime-specific effects.

To assess sensitivity, I split the sample by birth-report year: an early cohort (first birth reported in 2004-2012, before the acceleration of Womenomics-era reforms) and a late cohort (2013-2022). I additionally re-estimate the baseline specification excluding pandemic-era births (2020-2022) to rule out COVID-related distortions.

Table 40 reports the key event-study coefficients ($t = -2$, $t = 0$, $t = +5$) for each subsample. The results are reassuring. The near-lead coefficient ($t = -2$) is statistically indistinguishable from zero in all four samples, confirming that the pre-trend null holds within each cohort. The childbirth-year break ($t = 0$) is remarkably stable: -0.131 ($p < 0.001$) in the early cohort and -0.135 ($p = 0.002$) in the late cohort, with the no-COVID estimate (-0.145, $p < 0.001$) nearly identical to the pooled baseline. The long-horizon coefficient ($t = +5$) is small and insignificant in all splits.

Two implications follow. First, the magnitude of the motherhood employment penalty at birth has not detectably changed across the pre-2013 and post-2013 policy regimes in this sample, despite substantial expansions in childcare capacity and leave generosity over the period. This is consistent with the interpretation that structural labour-market segmentation - rather than policy generosity per se - is the binding constraint on employment continuity through childbirth. Second, excluding pandemic-era births does not alter the core estimates, indicating that COVID-related labour market disruptions do not drive the pooled results.

Table 40: Cohort-split sensitivity: key employment event-study coefficients by birth-year period (year fixed effects, heteroskedasticity-robust standard errors).

Sample	Event time	Estimate	SE	p-value	N obs	N women
Pooled	$t=-2$	-0.012	0.031	0.703	5,212	662
	$t=+0$	-0.132	0.027	<0.001	5,212	662
	$t=+5$	-0.028	0.032	0.384	5,212	662
Early (≤2012)	$t=-2$	0.020	0.041	0.627	3,062	390
	$t=+0$	-0.131	0.036	<0.001	3,062	390
	$t=+5$	-0.020	0.045	0.658	3,062	390
Late (≥2013)	$t=-2$	-0.063	0.047	0.186	2,150	272
	$t=+0$	-0.135	0.044	0.002	2,150	272
	$t=+5$	-0.025	0.056	0.657	2,150	272
No-COVID (≤2019)	$t=-2$	-0.004	0.032	0.906	4,822	577
	$t=+0$	-0.145	0.029	<0.001	4,822	577
	$t=+5$	-0.028	0.034	0.399	4,822	577

Notes: Each panel re-estimates the baseline employment event study (is_working $\sim i(\text{event\_time}, \text{ref}=-1) \mid \text{calendar\_year}$) on the indicated subsample. "Early" includes mothers whose first birth was reported in 2004–2012; "Late" in 2013–2022. "No-COVID" excludes birth years 2020–2022. All specifications use year fixed effects and heteroskedasticity-robust standard errors.

The same threshold pattern appears separately at $t=+3$, $t=+4$, and $t=+5$ (Figure 19). As noted in the main text, childless married employed women show at least as much concentration below these thresholds (49.7% at or below 103 man-yen, 60.0% at or below 130 man-yen), confirming that the bunching reflects secondary-earner constraints rather than a motherhood-specific behavioural response.

D.11 Industry-sector diagnostic

A natural concern is whether the motherhood employment penalty varies by pre-birth industry sector, and specifically whether women in sectors with stronger institutional protections (e.g., government, finance/insurance, utilities, and education) experience smaller employment breaks than those in other sectors.

To investigate, I assign each woman to one of three groups based on her last observed industry in the pre-birth window ($t=-5$ to $t=-1$): "stable sector" (government, finance/insurance, utilities, education), "other sector" (all remaining industries), and "unknown / no pre-birth industry" (women with no industry data in the pre-birth window, predominantly not employed before birth). This assignment avoids conditioning on employment at any single event time, which would create a mechanical negative pre-trend.

Figure 19: Earnings distribution by event time for employed mothers ($t = +3, +4, +5$), with reference lines at 103 and 130 man-yen. The threshold concentration pattern is visible in each post-birth horizon. Comparable bunching is observed among childless married employed women, indicating that this is a secondary-earner phenomenon rather than a motherhood-specific response.

I report two diagnostics. First, a full-sample event-study stratification retains all 662 mothers across the three groups (Table **??**). The "stable sector" ($N = 86$) and "other sector" ($N = 305$) subsamples both show significant employment declines at $t = 0$, but both also exhibit significant negative coefficients at $t = -2$, indicating that sector assignment based on pre-birth employment history induces composition differences that appear as pre-trends. The "unknown" group ($N = 271$) shows positive coefficients at $t = 0$ and $t = +5$, reflecting selection: these women were predominantly not employed before birth, and some enter employment over the panel window. These patterns confirm that sector-stratified event-study *levels* in this sample are not cleanly interpretable as sector-specific penalties.

Second, to avoid these composition artefacts, I estimate a conditional transition diagnostic restricted to the 328 women employed at $t = -1$ with an observed sector assignment (Table 42). Exit rates at birth are similar across sectors: 37.1% in the other-sector group versus 39.0% in the stable-sector group, a difference of 1.9 percentage points ($p = 0.764$). Employment rates at $t = +5$ are 67.5% in the other-sector group versus 57.4% in the stable-sector group, a difference of -10.1 percentage points ($p = 0.192$), suggestive of weaker long-run recovery in the stable sector but statistically imprecise.

Two implications follow. First, among women employed before birth, pre-birth sector type does not significantly predict exit probability at childbirth in this sample. This is consistent with the interpretation that the employment break at birth operates primarily through job-type segmentation (regular versus non-regular status) rather than through industry affiliation per se. Second, the imprecise long-horizon difference leaves open the possibility that stable-sector women who exit face different return barriers, though the small cell sizes ($N = 54$ and $N = 160$ at $t = +5$) preclude firm conclusions.

D.12 Standard-error sensitivity (heteroskedastic vs clustered)

The table 43 compares key employment event-study coefficients using heteroskedasticity-robust and woman-level clustered standard errors. The point estimates are unchanged by

Table 41: Industry stratification (revised): key employment event-study coefficients by pre-birth sector assignment.

Sample	Event time	Estimate	SE	p-value	N obs	N women
All women	$t=-2$	-0.012	0.031	0.703	5,212	662
	$t=+0$	-0.132	0.027	<0.001	5,212	662
	$t=+5$	-0.028	0.032	0.384	5,212	662
Stable sector	$t=-2$	-0.219	0.072	0.003	682	86
	$t=+0$	-0.316	0.067	<0.001	682	86
	$t=+5$	-0.357	0.075	<0.001	682	86
Other sector	$t=-2$	-0.083	0.036	0.020	2,492	305
	$t=+0$	-0.291	0.036	<0.001	2,492	305
	$t=+5$	-0.199	0.042	<0.001	2,492	305
Unknown / no pre-birth industry	$t=-2$	0.003	0.010	0.789	2,038	271
	$t=+0$	0.139	0.023	<0.001	2,038	271
	$t=+5$	0.357	0.038	<0.001	2,038	271

Notes: Sector is assigned using the most recent observed pre-birth industry in event times $[-5,-1]$ among years with employment observed. "Stable sector" includes government, finance/insurance, utilities, and education. Women with no observed pre-birth industry are retained as a separate "Unknown" group. This is a descriptive sensitivity check.

Table 42: Industry and birth-transition outcomes among women employed at $t=-1$.

Sector group	N women	Exit rate at $t=0$	N at $t=+5$	Employed rate at $t=+5$
Other sector	251	0.371	160	0.675
Stable sector	77	0.390	54	0.574

Difference (Stable − Other), linear probability model with heteroskedasticity-robust SEs:
Exit at t=0 (1=not working): 0.019 (SE 0.064, p=0.764, N=328)
Employed at t=+5 (1=working): -0.101 (SE 0.077, p=0.192, N=214)

Notes: Sector is assigned using the most recent observed pre-birth industry in event times $[-5,-1]$ among employed observations. "Stable sector" includes government, finance/insurance, utilities, and education. This table is a conditional transition diagnostic and not a full-sample sector event study.

construction, and inference is substantively identical across SE choices: the near lead at $t = -2$ remains null, the birth-year drop at $t = 0$ remains strongly significant, and the long-horizon coefficient at $t = +5$ remains statistically indistinguishable from zero.

Table 43: Sensitivity of key employment event-study coefficients to standard-error choice (year fixed effects).

	Heteroskedasticity-robust			Woman-clustered		
Event time	Estimate	SE	p	Estimate	SE	p
$t = -2$	-0.012	0.031	0.703	-0.012	0.024	0.620
$t = +0$	-0.132	0.027	< 0.001	-0.132	0.020	< 0.001
$t = +5$	-0.028	0.032	0.384	-0.028	0.032	0.387

D.13 Pre-pregnancy job-type stability ($t = -5$ to $t = -3$)

To address the concern that contract type in the pre-birth mechanism models may reflect immediate anticipatory switching, I report a stability diagnostic in the pre-pregnancy window. Specifically, I track broad job type (Regular vs Non-Regular) from $t = -5$ to $t = -3$ on common support and report persistence rates and transition counts. This is a descriptive check rather than an identification strategy: it does not remove long-run sorting into job types, but it provides suggestive evidence against abrupt compositional reclassification immediately before the mechanism margin. In the common-support sample observed at both $t = -5$ and $t = -3$ ($N = 32$), overall job-type stability is 81.2% (Regular→Regular: 77.8%; Non-Regular→Non-Regular: 85.7%). Some switching remains, but the pattern is more consistent with persistent segmentation than with abrupt pre-birth reclassification.

Table 44: Pre-birth job-type stability check ($t = -5$ to $t = -3$).

Metric	Value	Interpretation	Window
Women observed at both $t = -5$ and $t = -3$	32	Common-support sample	Pre-pregnancy
Overall job-type stability	81.2%	Same broad job type at $t = -5$ and $t = -3$	Pre-pregnancy
Regular → Regular	77.8%	Persistence among regular workers	Pre-pregnancy
Non-Regular → Non-Regular	85.7%	Persistence among non-regular workers	Pre-pregnancy

D.14 Core robustness summary and denominator map

To make inference priorities and sample definitions fully transparent, I report (i) one consolidated robustness table for the key employment coefficients at $t = -2$, $t = 0$, and $t = +5$, and (ii) a denominator map for the main samples used in the paper. The robustness table focuses on the same estimand across alternative sample/weighting/SE choices; the denominator map is a reading guide so each reported percentage is tied to the correct risk set.

Table 45: Master robustness table: key employment coefficients at $t = -2$, $t = 0$, and $t = +5$.

Specification	$t = -2$	$t = 0$	$t = +5$
Balanced short window ($t = -2$ to $t = +1$)	-0.058 (0.065)	-0.128 (< 0.001)	–
Baseline (year-FE robust)	-0.012 (0.702)	-0.132 (< 0.001)	-0.028 (0.384)
Clustered SE (woman)	-0.012 (0.620)	-0.132 (< 0.001)	-0.028 (0.387)
Early cohort (≤ 2012)	0.020 (0.627)	-0.131 (< 0.001)	-0.020 (0.658)
IPW (obs. at $t = +5$)	-0.024 (0.436)	-0.136 (< 0.001)	-0.057 (0.076)
JHPS only	-0.027 (0.576)	-0.136 (0.002)	-0.126 (0.027)
KHPS only	0.002 (0.966)	-0.134 (< 0.001)	0.010 (0.796)
Late cohort (≥ 2013)	-0.063 (0.186)	-0.135 (0.002)	-0.025 (0.657)
Leave-adjusted coding	–	-0.099 (< 0.001)	–
No-COVID sample	-0.004 (0.906)	-0.145 (< 0.001)	-0.028 (0.399)
Second-birth censored	-0.011 (0.724)	-0.135 (< 0.001)	-0.003 (0.935)

Notes: Cells report the coefficient with the p-value in parentheses. Leave-adjusted coding is defined for the childbirth margin ($t = 0$) only.

Table 46: Denominator map for main reported quantities

Quantity	Denominator	Definition
Main employment event study	662 women	Analytic first-birth sample in the event-study panel
Childbirth-margin mechanism ($t = 0$)	330 women	Women employed at $t = -1$ and observed at $t = 0$
Pre-birth mechanism ($t = -2$)	185 women	Women employed at $t = -2$ and observed at $t = -1$
Father leave (partner-linked analytic)	661 partners	Linked partner records in the analytic panel
Leave take-up tabulation (broader module)	1,183 couples	First-birth-linked leave-analysis module sample
Newlyweds education coverage	89 women	Mothers with non-missing Newlyweds-module education

Notes: Denominators differ by design because each statistic is defined on its own valid risk set or module coverage. Reported percentages should be interpreted within row, not compared mechanically across rows.

Appendix E. Extended institutional context

This appendix provides additional background on the institutional environment that shapes Japanese mothers' employment trajectories.

E.1 Dual labour market structure

Japan's labour market is divided between regular (*seishain*) and non-regular (*hi-seiki*) employment. Regular workers receive open-ended contracts, seniority-based wages, firm-specific training, and strong de facto job protection; non-regular workers are typically on fixed-term, part-time (*pāto*), or dispatch contracts with flat wage profiles and limited access to internal labour markets (Brinton, 1993; Yamaguchi, 2019). Note that the contract category "part-time" (*pāto*) does not necessarily imply fewer working hours: in this sample, 36 percent of non-regular workers work full-time hours. This segmentation is reinforced by firm size: large firms tend to provide more stable career ladders and formal leave systems, while small and medium-sized firms have thinner human resource capacity and less formalized leave management. The consequence is that employment continuity after childbirth depends on both individual preferences and on job status and firm-level institutional capacity.

E.2 Work-family policy and eligibility

The Childcare Leave and Family Care Leave Act grants leave for workers caring for a child under age one, with extensions to 18 or 24 months when childcare placement is unavailable. Eligibility for fixed-term employees requires at least one year of continuous employment and a reasonable expectation that the employment contract will continue beyond the child's first birthday (Ministry of Health, Labour and Welfare, 2019). This condition effectively excludes many fixed-term and short-hours workers, particularly those in non-regular jobs. Even when statutory rights exist, take-up depends on continued employment relationships and on firm-level willingness to accommodate long leave spells.

E.3 Long-hours norms and gendered care

Workplace norms in Japan remain characterized by long hours and a strong breadwinner model, especially in regular employment tracks. This norm constrains dual-earner equilibria even when formal leave is available (Nemoto, 2016; North, 2009). The persistent asymmetry in childcare time use in the data (wife-to-husband ratios around 5-to-1 after birth) is consistent with this institutional and cultural environment.

E.4 Fiscal incentives and secondary earners

Japan's tax and social-insurance system creates multiple earnings thresholds for secondary earners. OECD (2024) documents that these thresholds generate high effective marginal tax rates and encourage married women to cap earnings below the contribution cutoffs, often by selecting part-time or fixed-term jobs (OECD, 2024a). In this sample, earnings bunching at the 103 and 130 man-yen thresholds is at least as pronounced among childless married employed women (49.7% and 60.0% respectively) as among returning mothers (47.8% and 55.9%), suggesting that threshold responses are a pre-existing feature of secondary-earner labour supply rather than a behaviour unique to motherhood. These incentives interact with the dual labour market: women in non-regular jobs face both weaker job protection and stronger financial incentives to remain below thresholds, and motherhood pushes women into the reduced-hours earnings range where these constraints bind.

E.5 Implications for identification

These institutional features are important for interpreting the event-study results. The sharp exit at birth and the durable shift into reduced-hours work - often on non-regular contracts - are consistent with a system in which job status and firm-level constraints shape who can remain employed. The mechanisms do not require strong changes in preferences; instead, they reflect the structure of available jobs and the rules governing leave eligibility and earnings incentives.