Female Jobs and Fertility: Long-term Experimental Evidence from Ethiopian Factories*

Andreas Kotsadam[1], Janneke Pieters[2], Pauline Rossi[3], Espen Villanger[4], Céline Zipfel[5], and Tigabu Getahun[6]

[1]Ragnar Frisch Centre for Economic Research
[2]Wageningen University
[3]École Polytechnique-CREST
[4]Chr. Michelsen Institute
[5]Stockholm School of Economics
[6]EconInsight Center for Development Research

July 27, 2026

Abstract

We conducted an experiment with 27 factories in Ethiopia to test the widespread hypothesis that formal employment among women reduces fertility. Tracking respondents over eight years, we find that women who were randomly offered a job had *more* children than the control group. The effect on fertility emerged once the effect on wage employment faded away, around three years after baseline. There are clear effects on household income and savings, but not on bargaining power and marriage. These patterns are consistent with a Beckerian model in which the income effect dominates the substitution effect and households are credit constrained.

Keywords: Female employment; fertility; industrialization; sub-Saharan Africa
JEL Codes: J13, J16, J22, O15, O55.

*We are grateful to Alicia Adsera for valuable comments and suggestions, and to participants at seminars and conferences at Paris School of Economics, Paris Dauphine University, ULB-ECARES, ISER at Osaka University, The Rockwool foundation, WZB Berlin, Kristiania University of Applied Sciences, Norwegian School of Economics, Stockholm School of Economics, the University of Bordeaux, the Nordic Conference in Development Economics (NCDE) 2025, the European Development Research Network (EUDN) Conference 2026, the Jobs Workshop Bergen 2026, the Ragnar Frisch Centre for Economic Research, the Bristol Applied Economics Meeting 2026, and the Barcelona Summer Forum 2026 for helpful feedback. This research was supported by the Gender, Growth and Labour Markets in Low-Income Countries (G^2LM|LIC) programme. All errors are our own. The pre-analysis plan was registered at the AEA RCT Registry (AEARCTR-0014209): https://www.socialscienceregistry.org/versions/234939/docs/version/document.

1 Introduction

A long-standing consensus in economics is that expanding women's work opportunities reduces fertility. Correlations support this view: across countries and over time, fertility declines have tended to coincide with rising female labor force participation and wages (Schultz 1985; Goldin 1995; Mammen and Paxson 2000). Theoretical work interpreted these patterns through a change in women's opportunity cost of time (Galor and Weil 1996).

Whether this view holds in sub-Saharan Africa (SSA) is an open question (Van den Broeck et al. 2023; Zipfel 2025). On the one hand, high fertility and high female labor activity have long co-existed in the region, challenging the idea of a trade-off between productive and reproductive activities. On the other hand, women mostly engage in flexible self-employment or unpaid family work while formal wage jobs—especially in manufacturing—remain scarce (McMillan and Zeufack 2022; Bandiera et al. 2022). As such jobs expand, they may tighten the constraints women face in combining work and family. How will fertility choices adjust?

We study this question in the context of Ethiopia for two reasons. First, Ethiopia is the tenth most populated country in the world and it grows very quickly: the population is projected to roughly triple by 2100 up to more than 350 million people. Alone, the country is predicted to account for approximately 10% of the global population growth (United Nations 2019). Understanding the drivers of fertility in Ethiopia is therefore important. Second, the country is a forerunner in the industrialization of SSA. Since the mid-2000s, the government has set up various special economic zones, called industrial parks, to expand industrial production. A large fraction of workers hired in these factories are women, making Ethiopia the ideal setting to study fertility responses.

We cooperated with 27 manufacturing firms located in five industrial parks to conduct a randomized controlled trial (RCT) in 2016. Our sample consists of 1464 partnered women of reproductive age who applied for factory jobs. Eligible applicants were first screened by firms, and those meeting hiring criteria were placed into randomization lists specific to each factory and recruitment round. Within these lists, 50% of women were randomly selected to receive an offer of formal employment, while the other half served as controls. The jobs offered were typical of Ethiopia's expanding manufacturing sector—low-skilled, repetitive production positions requiring long hours and limited flexibility in exchange for a regular wage.

We followed respondents over eight to nine years, with several rounds of midline surveys, to estimate the long-term treatment effect on fertility. We registered a pre-analysis plan before adding a fertility module and collecting the latest wave of data in 2025. To minimize attrition at endline, we implemented an intensive tracking protocol involving repeated call-backs, tracing through family networks, and in-person follow-ups outside of respondents' original locations of residence. Overall, attrition at endline was 23%, with no differential attrition by treatment status.

We find that, over the study horizon, treated women worked about 8.3 additional months in formal wage jobs and earned more than control women. These average employment effects are large and precisely estimated; however, they are concentrated in the first three years and fade over time as many treated women exited factory jobs. Turning to fertility, we estimate a 5% increase in the number of births among treated women relative to the control group. The rate of childlessness is reduced by almost half: in the control group, 10% of women remained childless by endline, compared to less than 6% in the treatment group. The fertility effects are driven by those who had no child at baseline, suggesting that the job offer affected the timing of entry into motherhood rather than birth spacing among mothers.

To further examine the timing of fertility responses, we collected detailed fertility histories spanning the entire post-treatment period. Because we also observe when women enter and exit factory employment, we can link birth timing to employment trajectories and thereby examine the sequencing of work and childbearing. We find that the fertility response emerges gradually, once the treatment and control groups have largely converged in employment rates. These results highlight that the work-family trade-off is not a static choice; when jobs are rigid, employment and childbearing may occur sequentially for some women.

We use a surrogate analysis to predict how much of the treatment effect is likely to persist once women have completed their fertility, at around age 45. At endline, most women are in their thirties; we can observe the treatment effect on fertility by age. We estimate the relationship between fertility by age and completed fertility in another dataset and we use this estimation to predict completed fertility in our experimental sample. The extrapolation suggests that most of the treatment effect on childlessness is likely to disappear because control women are expected to catch up their first births; however, most of the treatment effect on number of births is likely to persist because treated women's earlier start is unlikely to be offset by fewer future births.

The positive fertility effect is at odds with conventional wisdom and at odds with the results we expected in the pre-analysis plan. This is because we assumed that the opportunity cost of time would drive fertility down. In fact, in a standard Beckerian framework, income effects can counteract substitution effects, making the fertility response theoretically ambiguous (Becker and Lewis 1973; Jones et al. 2011). We argue that a simple dynamic extension assuming that households engage in consumption smoothing under credit constraints can explain the timing of the effects: treated women have their first child earlier than control women but only once they have saved enough. We further provide empirical evidence supporting the income effect. We show that household income and savings are substantially higher in the treatment group and that the positive fertility response is concentrated among poorer women at baseline and in firms that pay a higher wage.

The data allow us to examine several alternative explanations. First, stated fertility preferences, such as ideal number of children and ideal age at first birth, are unaffected. This is consistent with a channel operating through a change in constraints rather than a change in preferences. Second, we find no evidence that changes in relative earnings altered bargaining power over fertility decisions nor intimate partner violence, and no support for the hypothesis of a backlash against women becoming financially independent. Third, we find that treated and control women are equally likely to be married at endline, to have stayed with their baseline partner, and to have ever migrated. Finally, we find no effect on aspirations for their children, maternal health nor child survival.

Our main contribution to the literature is to provide the first individual-level experimental evidence of the fertility impacts of female wage employment. The female opportunity cost of time argument plays a central role in the economics of fertility; this is "the second big idea" (after the quality-quantity trade-off) in traditional models of fertility choice according to Doepke et al. (2023). Yet, there is surprisingly little causal evidence on how access to new work opportunities shapes women's fertility trajectories over the life cycle. Prior work has focused on unmarried young women in low female labor force participation contexts, and relied either on clustered designs or observational variation. Jensen (2012) provides the only randomized evidence to date, based on a village-level expansion of job search assistance in India, which delayed marriage and first birth while lowering desired fertility and raising career aspirations; these effects are accompanied by increased educational attainment. In Bangladesh, Heath and Mobarak (2015) use non-experimental variation from the expansion of garment factories to show similar postponement of marriage and improved school enrollment for teenagers. Our experiment differs on two important dimensions. First, our sample includes relatively older women, aged between 20 and 32 at baseline, living with a partner. The channels of schooling and union formation are therefore not at play. Second, we estimate the individual effect of being offered a job among active job-seekers. This is different from estimating the macro effect of formal jobs expansion, which can change prices and expectations for all women, including those who do not work. By documenting a very strong first stage on employment and following women for a substantial share of their reproductive years, we provide a direct test of the opportunity cost of time channel. Our results suggest either that (i) the negative relationship between women's work opportunities and fertility is mostly mediated by other channels, or that (ii) in our context specifically, the income effect is large enough to dominate the substitution effect and the fertility response turns positive.

By discussing why the income effect dominates in Ethiopia, our study relates to the literature on the so-called "African exceptionalism" in demography (Bongaarts 2017). The fertility transition is particularly late and particularly slow in the region, and the response to usual drivers of fertility decline, such as education or contraception, is weaker than in other contexts, presumably due to cultural and institutional barriers (see references in Gobbi et al. (2026)). On fertility responses to female economic empowerment specifically, an emerging experimental literature has started to document positive effects in some contexts. In particular, interventions targeting married women engaged in self-employment, such as business training that raised entrepreneurial profits in Ethiopia and Togo, increased fertility at low levels of baseline income and parity, but had no effect among richer, older women (Donald et al. 2024).[1] These results are in line with ours. Our contribution is to study formal wage employment instead of informal self-employment. We show that, even when work is harder to combine with childcare, the fertility response is positive. This is partly because the combination can be sequential, some women quitting their jobs when they have saved enough to afford a child, and partly because the maternal time cost is low for other women, because they benefit from extensive informal childcare.

In terms of methodology, we add to the small set of long-term RCT follow-ups (see references in Bouguen et al. (2019)) by tracking 77% of the original sample eight to nine years after baseline. Long-term evaluations are particularly important for fertility, where short-run effects need not predict

[1] Interventions targeting adolescent girls and ultimately increasing women's incomes have delivered more mixed findings. Multi-faceted empowerment programs and secondary school scholarships, have been found to have either a negative effect (Bandiera et al. (2020) in Uganda, Duflo et al. (2024) in Ghana) or a positive effect (Berge et al. (2026) in Tanzania) on early fertility.

longer-run outcomes. For example, Duflo et al. (2024) show that secondary school scholarships in Ghana delayed fertility onset, yet fertility rates largely converged over time. In our setting, the fertility effects emerge only after three years, making this one of the few randomized evaluations able to capture the sequencing of work and childbearing.

Finally, this paper relates to the nascent experimental literature on the welfare effects of getting a job in Ethiopia's industrial parks. Blattman and Dercon (2018) find that industrial job offers did not sustainably raise wages or employment and came with meaningful health costs. Abebe et al. (2025) find positive short-run effects of a job-facilitation intervention on employment, earnings, and savings, but little evidence of persistent gains four years later. Studies based on earlier waves of our own RCT show no effect on physical intimate partner violence (Kotsadam and Villanger 2025) and reduced civic engagement despite income gains (Aalen et al. 2024). Fertility has not been examined so far, and our results suggest it is a margin on which the effects are both substantial and slow to materialize.

2 Setting and Field Experiment

By international standards, Ethiopia is a high-fertility country in which women marry early and, despite high overall labor force participation, rarely hold formal wage jobs. At the time of our baseline, the 2016 Ethiopia Demographic and Health Survey (EDHS) estimated a total fertility rate of 4.6 children per woman (Ethiopia Central Statistical Agency and ICF 2017). Only about one quarter of women aged 15–49 reported being paid in cash for their work. 77% of women aged 15 and older were economically active, but most worked in subsistence agriculture, family enterprises, or other informal activities (World Bank 2016).

The study took place during a period of rapid transition for the country, both in terms of fertility and industrialization. Fertility rates had been declining quickly, down from 6.7 children in 2000; meanwhile, female wage employment had grown up from below 10% of women aged 15–49 in 2007 (Jobs of the World Project 2026). The growth of the manufacturing sector was supported by government efforts to attract foreign investment. One component of this strategy was the establishment of industrial parks designed to host export-oriented manufacturing firms across the country.

We collaborated with 27 shoe and garment factories located in industrial parks across five regions of Ethiopia (Figure A.1): Tigray, Amhara, Oromia, Dire Dawa and SNNPR (Southern Nations, Nationalities and Peoples). All of these parks were located outside of the capital, Addis Ababa. The factories were a mix of domestic and foreign-owned firms and employed predominantly female workforces, with women accounting for roughly 72% of workers on average.

The initial data collection was conducted as part of a study on intimate partner violence (Kotsadam and Villanger 2025), and the sample was therefore restricted to women living with a partner at the time of recruitment (not necessarily married but in a stable relationship). The factories' standard procedure of hiring was to advertise bulks of positions by posting on the front gate, by word of mouth, and on local job boards. The applicants were asked to gather on a specific day and were screened for eligibility using verbal and physical tests. The factories we collaborated with accepted to slightly alter this process. They first assessed all job applicants and determined whether each applicant was eligible for the job or not. Then, from the pool of eligible applicants, we created lists of women living with a partner. These lists form our baseline sample. we randomly assigned half of the women to the treatment group, in which they all received a job offer in the given factory, and the remaining women to a control group, in which no one received a job offer. Randomization was feasible since the number of qualified applicants far exceeded the number of available jobs and the experiment did not change the number of jobs offered to women. The applicants on the lists were informed about the procedure before the randomization was conducted, and they perceived it as fair that all equally qualified candidates had an equal chance of receiving an offer. We regard the randomization as ethically justified for the same reason.[2]

For many applicants, these jobs represented a rare opportunity to enter formal wage employment. Workers received regular wages and worked in large, formal firms, but jobs were also characterized by long hours, repetitive tasks, limited opportunities for advancement, and fixed schedules. At the time of the first follow-up survey, workers earned an average of 1,021 Ethiopian birrs (ETB) per month (approximately $47 at the 2016 exchange rate, or $118 in PPP terms), worked at least eight hours per day and six days per week, and typically received only task-specific on-the-job training.

Qualitative interviews conducted with 24 women in the experiment and 11 of their partners across the five study regions provide additional insight into how workers experienced these jobs (Halvorsen 2021). The interviews suggest that many women entered factory employment with ex-

[2]Note that the procedure eliminated the scope for hiring based on good looks, corruption, and sexual extortion, which we heard stories about during our field visits. See Aalen et al. (2024) for a detailed discussion of ethical considerations and principles.

pectations that were not fully met, particularly regarding earnings and working conditions. The average woman expected to earn 40% more than her actual salary, and several interviewees described accepting positions with limited prior knowledge of the work itself. At the same time, the interviews reveal that applicants viewed factory jobs as valuable opportunities in a context where formal employment options were scarce. Many remained employed despite dissatisfaction with wages or conditions, often expressing hope that the jobs would provide future opportunities or greater economic security.

The jobs can be combined with family responsibilities, but not for all women. Factories offer paid maternity leave. When asked whether they would be able to continue working if they had a newborn child, 56% of women in our sample responded that they would not, 9% stated that they could do so part-time and 35% answered they could work full-time. More broadly, 60% described combining factory work with household duties as very difficult and another 23% as difficult.

3 Data and Empirical Strategy

3.1 Survey data

We surveyed women in 2016-2017 to collect baseline data (Wave 1) before they started working. The dates for the baseline data collection vary depending on when the firms hired. Some of the firms hired new batches of workers several times during the period. We collected the first follow-up data (Wave 2) on average 8 months after the first interview, and subsequent follow-up data at roughly 15, 22, 35 and 48 months after baseline (Waves 3-6). Our endline data (Wave 7) was collected in 2025, eight to nine years after baseline. We show the number of interviews over time by wave in Figures A.2 and A.3. Out of the 1464 women in our baseline sample, we managed to interview 1262 in the second wave and 976 in the sixth wave. Our intense tracking efforts at endline paid off: we managed to interview 1130 women corresponding to 77% of the baseline sample.[3]

The endline data was collected specifically to study long-term effects on fertility. The survey includes a detailed birth history module collecting dates of all live births. To elicit fertility preferences, we use the standard DHS question on the ideal number of children; we also ask about the appropriate age at first birth for a woman. In addition to fertility outcomes, we record complete em-

[3] Kotsadam and Villanger (2025) used Wave 4 to study impacts on intimate partner violence and Aalen et al. (2024) used Wave 5 to study impacts on civic engagement.

ployment and migration histories since baseline, including spells in factory work, self-employment, and unpaid family labor. We also gather detailed measures of earnings and household expenditures, which allow us to track changes in income and savings over time. To capture potential shifts in expectations, norms, and intra-household dynamics, the survey includes modules on women's aspirations and attitudes and modules on decision-making within the household.[4]

3.2 Attrition and balance

Out of 334 women missing in Wave 7, 201 were already missing in the first follow-up, an attrition rate of 14%, not differential between treatment and control (difference estimator of -0.0025). This is because the baseline survey was conducted when women were interviewed for the jobs, not at home. If contact details were incorrect, we were unable to find these women in the first follow-up. Out of the remaining 1263 women, we were able to track 1130 women by Wave 7, achieving an attrition of only 11% after seven to eight years. Over a comparable horizon, the best studies have an attrition rate between 5% and 15% (see review by Bouguen et al. (2019)).

Table C.1 examines attrition in Wave 7. It reports estimates from regressing an indicator for being missing in Wave 7 on treatment status, controlling for list fixed effects (column 1) as well as interactions between treatment and key baseline fertility and age variables (columns 2-3). Across all specifications, attrition is uncorrelated with treatment. Baseline parity and age are correlated with attrition — women who are older and have more children at baseline are easier to track, likely because they are more settled — but this is true equally in treatment and control. We discuss attrition in further detail in Appendix C.

Table 1 reports summary statistics for our Wave 7 sample (column 1) and separately for treatment and control groups (columns 2 and 3). At baseline, women were on average 25 years old; 69% of them had already given birth at least once, and they had 1.4 children on average. They wanted 4 children on average, with an appropriate entry into motherhood at age 23, and 21% reported that their partner wanted more children than them. 31% of the women had previously had a formal wage job. 14% were Muslim and 23% Protestant. They had an average of 9.3 years of education. In terms of time use, they spent 34 hours per week on leisure and 6.5 hours on social and religious activities.

[4]Appendix B provides more details concerning the measurement and definition of all variables.

Table 1: Baseline characteristics of the endline sample, by treatment arm

	All		Treated		Control	
	(1)		(2)		(3)	
	Mean	SD	Mean	SD	Mean	SD
Treatment and fertility variables						
Treatment	0.50	(0.50)	1.00	(0.00)	0.00	(0.00)
Any child	0.69	(0.46)	0.66	(0.47)	0.71	(0.45)*
Number of children	1.36	(1.49)	1.28	(1.42)	1.43	(1.55)*
Partner wants more kids	0.21	(0.41)	0.21	(0.41)	0.21	(0.41)
Wanted number of children	3.96	(1.84)	3.99	(2.03)	3.93	(1.62)
Appropriate age of first birth	22.70	(2.88)	22.81	(2.87)	22.59	(2.88)
Other variables						
Any wage job ever	0.31	(0.46)	0.31	(0.46)	0.30	(0.46)
Age	25.13	(6.20)	25.13	(6.48)	25.12	(5.92)
Muslim	0.14	(0.35)	0.12	(0.33)	0.15	(0.36)**
Protestant	0.23	(0.42)	0.23	(0.42)	0.24	(0.43)
Years of education	9.28	(3.09)	9.38	(3.11)	9.19	(3.07)*
Leisure time last 7 days (hours)	33.98	(22.87)	33.98	(22.39)	33.98	(23.36)
Social and religious activities last 7 days (hours)	6.51	(6.60)	6.29	(6.08)	6.72	(7.08)
N	1130		560		570	
F-value						1.47
P-value F-test						0.13

Notes: The table reports means and standard deviations of baseline characteristics, measured at Wave 1 before random assignment, for the 1,130 women in the Wave 7 endline sample: column 1 for the full sample, column 2 for women randomized to receive a job offer, and column 3 for the control group. Asterisks indicate the statistical significance of the treatment-control difference for each characteristic, based on a regression of the characteristic on the treatment indicator including the randomization-list (strata) fixed effects. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$. The F-tests reported at the foot are joint tests that all listed characteristics are unrelated to treatment.

Even though an F-test cannot reject that both treated and control groups are overall balanced (p-value=0.13), there are some statistically significant differences. In particular, control women have 0.15 more children at baseline. This imbalance is entirely driven by the extensive margin: 71% of control women have a child against 66% of treated women. To address this, we control for the baseline number of children throughout. We also examine mothers and non-mothers separately to verify that the estimates are not driven by differences in sample composition across treatment and control. Controlling for the baseline outcome and splitting the sample by baseline parity were both specified in the pre-analysis plan.

3.3 Representativeness

Our sample of job-seekers is representative of the broader population of young, married, urban women living near the industrial parks on many dimensions. Appendix Table A.1 reports summary statistics using data from the 2016 EDHS, the closest wave in time to our baseline, and restricting the sample to married women aged 18-30 residing in an urban area within a 50-kilometer radius of the five industrial parks covered by our study. The average woman in that subsample is 25.5 years old, has 8.1 years of education, has already 1.2 children and want 3.8 children. 72% have had at least one child, 17% have a partner who wants more children than them, and about 27% engage in wage work. 7% are Protestant and 23% are Muslim. These orders of magnitude are very close to ours, except for religion and to a smaller extent education. The similarity between both samples suggests that there is no strong selection into *applying for* a job in terms of desired and realized fertility. This does not imply that there is no selection into *taking up* a job, as we will see.

Focusing on job-seekers is relevant from a scientific perspective and from a policy perspective. First, to test the opportunity cost of time channel, we need a substantial variation in this cost. If we had offered jobs to women who would never consider working, we would not have been powered to detect downstream fertility responses. Second, women who respond to our intervention are the marginal women of interest to policy makers. They are the ones who would adjust their fertility choices if more industrial jobs were created. Infra-marginal women have already adjusted and non job-seekers will not respond.

3.4 Pre-registered empirical specifications

The main specification in our pre-analysis plan is the following intent-to-treat (ITT) model:

$$Y_{i,t6} = \alpha Y_{i,t0} + \beta Treatment_i + \gamma X_{i,t0} + \delta List_i + \epsilon_{it},$$

where i indexes individuals, $t0$ refers to baseline values, and $t6$ is the sixth follow-up (Wave 7). $Treatment_i$ is a dummy variable equal to 1 if the woman was randomized to get the job offer and zero if not. The main coefficient of interest is β, which captures the average effect of receiving a job offer on the outcome of interest. We always include list fixed effects (blocking variables) as women are randomized within this unit. We select control variables $X_{i,t0}$ using a post-double LASSO selection approach (Belloni et al. 2014). As robustness, we present results with only the list fixed effects and with the full set of controls. We use heteroskedasticity robust standard errors in all estimations.[5]

In addition to the ITT specification, the pre-analysis plan includes an instrumental variable (IV) analysis and a duration analysis. The goal of the IV analysis is to estimate the causal effect of having a job on fertility – as we will show, not all treated women accepted the job offer and some control women were able to find another job. To estimate the causal effect of having a job, we therefore use the randomized job offer as an instrument for the actual employment status. The goal of the duration analysis is to study the *timing* of fertility responses. We estimate a non-parametric Kaplan-Meier model and a semi-parametric Cox model. In our main specification, we pool all parities and include dummies to control for the parity at baseline in $X_{i,t0}$. We then run separate regressions for women without a child at baseline and for women with at least one child.

4 Results

4.1 First stage results on employment

Table 2 shows the first-stage effects on employment. It reports the effect of being offered a job on the probability of holding a formal wage job in the past six months, by survey wave. By the first follow-up, roughly eight months after randomization (Wave 2), the treatment raises the probability of formal employment by 38 percentage points from a control mean of 29%, a very large effect.

[5] See Appendix B for more details on the pre-registered specification.

Table 2: First stage effect on wage employment, by wave

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.15***	0.38***	0.25***	0.16***	0.12***	0.032	-0.0077
	(0.021)	(0.028)	(0.031)	(0.033)	(0.034)	(0.033)	(0.028)
Control mean	0.32	0.29	0.31	0.35	0.34	0.33	0.32
N	5805	1037	989	922	860	874	1123
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso
Wave	2-7	2	3	4	5	6	7

Notes: The dependent variable is an indicator equal to 1 if the woman held a formal wage job in the six months before the interview. Column 1 reports the pooled effect across all post-baseline waves; columns 2 to 7 report the effect separately in Waves 2 through 7. Each column is a separate regression estimated by post-double-selection LASSO, with block (randomization-list) fixed effects always partialled out. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

The gap then narrows steadily and, by endline, the difference is close to zero. The last wave in which we can detect a significant effect is Wave 5, roughly 35 months after baseline. Thus, it takes three to four years for the employment effect to gradually fade away. Over the whole study period, the average effect is an increase by 15 percentage points, significant at 1%. The control mean is remarkably stable over time, meaning that the convergence is driven by treated women leaving their factory jobs.

Our findings are consistent with those of Abebe et al. (2025): they document that the positive employment effects had persisted by the first follow-up, eight months after baseline, but not by the four-year follow-up. They are different from Blattman and Dercon (2018), whose treated workers had mostly quit industrial employment already after one year. Two features plausibly explain the difference: our factories are more geographically dispersed in less urban areas, and our sample consists of partnered women whereas their applicants were predominantly unmarried. As a result, they were plausibly more mobile and had greater outside options.[6]

Given that the fertility analysis presents results for the sample of women who had no child at baseline, Appendix Table A.7 reports the employment effects for this sub-sample. They are more short-lived than in the full sample: the treatment effect is large in Wave 2 (28 percentage points, significant at 1%), decreases but remains significant in Wave 3, and disappears afterwards. Finally,

[6] We used the authors' replication package to produce Appendix Table A.2, which provides support for this interpretation. In their sample, married women exhibited significantly longer industrial employment spells in response to the job offer than unmarried women, accumulating roughly 4.3 months of industrial employment compared to 2.5 months among unmarried women. They were also more likely to remain employed in industrial work at endline. These patterns are consistent with married women having fewer outside options. 80% of their endline sample of women were unmarried. Note that their endline sample was 81% female; we did not find differential effects on job tenure by gender.

exploiting time use data in the first follow-up survey (Wave 2), we can show that paid work mainly crowded out leisure, household work, and social and religious activities (Appendix Table A.8).

4.2 Pre-registered results on fertility

All results presented in this section strictly follow the pre-analysis plan.

ITT estimates We first estimate the ITT effect of being randomly offered a factory job on women's fertility outcomes eight to nine years after baseline. Table 3 presents the results using our main specification. We find that job offers increased both the probability of having any child and the total number of children by endline, with effects concentrated among women who were not yet mothers at baseline (we call them "baseline non-mothers" for brevity). In the whole sample (columns 1 and 3), treated women are 4.1 percentage points more likely to have at least one child by endline ($p<0.01$), and have 0.12 additional children on average ($p<0.05$), relative to control means of 90 percentage points and 2.51 respectively. These aggregate effects are not driven by the imbalance in composition documented earlier. When we restrict the sample to baseline non-mothers, in columns 2 and 5, treatment raises the probability of becoming a mother by 15 percentage points and increases total births by 0.31 children ($p<0.01$), suggesting that the job offer accelerated entry into motherhood for this group. By contrast, effects on higher-order births among women who were already mothers at baseline are smaller in magnitude and imprecisely estimated (column 4).[7]

We find no treatment effects on wanted number of children or appropriate age at first birth (columns 6 and 7). If we examine mothers and non-mothers separately, estimates are small and insignificant in both samples. The stability of these stated preferences across treatment and control groups suggests that job offers did not alter women's fertility goals. Instead, treated women appear to be further along in achieving these goals by the time we observe them.

IV estimates To relate the changes in fertility to employment, we instrument months worked with treatment assignment. Figure A.4 displays the distribution of months worked in a formal wage

[7] Alternative specifications are shown in Appendix Table A.4. We see that controls matter for the effects on number of children. This sensitivity reflects the imbalance across treatment and control groups documented in Table 1. Note that we never find a negative treatment effect on endline number of children even though the treated group started with fewer children. Randomization inference corroborates these results: permuting treatment within randomization lists 1,000 times and re-estimating the full selection procedure in each permutation yields p-values of 0.011 (any child) and 0.039 (number of children) in the whole sample, and below 0.005 for both outcomes among baseline non-mothers.

Table 3: ITT estimates on pre-registered fertility outcomes

	(1) Any children	(2) Any children	(3) Number of children	(4) Number of children	(5) Number of children	(6) Wanted nr children	(7) Appropriate age first child
Treatment	0.041***	0.15***	0.12**	0.066	0.31***	-0.13	0.17
	(0.016)	(0.048)	(0.057)	(0.071)	(0.10)	(0.11)	(0.18)
Control mean	0.90	0.65	2.51	3.07	1.11	4.83	23.42
Sample	All	Non-mothers	All	Mothers	Non-mothers	All	All
N	1130	350	1130	773	350	1130	1130
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso

Notes: Each column reports the intention-to-treat effect of the random job offer on a fertility outcome measured at the Wave 7 endline. Columns 1 and 2 use an indicator for having any child (the extensive margin); columns 3 to 5 the number of children; column 6 the wanted number of children; column 7 the reported appropriate age at first birth. Columns 1, 3, 6, and 7 use the full endline sample; columns 2 and 5 restrict to women who were not mothers at baseline, and column 4 to women who were already mothers. The control mean row reports the outcome mean in the control group and N the number of women. Each column is a separate regression estimated by post-double-selection LASSO with block fixed effects partialled out. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$. Randomization-inference p-values (1,000 within-list permutations, controls re-selected per permutation): any child 0.011 (full), <0.001 (non-mothers); number of children 0.039 (full), 0.004 (non-mothers).

job over the nine-year period, separately for treatment and control. The treatment effect operates primarily at the extensive margin: over half of control women worked zero months across the entire period, compared to roughly one quarter of treated women. Conditional on working at all, the distribution is broadly similar across groups, with most workers accumulating between one and fifty months. As shown in Table 4, the first stage is strong: treatment increases months worked by 8.3 months (column 1) and raises the probability of working at least one month by 23 percentage points (column 4). Both estimates are significant at 1% and the F-stat are equal to 25 and 64, respectively.[8]

Table 4 further shows that OLS correlations between work and fertility are negative, while the IV estimates are positive. Panel A looks at any child while Panel B looks at number of children. The OLS coefficients indicate that working more months is associated with lower fertility, which likely reflects negative selection: women who choose to work more, or who remain employed longer, may be those who systematically prefer smaller families. The IV estimates correct for this selection and reveal the positive causal effect of employment on fertility. One additional month of formal employment raises the number of children by approximately 0.02 (Panel B column 3), while working at least one month in total raises the number of children by 0.72 (Panel B column 6). On the extensive margin, one additional month of formal employment raises the probability of having a

[8]For completeness, we present the first-stage effects separately for baseline mothers and non-mothers in Appendix Table A.3. They are large and statistically significant in both subsamples; magnitudes are smaller in the sample of non-mothers.

Table 4: IV analysis: instrumenting wage employment with treatment

Panel A. Any child						
	(1) First stage: Months worked	(2) OLS	(3) IV	(4) First stage: Any month	(5) OLS	(6) IV
Months worked		-0.00048*	0.0054**			
		(0.00029)	(0.0024)			
Any month					-0.020	0.19**
					(0.015)	(0.077)
Treatment	8.33***			0.23***		
	(1.67)			(0.029)		
Control mean	18.47	0.90	0.90	0.47	0.90	0.90
F-value	25			64		
N	1046	1046	1046	1046	1046	1046
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso
Panel B. Number of children						
Months worked		-0.0028**	0.020**			
		(0.0012)	(0.0091)			
Any month					-0.17***	0.72**
					(0.064)	(0.30)
Treatment	8.33***			0.23***		
	(1.67)			(0.029)		
Control mean	18.47	2.49	2.49	0.47	2.49	2.49
F-value	25			64		
N	1046	1046	1046	1046	1046	1046
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso

Notes: The table instruments formal wage employment with random treatment assignment, using two employment measures: the number of months worked and an indicator for having worked at least one month. Panel A uses having any child as the outcome, Panel B the number of children. Within each panel, columns 1 and 4 report the first-stage effect of treatment on the two employment measures, columns 2 and 5 the OLS association between employment and fertility, and columns 3 and 6 the corresponding instrumental-variable (two-stage least squares) estimates. The sample is the 1,046 women with complete work histories. Each equation is estimated by post-double-selection LASSO with block fixed effects partialled out; first-stage F-statistics are reported in each panel. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

child by about half a percentage point (Panel A column 3), while working at least one month raises it by roughly 19 percentage points (Panel A column 6).[9]

Duration estimates Finally, we examine the timing of fertility responses using birth histories. Figure 1 displays Kaplan–Meier estimates of the survival function for time between baseline and next birth. The curve indicates the fraction of women who have not given birth to another child yet, at any point in time after baseline, separately in the treatment group (red curve) and in the control group (blue curve). The steeper the curve, the higher the birth rate. By construction, the fraction is equal to 1 at baseline. Between baseline and month 9, we expect no difference between

[9]The IV estimates differ slightly from the ratio of reduced-form to first-stage estimates for two reasons: the IV sample excludes 84 women with incomplete work histories (N=1046 vs. 1130 in Table 3), and the post-double-selection procedure selects controls separately in each equation.

both groups because these children were conceived before the job offer. This is a test of balance and indeed, when we estimate a Cox model including the full set of baseline controls, we find no significant difference in hazard ratios between baseline and month 9: the p-value is equal to 0.65 for all women (Panel A) and to 0.32 when we restrict the sample to baseline non-mothers (Panel B).

After month 9, the survival curves are very close to each other during the period when employment differences are large and significant – up to Wave 5, 35 months after baseline, in the full sample, and up to Wave 3, 15 months after baseline, in the sample of baseline non-mothers. Afterwards, the curves start diverging: treated women are consistently more likely to give birth. The estimated Cox hazard ratio after month 35 is equal to 1.15 (p-value=0.12) in the full sample. The pattern is markedly stronger among baseline non-mothers, with a ratio of 1.72 after month 15 (p-value=0.004). Among those who were childless at baseline, only 20% of treated women had not given birth to a child by endline against 35% of control women. The duration analysis thus confirms the interpretation that job offers primarily affected the timing of first births. Moreover, we learn that the fertility response can only be detected once a substantial share of treated women have exited formal employment.[10]

4.3 *Tempo* or *quantum*? A surrogate analysis

Demographers distinguish between *tempo* effects affecting the timing of births without changing the final number of births, and *quantum* effects affecting completed fertility, i.e. the total number of children born to a woman at the end of her reproductive life (conventionally set at age 45). At endline, women in our sample are between 28 and 40 years old: we do not observe their completed fertility. In this section, we use a surrogate analysis to predict the treatment effect on completed fertility. In particular, we want to know whether the effects on “any children" and “number of children" estimated at endline are likely to persist or wash out once women turn 45 years old.

The intuition of the methodology is simple. We use the experimental sample to estimate the treatment effect on number of births by age. Then we use an observational sample to estimate the relationship between number of births by age and completed fertility. Finally we use the second

[10] Alternatively, we can study the timing of fertility responses by reconstructing panel data from retrospective birth histories and estimating ITT effects year-by-year. Results for the full sample are reported in Appendix Table A.5. The effects on childlessness emerge 4 years post-baseline (Panel A) and the effects on number of children emerge 7 years post-baseline (Panel B). When we restrict the sample to baseline non-mothers in Appendix Table A.6, we find that the effect on fertility onset builds up over time, slowly in the first two years then more quickly in the third and fourth years, generating a persistent gap of 14–18 percentage points ($p<0.01$) from 4 years post-baseline onwards.

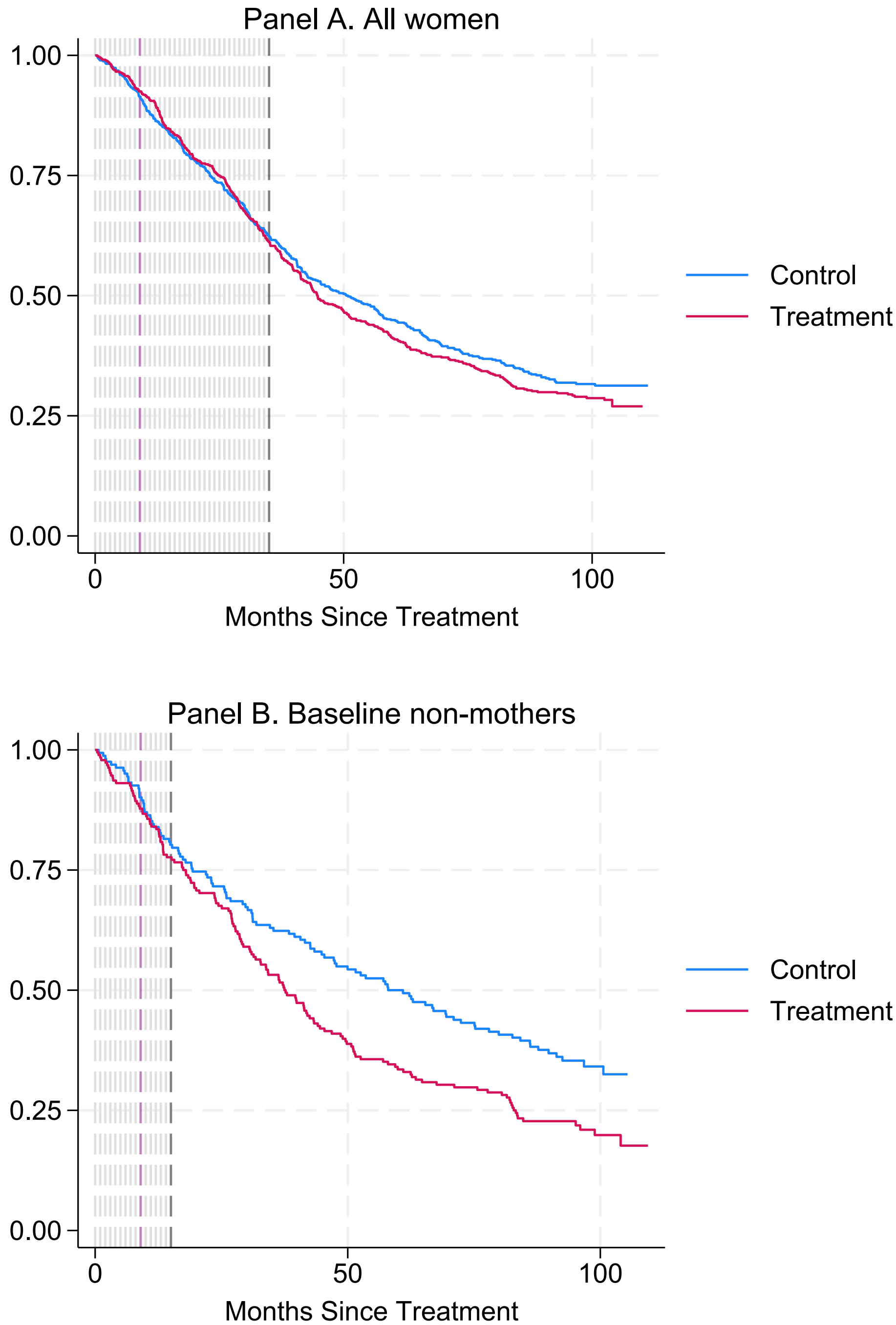


Figure 1: Duration analysis: time since baseline to next birth, by treatment arm

Notes: Kaplan–Meier survival estimates for months since baseline to next birth, by treatment status. In both panels the first dashed line marks 9 months after baseline: births before this point were conceived before the job offer. **Panel A (all women):** the second line at 35 months marks the average time since baseline in Wave 5, after which the first-stage employment effect disappears; the null of identical survival curves cannot be rejected (log-rank p-value=0.34); a Cox model including the full set of baseline controls and allowing the treatment effect to vary over time gives a hazard ratio of 0.90 (p-value=0.65) between baseline and 9 months, 1.07 (p-value=0.61) between 9 and 35 months, and 1.15 (p-value=0.12) after 35 months. **Panel B (baseline non-mothers, i.e. everyone not having a child before baseline):** the second line at 15 months marks the average time since baseline in Wave 3, after which the employment effect disappears; the null of identical survival curves can be rejected at 1%; the corresponding hazard ratios are 1.49 (p-value=0.32) between baseline and 9 months, 0.76 (p-value=0.46) between 9 and 15 months, and 1.72 (p-value=0.004) after 15 months.

step to predict completed fertility in the experimental sample and estimate the treatment effect on predicted completed fertility. The method is valid under two assumptions (Athey et al. 2025). First, surrogacy: the treatment affects completed fertility only through the number of children at endline. In other words, we assume that the differential process has settled by endline and that treated and control women will not differ in future childbearing beyond what number of births by age captures. Second, comparability: the experimental sample and the observational sample should be comparable. This is why we use the most recent wave of EDHS (2024) and restrict the dataset to urban areas.[11] Since both assumptions are strong, we consider the following analysis as suggestive.

Formally, in our endline data, we observe (n_i, a_i), the number of children and the age of each woman. Her completed fertility is equal to $CF_i = n_i + R_i$ where R_i is the remaining number of births between endline and age 45. Next, we turn to EDHS and focus on women aged 45 to 49 (completers). For each woman, we exploit her birth history to reconstruct what her number of births was at any past age and infer $R_i(n, a)$, her remaining number of births. We regress R_i on n and a using a flexible specification (quadratics plus interaction). Finally, we go back to the experimental sample and predict each woman's completed fertility as $\hat{CF}_i = n_i + \hat{R}(n_i, a_i)$.

We can now estimate the treatment effect on predicted completed fertility using our usual specification. The suggestive results are reported in Table 5 Panel A. We find that the extrapolated treatment effect on completed fertility is very close to the treatment effect on number of births at endline: 0.13 significant at 5% in the whole sample and 0.32 significant at 1% in the sample of non-mothers.[12] This is because, in the EDHS sample of completers, the number of births by age has no significant effect on the remaining fertility. Indeed, when we estimate the derivative of $\hat{R}(n, a)$ with respect to n and averaged over a, denoted κ, we find $\kappa = 0.017$, not significantly different from zero. This means that there is no replacement: each extra birth at a given age translates into exactly one extra birth at age 45. In contrast, for our endline result to vanish at age 45, we would need $\kappa = -1$: each extra birth at a given age should crowd out exactly one future birth (full replacement).

We repeat the exercise for childlessness. Instead of estimating the remaining number of births $R(n, a)$, we estimate the persistence probability $P(a)$, the probability of being childless at age

[11] As alternative observational samples, we used EDHS 2016 alone or pooled with 2024, with or without the restriction to urban areas. The results are very similar.

[12] The standard errors presented here do not account for the fact that the outcome is predicted. If we adjust them using the delta method, they increase by about 0.1%, which is innocuous.

Table 5: Surrogate analysis: suggestive extrapolation to completed fertility

Treatment effect on ...	Full sample ATE (SE)	Baseline non-mothers ATE (SE)
Panel A: number of children		
Observed endline number of children	0.117 (0.057)	0.314 (0.103)
Predicted completed fertility	0.132 (0.063)	0.320 (0.107)
Panel B: any children		
Observed endline childlessness	-0.041 (0.016)	-0.151 (0.048)
Predicted completed childlessness	-0.012 (0.006)	-0.051 (0.019)

Notes: ATE controlling for block fixed effects and controls selected by Lasso.
Predicted completed fertility = endline number of children + predicted remaining births, where $E[\text{remaining births} \mid \text{endline number of children, age}]$ is estimated among women aged 45–49 in 2024 EDHS.
Predicted completed childlessness = $\mathbf{1}\{\text{endline childless}\} \times \hat{P}(\text{age})$, where $\hat{P}(\text{age})$ is the probability of being childless at age 45 conditional on being childless at a given age and is estimated among women aged 45–49 in 2024 EDHS.

45 given childlessness at age a in the EDHS sample of completers. Then we predict completed childlessness in the experimental sample as $\hat{C0}_i = \mathbf{1}\{n_i = 0\} \times \hat{P}(a_i)$. The suggestive results are reported in Table 5 Panel B. The extrapolated effect on completed childlessness is much smaller than the effect at endline: -1.2pp significant at 5% in the whole sample and -5.1pp significant at 1% in the sample of non-mothers. Only one third of the endline gap between treated and control women persists. Contrary to the number of births, persistence is much weaker for childlessness because the vast majority of women will eventually have a child.

The surrogate analysis suggests that the treatment effect on childlessness is most likely a tempo effect while the one on number of births is most likely a quantum effect. This is not a contradiction. Our main result is that treated women have their first birth earlier. Control women are expected to catch up only partially: they will have children, but not as many as treated women. Therefore, the intensive margin gap may survive even as the extensive margin gap closes.

5 Mechanisms

The goal of this section is not to argue that we need a brand new model to account for surprising experimental results. Instead, we want to highlight that a standard Beckerian framework can explain a lot. We proceed in three steps. First, we revisit the theoretical conditions under which the income effect dominates the substitution effect in a static framework. We then present a simple dynamic extension to account for the timing of the effects. Second, we provide empirical evidence that the income effect is indeed large in our context. Third, we investigate alternative mechanisms, in particular female bargaining power, and show that they are not supported by the data.

5.1 Conceptual framework

Insights from a standard static model We start with a standard model of fertility choices and female work opportunities. Following the review by Doepke et al. (2023), we write the couple's maximization problem as:

$$\max_{c,n,L} log(c) + \delta log(n) \text{ s.t. } c + pn \leq I + Lw \text{ and } L + n\phi \leq 1$$

Couples choose their consumption c and their number of children n subject to a budget constraint and a time constraint. Each child has a monetary cost of p (the price of adult's consumption is normalized to 1). The household has two sources of income: (i) an external income I which does not depend on the number of children, e.g. income from men's work; and (ii) women's wage income which is equal to her wage w multiplied by her labor supply L. We normalize the unit of time to 1 and we assume that each child has a time cost of ϕ: female wage income is therefore equal to $(1 - \phi n)w$. δ is the weight attached to children in the utility function; the weight attached to consumption is normalized to 1. The optimal choice for fertility is:

$$n^* = \frac{\delta}{1+\delta}.\frac{I+w}{p+\phi w}$$

In the absence of female work opportunities, $w = 0$ and $n^* = \frac{\delta}{1+\delta}\frac{I}{p}$. Children are normal goods. Couples spend a share $\frac{\delta}{1+\delta}$ of their income on children and a share $\frac{1}{1+\delta}$ on consumption. The number of children increases with income, decreases with the price of children relative to consumption and increases with the taste for children relative to consumption.

Introducing female work opportunities ($w > 0$) has two opposite effects. First, an *income effect*, raising the potential total income from I to $(I + w)$, which increases n^*. Second, a *substitution effect*, raising the price of a child from p to $(p + \phi w)$, which decreases n^*. Under which conditions does the income effect dominate? Taking the derivative of n^* with respect to w, we find that:

Proposition 1. *The income effect dominates the substitution effect iff* $\frac{p}{I} > \phi$

Intuitively, $\frac{p}{I}$ captures the cost of child as a share of men's income, while ϕ captures the cost of child as a share of women's time. The income effect is more likely to dominate when a child is more costly in monetary terms than in time terms.

Turning to labor supply, $L^* = 0$ if $w = 0$ and $L^* = 1 - \phi n^*$ if $w > 0$. If the income effect dominates, an increase in wages reduces L^* (through an increase in n^*) - except at 0, where introducing female work opportunities discontinuously raises labor supply from zero to a positive number. This implies that women should exit the labor market earlier when new work opportunities pay a high wage compared to a low wage.

We interpret our experiment as raising the probability of receiving a job offer from $\pi < 1$ in the control group to 1 in the treatment group; all offers pay the same wage $\bar{w}$. We model the treatment as changing the probability rather than changing the wage because conditional on working, women earn the same wage in both groups. The intervention does not give them access to a better job than the counterfactual job, it gives them access to a job for sure.[13] In other words, it raises w from 0 to $\bar{w}$ for a fraction $(1 - \pi)$ of women in the treatment group.

The sign of the ITT estimate on fertility is informative about the relative magnitude of $\frac{p}{I}$ and ϕ. Fertility increases iff the income effect dominates the substitution effect. This simple framework delivers two additional testable implications.

Heterogeneous treatment effect by baseline income. The income effect is less likely to dominate when I is large. We expect the treatment effect on fertility to be less positive and even potentially turn negative as baseline external income increases.

Heterogeneous treatment effect by wage offered. We expect heterogeneous responses depending on the level of $\bar{w}$. If the income effect dominates, we should observe a larger treatment effect on fertility and a smaller treatment effect on labor supply when $\bar{w}$ is higher.

Insights from a simple dynamic extension The static model cannot account for important findings on the timing of births: (i) the fertility response is concentrated on non-mothers at baseline, suggesting that the relevant margin is the timing of the first birth; (ii) by the end of reproductive life, the effect on childlessness is likely to vanish while the effect on the number of births is likely to persist; (iii) in the short-run, there is an effect on employment and no effect on fertility whereas the

[13] Using the first stage on having a wage job, we infer that the proportion of never-takers is equal to 33% and the proportion of always-takers is equal to 29%. This implies that the probability of getting an offer in the control group is $\pi = \frac{0.29}{1-0.33} = 0.43$. 43% of women get an offer and 67% accept it, which leads to a formal employment rate of 29% in the control group.

opposite is true in the long-run. In this section, we argue that we can rationalize these patterns by making two assumptions within a dynamic framework: households want to smooth consumption over time and they can save but not borrow. The first assumption is standard. The second assumption is plausible in our context: 88% of women have a savings bank account but only 3% have borrowed money from a bank or a money lender.

We extend the static model by assuming that there are two periods. In each period, couples decide whether to have a child or not: $d_t = \{0;1\}$ is a discrete choice. For simplicity, we assume that there is no discounting. The maximization problem is:

$$\max_{c_t, d_t, L_t, s} log(c_1) + log(c_2) + \delta log(d_1 + d_2)$$

$$\text{s.t.} \quad \mathrm{c}_1 + pd_1 + s \leq I + wL_1 \text{ and } c_2 + pd_2 \leq I + wL_2 + s \text{ and } s \geq 0$$
$$\text{and} \quad \mathrm{L}_1 + d_1\phi \leq 1 \text{ and } L_2 + d_2\phi \leq 1$$

Couples care about their consumption in each period, c_t, and about their total number of children, $(d_1 + d_2)$. They can save in period 1 and consume their savings in period 2, but they cannot borrow: s cannot be negative. They have one unit of time in each period. If they decide to have a child in period t, they have to pay the monetary cost and the time cost in that same period.

In this model, due to decreasing marginal returns to consumption, couples maximize their intertemporal utility by using their savings to equalize c_1 and c_2. They have incentives to postpone their expenditures to period 2 in order to save in period 1 and dissave in period 2.

Since fertility choices are discrete, there are four options for (d_1, d_2). Option $(0,0)$, having no child at all, is dominated by assumption because of the log in the utility function. Option $(1,0)$, having a single child in period 1, is dominated by option $(0,1)$, having a single child in period 2, because $log(I + w - \phi w - p) + log(I + w) < 2log(I + w - \frac{\phi w + p}{2})$. The intuition is that couples cannot smooth consumption if expenditures are higher in period 1. The only relevant comparison is between option $(0,1)$ and option $(1,1)$, having a child in both periods. Couples are better off with two children than with a single child iff:

$$2log(I + w - \frac{\phi w + p}{2}) < 2log(I + w - \phi w - p) + \delta log(2)$$

$$\Leftrightarrow \delta > \tilde{\delta} = \frac{2}{log(2)} log(1 + \frac{p+\phi w}{2(I+w-(p+\phi w))})$$

Only couples who value children enough – i.e. with a utility weight on children δ larger than a given cutoff $\tilde{\delta}$ – have a child in period 1. All couples have a child in period 2. In this model, couples have incentives to postpone births as much as possible in order to save and smooth consumption. Only those who want several children start early because they need time to have them all.

We have the same predictions as in the static model. When $w = 0$, $\tilde{\delta}(I,p) = \frac{2}{log(2)} log(1 + \frac{p}{2(I-p)})$. The cutoff decreases, hence fertility increases, when I is higher and when p is smaller. Our experiment raises the potential income to $(I+w)$ and raises the price of a child to $(p+\phi w)$. In net, the combination of both effects lowers the cutoff under the same condition as before (see Figure D.1 in Appendix D for a graphical illustration):

$$\tilde{\delta}(I+w, p+\phi w) < \tilde{\delta}(I,p) \Leftrightarrow \frac{p}{I} > \phi$$

When the income effect dominates, treated couples with $\delta \in (\tilde{\delta}(I+w, p+\phi w); \tilde{\delta}(I,p))$ switch from $(0;1)$ to $(1;1)$. In other words, the treatment induces some couples to have a child in the first period. Therefore, at the end of period 1, there is a positive treatment effect on the number of children and on having any children. At the end of period 2, the effect on childlessness disappears entirely whereas the effect on births persists entirely, because all couples have exactly one child in the second period. This is consistent with the treatment affecting mostly the timing of first births and with the control group catching up at the extensive margin but not at the intensive margin.

Finally, a similar argument about consumption smoothing under credit constraints can explain why the effects on employment and fertility are sequential (see Appendix D for a formal exposition). Recall that, when women have a child in period t, they have to devote a fraction of time ϕ to childcare and can only work a fraction of time $(1-\phi)$. Suppose that some women cannot simultaneously combine childcare and work activities: at a given point in time, they either work or care for the child. When is the optimal moment to give birth over the course of the period? Again, if women can save but not borrow, the only way to keep the flow of instantaneous expenditures constant during the whole period is to work first during $(1-\phi)$ unit of time, in order to accumulate savings, and then exit the labor market for the remaining ϕ unit of time in order to have children. This implies that, during the initial spell of the period, we should observe an effect on labor supply and no effect on fertility, whereas in the final spell of the period, we should observe no effect on labor supply and an effect on fertility.

This conceptual framework is consistent with qualitative work in Ethiopia documenting that women consider factory jobs as a temporary, savings-generating bridge to subsequent life choices, like child-rearing and/or self-employment. For instance, Blattman and Dercon (2018, p.31-32) write that *"It is worth noting that nearly all of the applicants we interviewed described industrial jobs as temporary in the sense that they did not expect to work there for more than a few years. For example, women commonly explained that they planned to work in the formal sector for only a few years, until they had children and took on child-rearing, household occupations, and more flexible part-time self-employment (...) Many young people, perhaps even a majority, also expressed a preference for one day running their own businesses. Formal sector jobs were temporary jobs while they accumulated enough savings to start an enterprise".*

In terms of heterogeneity, the insights from the static model still hold: (i) the income effect is more likely to dominate when the baseline income is low; and (ii) if the income effect dominates, treated women are more likely to have a child early, and consequently to exit employment early, when the wage is high. The dynamic model delivers two additional predictions. First, switchers are in the middle of the δ distribution. Women with a very low taste for children never have a birth in the first period whereas women with a very high taste always have one. Second, the sequential effect should be driven by women who cannot combine childcare and work activities. We test all these predictions in the next section.

5.2 Supporting evidence for the income effect

Large treatment effect on household income and savings We start by showing that the treatment indeed made households persistently richer. In each wave, we asked households about their income and savings over the previous six months. To capture the cumulated gains over time, we sum each variable across waves: in Table 6, column 1 reports the effect for Wave 2 alone, column 2 for Waves 2–3, column 3 for Waves 2–4, and so on until Wave 7. This gives us a lower bound of the total gains over the whole study period.[14] We restrict the sample to households observed in all waves (N=795).

[14]The time length between two waves exceeds six months so we do not observe the whole study period. We could have interpolated the data to fill in the missing months; however, we do not know the exact dynamics of income and savings over time. This is why we prefer to show the raw data and interpret our estimates as a lower bound.

Table 6: Mechanism: effects on cumulated household income and savings.

Panel A. Household income						
	(1)	(2)	(3)	(4)	(5)	(6)
	Wave 2	2-3	2-4	2-5	2-6	2-7
Treatment	2525.0***	4290.3***	5742.2***	6991.3***	8773.7***	10001.8***
	(776.7)	(1338.3)	(1839.8)	(2420.0)	(2946.1)	(3276.3)
Control mean	17038.46	34695.98	52632.85	73041.25	90860.60	103894.14
N	795	795	795	795	795	795
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso
Panel B. Savings						
Treatment	2305.8***	3434.5***	4303.1***	5494.7***	7011.6***	7651.4***
	(641.7)	(1065.8)	(1417.7)	(1887.0)	(2207.3)	(2440.7)
Control mean	1903.81	4427.53	8141.48	12454.73	16130.14	15842.77
N	795	795	795	795	795	795
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso

Notes: Panel A reports effects on cumulated household income and Panel B on cumulated household savings, each measured over the previous six months in every wave. Column 1 reports the effect for Wave 2 alone; each subsequent column adds one wave, cumulating the amount across waves, up to Waves 2 through 7 in column 6. Each column is a separate regression estimated by post-double-selection LASSO with block fixed effects partialled out. The control mean row reports the cumulated amount in the control group. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$. Amounts are expressed in current Ethiopian birr. The balanced sample is restricted to the 795 women observed in all waves.

In Wave 2, treated households report earning an additional 2500 ETB compared to control households in the past six months (Panel A, column 1). This is very close to the expected change in women's income given the treatment effect on employment and the average monthly wage: $0.38 \times 1021 \times 6 = 2328$. This suggests that other household members have not adjusted their labor supply in response to the treatment. Almost all of the additional income was saved: treated households report saving 2300 ETB more than control households during the same period (Panel B, column 1).

Treatment effects accumulate over time and by endline, treated households have earned at least 10000 ETB more than control households (Panel A, column 6), which represents a 10% increase relative to the control mean. A substantial portion of this gain was saved: Panel B shows that cumulative savings in treated households exceed those in control households by at least 7500 ETB, about half of the control group's total (Panel B, column 6). Roughly 70% of the gains are concentrated in Waves 2 to 5, when we observe a significant effect on female wage employment. If we used constant ETB instead of current ETB, this fraction would be much larger because inflation was particularly high between Wave 5 and Wave 7. Financial data in low-income settings are noisy, but the consistency of these patterns supports the interpretation that treated households became permanently richer, despite only transitory increases in women's labor supply.

Heterogeneity by baseline income Our first main prediction is that the positive fertility response should be concentrated among poorer households. Table 7 splits households into quintiles of baseline income and shows three patterns consistent with this prediction. First, in the control group, women in the richest quintile are significantly more likely to have transitioned into motherhood by endline – consistent with children being normal goods. Second, the treatment effect on entry into motherhood is positive for women in the bottom four quintiles but turns negative for the top quintile, suggesting that for the richest households the income effect is not large enough to offset the substitution effect.[15] Third, the positive treatment effect has the same magnitude as the top quintile advantage in the control group. Taken together, these patterns imply that the treatment allows the bottom 80% of the income distribution to catch up with the top 20%. The same results hold with less precision when we look at the number of children (column 3) and they are starker in the sample of non-mothers (columns 2 and 4).

Heterogeneity by wage level Our second main prediction is that, conditional on the income effect dominating, the fertility response should be stronger – and the labor supply response weaker – in higher-wage firms. To test this, we exploit the diversity of the 27 firms in our sample and split them into those paying above and those paying below the median wage.[16] This split is not experimental and is correlated with other firm characteristics: relative to low-wage firms, high-wage firms are less likely to have workplace conflicts, have operated for longer, and are more likely to be Ethiopian-owned rather than Chinese-owned.

We find that our results are driven by the high-wage firms. This pattern comes across clearly in Figure 2, which shows duration analyses for both groups of firms separately. In high-wage firms (Panel A), the survival curves for treatment and control diverge more sharply than in the pooled sample. In low-wage firms (Panel B), by contrast, the two curves track each other closely throughout. Regarding first-stage effects, they are very similar in both groups up to Wave 5, after which they disappear in high-wage firms (like in the poold sample) whereas they persist until endline in low-wage firms.

The heterogeneous effect on employment is itself consistent with the income effect dominating

[15]The heterogeneous fertility effect is not driven by heterogeneous take-up of the job offer. In the first-stage regression, the treatment coefficient is the same across all quintiles (Appendix Table A.9). Richer women are equally likely to take up the job as other women; however, their fertility response is different.

[16]Specifically, we compute the average worker-reported basic monthly salary from Wave 2 (removing the firms with less than 10 observations), and split them at the median of these firm-level mean wages.

Table 7: Heterogeneity by baseline household income quintiles

	(1) Any children	(2) Any children	(3) Number of children	(4) Number of children
Treatment	0.067**	0.26***	0.14	0.37**
	(0.031)	(0.095)	(0.12)	(0.18)
Income Q2	-0.038	-0.098	-0.16	-0.037
	(0.039)	(0.11)	(0.12)	(0.21)
Income Q3	0.014	0.045	0.044	0.24
	(0.035)	(0.12)	(0.12)	(0.23)
Income Q4	-0.015	-0.097	0.097	-0.21
	(0.035)	(0.11)	(0.12)	(0.22)
Income Q5 (richest)	0.065**	0.23**	0.27**	0.23
	(0.028)	(0.092)	(0.13)	(0.20)
Treatment × Income Q2	-0.00017	-0.059	0.13	-0.14
	(0.049)	(0.13)	(0.17)	(0.28)
Treatment × Income Q3	-0.037	-0.20	-0.13	-0.38
	(0.045)	(0.15)	(0.18)	(0.30)
Treatment × Income Q4	-0.015	-0.058	-0.019	0.34
	(0.046)	(0.14)	(0.18)	(0.31)
Treatment × Income Q5 (richest)	-0.089**	-0.36***	-0.15	-0.32
	(0.039)	(0.13)	(0.17)	(0.29)
Control mean	0.90	0.65	2.51	1.11
Sample	All	Non-mothers	All	Non-mothers
N	1130	350	1130	350
Controls	Lasso	Lasso	Lasso	Lasso

Notes: The table interacts treatment with indicators for quintiles of baseline household income; the omitted category is the poorest quintile (Q1). Columns 1 and 2 use having any child as the outcome and columns 3 and 4 the number of children. Columns 1 and 3 use the full endline sample; columns 2 and 4 restrict to women who were not mothers at baseline. Each "Income Qk" row gives the control-group difference for quintile k relative to Q1, and each "Treatment × Income Qk" row the differential treatment effect for that quintile. Each column is a separate regression estimated by post-double-selection LASSO with block fixed effects partialled out. The control mean is the outcome mean in the control group. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

in our setting: in low-wage firms, the wage is not high enough to push women into childbearing, so they remain employed longer. When we look at accumulated income and savings, the treatment effect on income is 3.5 times larger in high-wage firms than in low-wage firms, and there is no effect on savings in low-wage firms. This is in line with the idea that women exit employment once they have saved enough, and this takes longer in low-wage firms.

Appendix E provides more details: we present the correlates of high-wage and low-wage factories, we show that our sample is balanced between treatment and control within each subsample and we report the treatment effects on the main fertility outcomes, on employment by wave and on cumulated income and savings for each subsample. We also show that these heterogeneous treatment effects are robust to classifying firms as high- or low-wage using the regional median wage instead of the national one in order to absorb some of the unobserved heterogeneity.

Heterogeneity by taste for children Another prediction is that only couples with an intermediate taste for children should switch the timing of their first birth in response to the offer. Using the desired number of children at baseline as a proxy for δ, we find exactly this: the treatment effect on "any child" is concentrated among women who want three or four children. The effect is smaller and not significant among those who want fewer than three or more than four (Table A.10).[17]

Sequence of events at the individual level and heterogeneity by time costs A final prediction is about heterogeneity in women's ability to combine work and childcare. We use retrospective information about job history and birth history at endline to study individual trajectories in the sample of non-mothers. We observe that in roughly half of the cases, women quit their jobs and then have their first child, whereas the other half have their first child while working and quit later. This suggests that for the second half of the sample, $\phi \approx 0$, i.e. women can combine work and childcare. For them, there is no trade-off: the income effect unambiguously dominates and we expect the positive effect on fertility to materialize very quickly. For the first half, who quit and then have a child, $\phi > 0$ and the trade-offs described in our dynamic model are relevant. We expect them to drive the sequential effects observed in the aggregate data.

[17] Our pre-analysis plan also specifies an omnibus test for treatment-effect heterogeneity, using the generic machine learning approach of Chernozhukov et al. (2023) in an honest random forest (Athey et al. 2019). Applied to the full set of baseline covariates, neither approach detects systematic heterogeneity in the fertility effects (all heterogeneity-test p-values exceed 0.18); we therefore do not proceed to the pre-specified characterization of the most and least affected women.

Unfortunately, we do not have a good baseline proxy of ϕ. The best we can do is using the endline question asking women whether they could work at the factory with a newborn: the sample is also split roughly equally between "yes" (45%) and "no" (55%). When plotting the survival curves for each sub-sample among non-mothers, we find that treated and control groups diverge from the beginning among women who can combine work and childcare – in line with an immediate income effect (Panel A of Figure A.5). In contrast, the divergence appears only in the medium run among women who cannot combine (Panel B of Figure A.5) – in line with the mechanism that women need to save up for a child when they anticipate exiting the labor market after childbirth.

5.3 Alternative mechanisms not supported by the data

Taken together, the patterns presented in section 5.2 make it plausible that the fertility effects operate through improved household resources. We now consider a series of alternative channels through which the treatment could have affected fertility, and show that none of them are supported by the data. This strengthens our confidence that income effects are the dominant mechanism.

Intra-household bargaining and conflict The first set of alternative mechanisms relates to the balance of power within households. This is a natural margin to consider, given that the job offers substantially increased women's earnings and that a non-trivial share of treated women came to contribute as much as or more than their partners to total household earnings. Higher relative earnings could shift intra-household bargaining power, potentially altering fertility outcomes if women and their partners have different fertility preferences (Rasul 2008; Ashraf et al. 2014; Doepke and Tertilt 2018; Doepke and Kindermann 2019). In standard economic models of the household, an increase in female relative earnings is predicted to shift outcomes towards women's preferences. We can get the opposite prediction in models with domestic violence if women are punished for their financial independence.

We collected detailed information on intimate partner violence using standard questions to elicit the prevalence of physical and emotional abuse in the past 3 months. We also collected detailed information on fertility decision-making within the household, including who makes decisions about having children and women's perceptions of their partner's desired fertility. Specifically, we ask who in the household usually has the final say on a set of domains, including whether to have children or more children, which family planning method to use, what to do if the respondent falls sick, whether she may earn money outside the home, whether she may visit relatives, and several financial decisions. Each item is coded as one if the husband has the final say, and zero otherwise. Joint decisions are coded as zero if the woman reports having "a lot" of input. We construct a *Husband decides index* averaging across all domains.

We find no evidence that factory job offers affected intra-household conflict or bargaining power. Results are reported in columns 1–6 of Table 8. There is no treatment effect on physical or emotional abuse (columns 1 and 2), decision-making on fertility and contraception (columns 3 and 4) or decision-making in general (column 5). Likewise, women's confidence that their opinions are taken seriously in fertility decisions also show near-zero effects. Across all outcomes the estimates are small, statistically indistinguishable from zero, and mixed in sign. These results are in line with the conclusions of the initial study using data from earlier waves (Kotsadam and Villanger 2025). On contraception more specifically, we asked whether women were using any contraceptive method in Waves 2–6. Appendix Table A.11 shows no treatment effect on contraceptive use in any wave; estimates for baseline non-mothers are similarly close to zero. This suggests that the fertility response operates through a change in *traditional* birth control methods, which require the cooperation of both partners.

A related concern is that women taking up factory work could face pressure, not only from their husbands but also from the whole community, to compensate by conforming more strongly to traditional wife-and-mother roles, leading to higher fertility as a form of social signaling. This mechanism requires conservative baseline norms regarding female labor force participation, for which we find no support in our data. When asked whether "women who work for a salary outside the home are more respected in the local community," 92% of respondents agree at baseline. Conversely, only 11% agree that "it is generally preferred that married women should work on household tasks such as taking care of children, collecting firewood, cleaning and cooking, and that they should not take salaried employment away from home." Women's own attitudes are similarly progressive: 99% agree that "it is okay for women to work outside of the home," and 92% agree that "men should be responsible to help with childcare when his wife is busy with business or factory job." With baseline attitudes this strongly favorable to female employment, there is little scope for a backlash mechanism to operate in our context.

Table 8: No effect on other potential mechanisms

	Bargaining and abuse						Marriage and mobility			Health	
	(1) Physical abuse	(2) Emotional abuse	(3) Husband decides no. children	(4) Husband decides contraception	(5) Husband decides index	(6) Confident in fertility decisions	(7) Married	(8) Same partner	(9) Never moved	(10) Number of children dying	(11) Food insecurity index
Treatment	0.0050	0.0012	0.016	-0.0079	0.0025	-0.032	0.0026	0.0036	-0.035	0.019	-0.066
	(0.015)	(0.021)	(0.023)	(0.019)	(0.014)	(0.028)	(0.025)	(0.027)	(0.029)	(0.025)	(0.048)
Control mean	0.08	0.16	0.26	0.12	0.27	0.66	0.79	0.73	0.68	0.16	0.04
N	1130	1130	1043	1036	1022	1130	1130	1130	1130	1130	788
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso

Notes: Each column is a separate regression of the outcome named in the header on treatment, estimated on the Wave 7 endline sample with the same post-double-selection LASSO specification and block fixed effects as the main tables. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$. Columns 1–6 report bargaining and abuse outcomes: physical and emotional abuse are measured over the three months before the interview, columns 3–5 are indicators for the husband deciding on the number of children, on contraception, and an index averaging 10 DHS-type husband-decides questions, and column 6 is whether the woman feels confident voicing her opinion on fertility decisions. Columns 7–9 report marriage and mobility outcomes: whether the woman is currently married, has the same partner as at baseline, and has never moved residence. Columns 10–11 report health outcomes: the number of the woman's children who have died, and a food-insecurity index. The food-insecurity index (column 11) is the average across Waves 2–7 of the wave-specific Household Food Insecurity Access Scale, each standardized on the control group, for the balanced panel of women observed in all six waves ($N = 788$).

Marriage and migration The second set of alternative mechanisms is about stability. Job offers may encourage marriage and discourage geographic mobility, all of which is favorable to high fertility. Table 8 shows no effect on the probability of being currently married (column 7), having remained with the same partner since baseline (column 8), or never having moved (column 9). Partnership instability may explain why some women are still childless at endline: among them, only 32% have kept the same partner during the whole period. However, it does not explain why childlessness is lower in the treatment group.

Aspirations A third potential explanation is that exposure to factory work could shift women's aspirations—about their children's futures or about their own life trajectories—in ways that raise fertility. This channel is plausible precisely because the jobs are of low quality: as documented in Section 2, they are monotonous, offer few prospects for advancement, and have been shown to reduce women's civic engagement even as they raise income (Aalen et al. 2024). Confronted with such jobs, women may revise their aspirations downward, which could affect fertility through two routes. First, they could lower their perceived returns to education, which would tilt the quantity-quality trade-off towards quantity. We find no evidence for this: treatment does not affect women's educational aspirations nor family aspirations for their daughters or sons (Appendix Table A.12). Second, they might give up hopes of a career and substitute toward family formation as an alternative source of meaning or status. This has a testable implication across the wage distribution of treatment firms: women in low-wage firms have greater scope for disappointment than those in high-wage firms, and so should be more likely to quit early and have additional children. Our heterogeneity in retention effects goes in the opposite direction (Figure 2), which weighs against this interpretation.

Child and maternal health A final channel is health. The effect of jobs is a priori unclear: they can improve health through higher income or deteriorate health through poor working conditions. The positive effect on number of children requires an improvement in child survival (which mechanically increases the number of surviving births) or an improvement in maternal health or nutrition, which could increase fecundity.[18] Table 8 column 10 shows no treatment effect on the number of children who died. Using data from previous waves, we also find a small and insignificant effect on

[18] The reproductive health literature links undernutrition to reduced fecundity and elevated miscarriage risk (ESHRE Capri Workshop Group 2006), though Bongaarts (1980) suggests the effect on actual fertility is likely small.

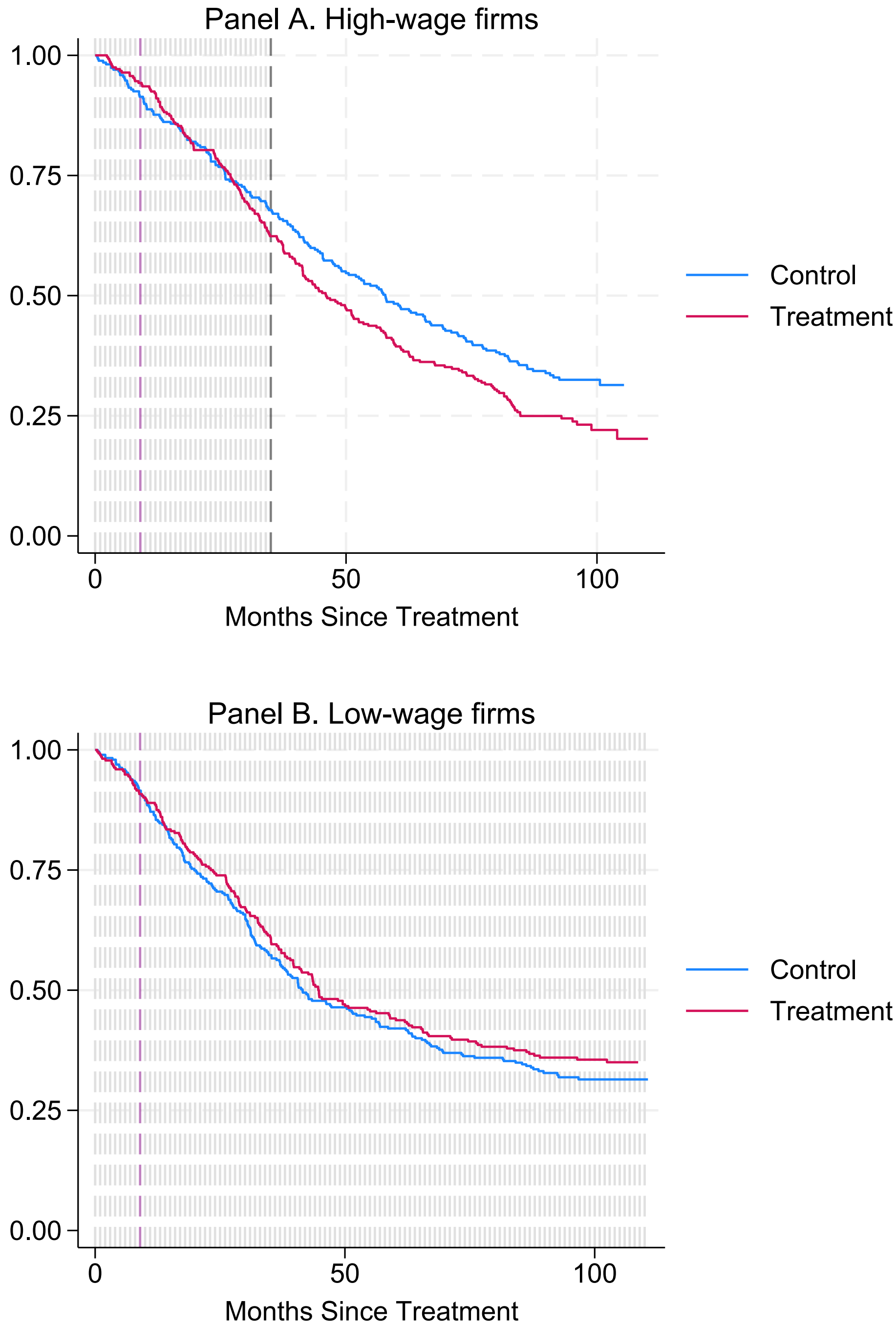


Figure 2: Heterogeneity by firm wage level

Notes: Kaplan–Meier survival estimates for months since baseline to next birth, by treatment status, split by firm wage level. In both panels the first dashed line marks 9 months after baseline: births before this point were conceived before the job offer. **Panel A (high-wage firms):** the second line at 35 months marks the point after which the first-stage employment effect disappears. **Panel B (low-wage firms):** the first-stage employment effect never disappears for this sample, so no second line is shown.

a food-insecurity index summarizing the six post-baseline waves (Table 8, column 11), suggesting that the treatment did not operate through changes in maternal nutritional status either.

6 Discussion and Conclusion

This section discusses external validity and general equilibrium effects. We consider three questions: in which contexts do we expect the income effect to dominate the substitution effect? Which additional channels do we expect to matter in other contexts? How to reconcile the positive micro estimates with the macro observation that fertility declined as Ethiopia industrialized?

First, our model predicts that the income effect dominates the substitution effect when the monetary cost of a child is high relative to male income and when the maternal time cost of a child is low ($\frac{p}{I} > \phi$). These conditions appear to hold in our context. On the one hand, using our data on household expenditures in the control group at endline, we estimate that one additional child is associated with an increase in annual expenditures by 4,240 ETB in 2016 prices.[19] This amounts to $\frac{p}{I} = 16\%$ of total household expenditures. On the other hand, roughly half of respondents report that they could continue working at the factory with a newborn child. These women would have to quit the labor market for less than 1 year when they have a child, or $\phi = 5\%$ of their reproductive life.

How does Ethiopia compare to other low- and middle-income countries (LMIC) in terms of maternal time cost? The availability of informal childcare depends on whether cultural norms favor collective child-rearing or maternal investments. The starkest form of delegated childcare is fostering – a child living away from their mother altogether. It is common in Ethiopia: in the most recent EDHS, 13% of mothers have at least one child aged 0–14 currently living away from them.[20] This fraction is roughly twice as low as the average rate in SSA and twice as high as the average rate outside SSA (respectively 23% and 6% as shown in Appendix Figure F.1). The high prevalence of child fostering in SSA, ranging from 10% to 44%, makes it possible for women to have children *and* work, even in inflexible conditions. In contrast, in contexts where fertility responses to female employment have been shown to be negative, the fostering rate is low, e.g. 6% in Bangladesh and

[19] This number is close to Ethiopia's national poverty line of 7,184 ETB per adult per year (in 2016 prices), which translates into 3,500-4,500 ETB per child given the national adult-equivalence scale of 0.5-0.6.

[20] Over the life cycle, this fraction would be much higher because fostering is rarely permanent. Children go in and out of their mother's home depending on economic circumstances.

4% in India. To the extent that fostering proxies for the availability of informal childcare, we may expect substitution effects to be lower than in our study elsewhere in SSA and higher in other LMIC.

The income effect is consistent with macro evidence that the income–fertility relationship is positive at early stages of development (Vogl 2016). But is there causal evidence from other contexts that income effects can indeed be large? One way to cleanly quantify the income effect would be to leverage the large experimental literature evaluating cash transfer programs. Yet very few papers study births as an outcome of interest, perhaps due to reporting bias, data limitations, or a focus on child-level outcomes. Notable exceptions have produced mixed results, showing delays (Baird et al. 2011) or accelerations of birth timing (Carneiro et al. 2026; Walker et al. 2026) with no clear effect on completed fertility. Two differences with interventions raising women's *earned* income are worth noting, however. First, the earned income boosts are large relative to typical cash transfers. For example, women's factory wages amount to roughly a third of baseline household income in our sample; studies documenting positive fertility responses to improved self-employment report female income gains close to 30% (Berge et al. 2026; Donald et al. 2024). Second, factory employment may affect women's sense of control over their income stream and their ability to plan the sequencing of work and childbearing in ways that a passive transfer does not.[21] Viewed together, the studies summarized in Appendix Table F.1 suggest that positive fertility responses are most likely to emerge when substantial and stable income gains accrue to women who remain far from their desired fertility.

Second, two characteristics of our sample are important to keep in mind when extrapolating the effect to other contexts. The first one is that women in our experiment have already started their adult life at baseline: they live with a partner and look for a job. This shuts down the schooling and marriage channels that have been shown to matter when new employment opportunities became available to young women in India and Bangladesh (Jensen 2012; Heath and Mobarak 2015), or when educational interventions expanded future labor market opportunities, as in the secondary school scholarship experiment in Ghana (Duflo et al. 2024). More generally, studies documenting delayed fertility in response to female economic empowerment interventions typically target adolescents or unmarried young women and simultaneously affect schooling and marriage timing (Bandiera et al. 2020; Baird et al. 2011).

[21] Evidence from a different context points in this direction: exploiting a large regularization of undocumented immigrants in Spain, Elias et al. (2023) find that gaining access to stable formal employment significantly increased fertility among affected women.

The other important dimension is that female desired fertility is high. At baseline, women in our sample want four children on average, and only one woman in five reports that her partner wants more children than her. This leaves little room for increased female earnings to reduce fertility by changing bargaining power within the household. Women's ideal number of children varies starkly across LMIC, from above five to below two; it is substantially higher in SSA than in other countries, particularly among poorer households (Zipfel 2025). Recent work argues that this variation is rooted in family institutions governing marriage and inheritance that generate different incentives for women to want many children (Gobbi et al. 2026). With an ideal family size of four, our study population is at the lower end among SSA countries and at the upper end among other LMIC, suggesting that the bargaining power channel might be more relevant outside SSA.

Third, our study brings new evidence to debates on how the transformation of labor markets interacts with demographic change. In Ethiopia, fertility has declined relatively faster than in the average SSA country in recent decades. This rapid transition coincided with the ambitious industrialization policy, among many other changes. There are two ways to reconcile this observation with our positive experimental results. Either the macro correlation is driven by other transformations of the Ethiopian society. Or the macro effect of industrialization is negative whereas the micro effect of having an industrial job is positive. Conceptually, both effects are distinct: the micro effect is only one of the channels through which industrialization operates. When new factories open, they may shift prices, wages, networks and expectations in ways that affect fertility behavior among the broader population, including women who do not work in these factories. For example, improvements in female education documented in the literature reflect changes in expected returns to education.

This interpretation is supported by recent evidence from Kenya suggesting that the expansion of the cut-flower industry reduced fertility while simultaneously impacting local development (Bishop, Moneke and Zipfel 2026). More broadly, historical evidence shows that structural transformation can affect fertility through channels operating well beyond directly treated workers, including reductions in the economic value of children (Ager, Herz and Brueckner 2020). All these patterns are consistent with macro evidence that fertility responds positively to short-run income gains but negatively to long-run economic growth in developing economies (Chatterjee and Vogl 2018).

In conclusion, our study provides the first long-term experimental evidence on how formal wage employment affects women's fertility in SSA. Using individual-level randomization of factory job offers to partnered female jobseekers in Ethiopia, we show that access to a formal job increases fertility over an 8-9 year horizon. These effects are concentrated among women who were childless at baseline and take three years to emerge.

The positive fertility response diverges from conventional expectations about work-family trade-offs. We argue that income effects are likely to dominate forces driving fertility down in contexts where informal childcare is widespread and female desired fertility is high. These findings speak to debates about the drivers of SSA's "unique" fertility transition. As formal employment opportunities expand in the region's industrializing economies, understanding how new forms of work interact with family formation decisions becomes increasingly important. Our evidence suggest that the contribution of industrialization to fertility decline might be weaker than in other contexts.

References

Aalen, Lovise, Andreas Kotsadam, Janneke Pieters, and Espen Villanger, "Jobs and Political Participation - Evidence from a Field Experiment in Ethiopia," *The Journal of Politics*, 2024, *86* (2), 656–671.

Abebe, Girum, Niklas Buehren, and Markus Goldstein, "Do Factory Jobs Improve Welfare? Experimental Evidence from Ethiopia," *World Bank Economic Review*, 2025, *39* (1), 124–142.

Ager, Philipp, Benedikt Herz, and Markus Brueckner, "Structural Change and the Fertility Transition," *The Review of Economics and Statistics*, 10 2020, *102* (4), 806–822.

Ashraf, Nava, Erica Field, and Jean Lee, "Household Bargaining and Excess Fertility: An Experimental Study in Zambia," *American Economic Review*, July 2014, *104* (7), 2210–37.

Athey, Susan, Julie Tibshirani, and Stefan Wager, "Generalized random forests," *The Annals of Statistics*, 2019, *47* (2), 1148–1178.

_ **, Raj Chetty, Guido W. Imbens, and Hyunseung Kang**, "The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely," *The Review of Economic Studies*, 2025. Advance article.

Baird, Sarah, Craig McIntosh, and Berk Özler, "Cash or Condition? Evidence from a Cash Transfer Experiment," *The Quarterly Journal of Economics*, 2011, *126* (4), 1709–1753.

Bandiera, Oriana, Ahmed Elsayed, Andrea Smurra, and Céline Zipfel, "Young Adults and Labor Markets in Africa," *Journal of Economic Perspectives*, February 2022, *36* (1), 81–100.

_ **, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman**, "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa," *American Economic Journal: Applied Economics*, January 2020, *12* (1), 210–59.

Becker, Gary S. and H. Gregg Lewis, "On the Interaction between the Quantity and Quality of Children," *Journal of Political Economy*, 1973, *81* (2, Part 2), S279–S288.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 2014, *81* (2), 608–650.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Fortunata Makene, Linda Helgesson Sekei, Vincent Somville, and Bertil Tungodden, "On the Doorstep of Adulthood: Empowering Economic and Fertility Choices of Young Women," *The Economic Journal*, 2026.

Bishop, Menna, Niclas Moneke, and Celine Zipfel, "Female Wage Labour and Rural Development: Evidence from Kenya's Cut-Flower Boom," 2026. Working paper.

Blattman, Christopher and Stefan Dercon, "The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia," *American Economic Journal: Applied Economics*, 2018, *10* (3), 1–38.

Bongaarts, John, "Does malnutrition affect fecundity? A summary of evidence," *Science*, 1980, *208* (4444), 564–569.

_ , "Africa's Unique Fertility Transition," *Population and Development Review*, 2017, *43* (S1), 39–58.

Bouguen, Adrien, Yue Huang, Michael Kremer, and Edward Miguel, "Using Randomized Controlled Trials to Estimate Long-Run Impacts in Development Economics," *Annual Review of Economics*, 2019, *11*, 523–561.

Carneiro, Pedro, Lucy Kraftman, Imran Rasul, Francesca Salvati, and Molly Scott, "Accelerating birth timing to access cash transfers? Evidence from households in extreme poverty," *Journal of Health Economics*, March 2026, *106*, 103098.

Chatterjee, Shoumitro and Tom Vogl, "Escaping Malthus: Economic Growth and Fertility Change in the Developing World," *American Economic Review*, June 2018, *108* (6), 1440–67.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Iván Fernández-Val, "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments," Working Paper 24678, National Bureau of Economic Research 2023.

den Broeck, Goedele Van, Talip Kilic, and Janneke Pieters, "Structural Transformation and the Gender Pay Gap in Sub-Saharan Africa," *PLOS ONE*, 2023, *18* (4), e0278188.

Doepke, Matthias and Fabian Kindermann, "Bargaining over Babies: Theory, Evidence, and Policy Implications," *American Economic Review*, September 2019, *109* (9), 3264–3306.

_ **and Michèle Tertilt**, "Women's Empowerment, the Gender Gap in Desired Fertility, and Fertility Outcomes in Developing Countries," *AEA Papers and Proceedings*, May 2018, *108*, 358–62.

_ **, Anne Hannusch, Fabian Kindermann, and Michèle Tertilt**, "The Economics of Fertility: a New Era," in Shelly Lundberg and Alessandra Voena, eds., *Handbook of the Economics of the Family*, North-Holland, 2023.

Donald, Aletheia, Markus Goldstein, Tricia Koroknay-Palicz, and Mathilde Sage, "The Fertility Impacts of Development Programs," 2024. World Bank Policy Research Working Paper 10848; IZA DP 17500.

Duflo, Esther, Pascaline Dupas, Elizabeth Spelke, and Mark Walsh, "Intergenerational Impacts of Secondary Education: Experimental Evidence from Ghana," July 2024. Working paper.

Elias, Ferran, Joan Monras, and Judit Vall Castello, "Labor Market Stability and Fertility Decisions," Working Paper 2023-36, Federal Reserve Bank of San Francisco 2023.

ESHRE Capri Workshop Group, "Nutrition and reproduction in women," *Human Reproduction Update*, 2006, *12* (3), 193–207.

Ethiopia Central Statistical Agency and ICF, "Ethiopia Demographic and Health Survey 2016," Final Report, CSA and ICF, Addis Ababa, Ethiopia, and Rockville, Maryland, USA 2017.

Galor, Oded and David N. Weil, "The Gender Gap, Fertility, and Growth," *American Economic Review*, 1996, *86* (3), 374–387.

Gobbi, Paula E., Anne Hannusch, and Pauline Rossi, "Family Institutions and the Global Fertility Transition," *Journal of Economic Perspectives*, 2026, *40* (1), 47–70.

Goldin, Claudia, "The U-Shaped Female Labor Force Function in Economic Development and Economic History," in T. Paul Schultz, ed., *Investment in Women's Human Capital and Economic Development*, Chicago: University of Chicago Press, 1995, pp. 61–90.

Halvorsen, Sandra Kristine, "Labour Turnover and Workers' Well-being in the Ethiopian Manufacturing Industry," WIDER Working Paper 2021/36, United Nations University – UNU-WIDER February 2021.

Heath, Rachel and A. Mushfiq Mobarak, "Manufacturing Growth and the Lives of Bangladeshi Women," *Journal of Development Economics*, 2015, *115*, 1 – 15.

Jensen, Robert, "Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India," *The Quarterly Journal of Economics*, 03 2012, *127* (2), 753–792.

Jobs of the World Project, "Jobs of the World Project," https://jwp.iza.org/ 2026. Accessed June 2026.

Jones, Larry, Alice Schoonbroodt, and Michèle Tertilt, "Fertility Theories: Can They Explain the Negative Fertility-Income Relationship?," in John B. Shoven, ed., *Demography and the Economy*, University of Chicago Press, 2011, pp. 43–100.

Kotsadam, Andreas and Espen Villanger, "Jobs and Intimate Partner Violence - Evidence from a Field Experiment in Ethiopia," *Journal of Human Resources*, 2025.

Mammen, Kristin and Christina Paxson, "Women's Work and Economic Development," *Journal of Economic Perspectives*, December 2000, *14* (4), 141–164.

McMillan, Margaret and Albert Zeufack, "Labor Productivity Growth and Industrialization in Africa," *Journal of Economic Perspectives*, February 2022, *36* (1), 3–32.

Molina-Millán, Teresa and Karen Macours, "Attrition in Randomized Controlled Trials: Using Tracking Information to Correct Bias," *Economic Development and Cultural Change*, 2025, *73* (2), 811–834.

Rasul, Imran, "Household Bargaining over Fertility: Theory and Evidence from Malaysia," *Journal of Development Economics*, 2008, *86* (2), 215 – 241.

Schultz, T. Paul, "Changing World Prices, Women's Wages, and the Fertility Transition: Sweden, 1860-1910," *Journal of Political Economy*, 1985, *93* (6), 1126–1154.

United Nations, "Department of Economic and Social Affairs, Population Division. World Population Prospects, Volume I: Comprehensive Tables.," 2019.

Vogl, Tom S., "Differential Fertility, Human Capital, and Development," *Review of Economic Studies*, 2016, *83* (1), 365–401.

Walker, Michael, Nick Shankar, Edward Miguel, Dennis Egger, and Grady Killeen, "Can Cash Transfers Save Lives? Evidence from a Large-Scale Experiment in Kenya," 2026. NBER Working Paper No. 34152.

World Bank, "Labor Force Participation Rate, Female (% of Female Population Ages 15+) (Modeled ILO Estimate) — Ethiopia," World Development Indicators, series SL.TLF.CACT.FE.ZS, citing International Labour Organization, ILOSTAT database 2016. Accessed June 2026.

Zipfel, Céline, "The Demand Side of Africa's Demographic Transition: Desired Fertility, Wealth, and Jobs," *Working Paper*, 2025. Forthcoming, Journal of Human Resources.

APPENDIX

A Figures and Tables referred to in the text

Figure A.1: Industrial Parks

Notes: Location of industrial parks and 50 km buffer zones.

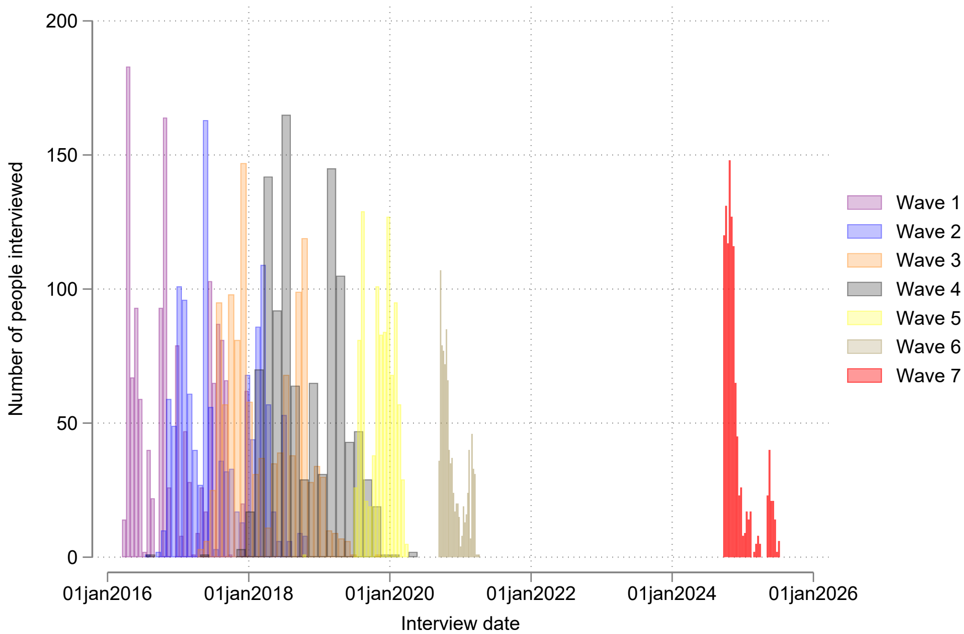


Figure A.2: Timing of interviews, by wave

Notes: The bars show the number of people interviewed over different waves and the colors show the different waves.

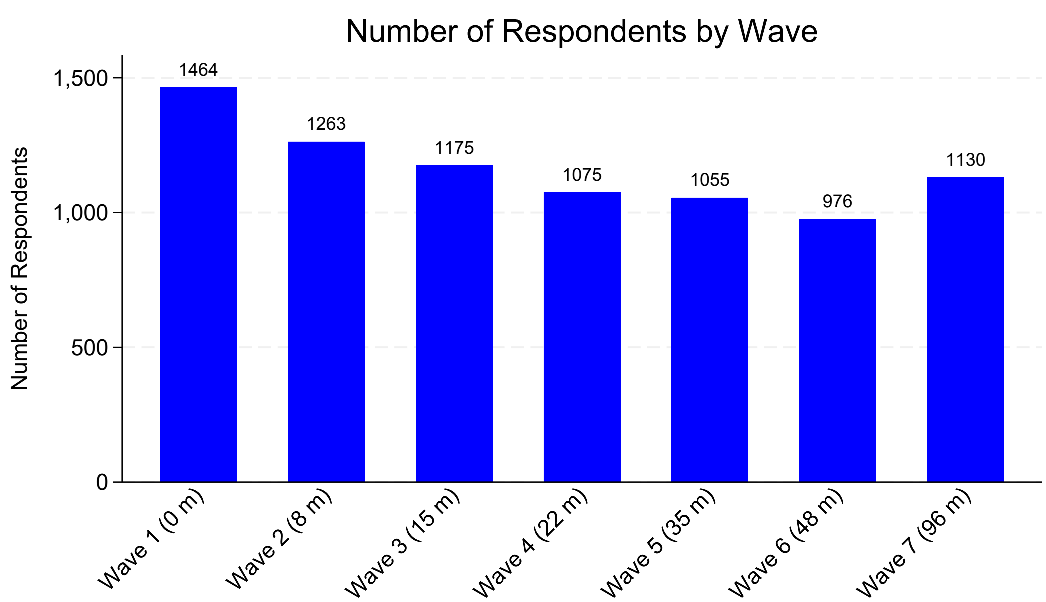


Figure A.3: Number of respondents, by wave

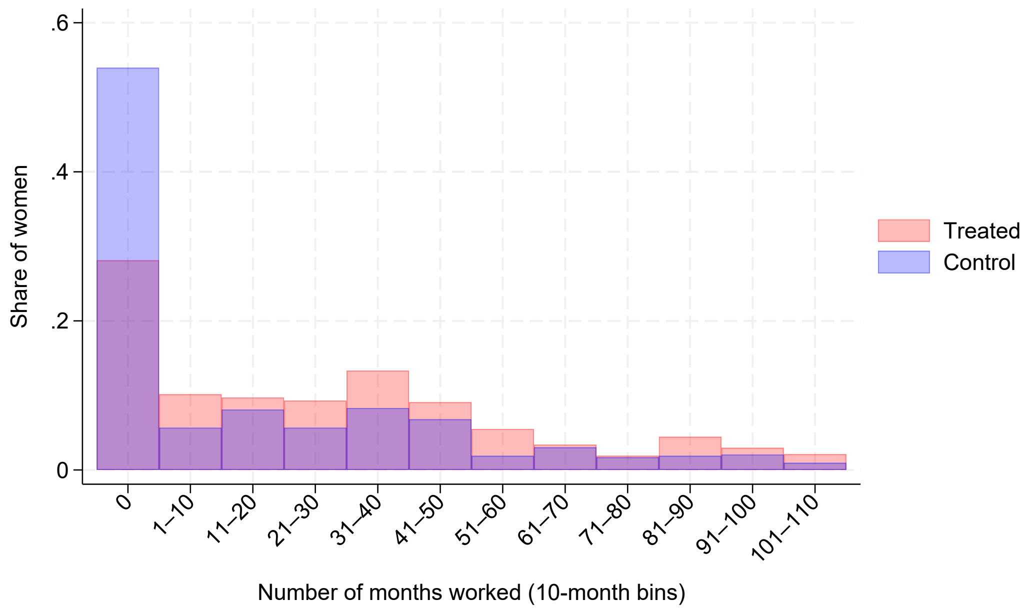


Figure A.4: Distribution of months worked (0, 10-month bins), by treatment status.

Table A.1: DHS 2016. Summary statistics. Married women aged 18-30 residing in an urban area within 50km of the industrial parks in the RCT sample.

	RCT baseline	EDHS 2016	Diff.
Any child	0.704	0.721	0.017
Number of children	1.281	1.225	-0.056
Partner wants more kids	0.211	0.172	-0.039
Wanted number of children	3.964	3.782	-0.182
Share in wage work	0.305	0.268	-0.038
Age	25.126	25.472	0.346
Muslim	0.138	0.230	0.092***
Protestant	0.234	0.072	-0.162***
Years of education	9.285	8.145	-1.140***
Observations	1130	750	

Table A.2: Blattman and Dercon (2018): Heterogeneity in employment effects, by baseline marital status.

	(1) Worked in any industrial firm ≥30 days	(2) Months worked in industrial firm	(3) In industrial firm at endline
Treatment	0.358***	2.547***	0.110**
	(0.046)	(0.432)	(0.046)
Married at baseline	-0.124*	-1.300**	-0.159***
	(0.075)	(0.569)	(0.054)
Treatment × Married at baseline	0.214**	1.731*	0.194**
	(0.094)	(0.911)	(0.092)
Control mean	0.379	2.475	0.185
R-squared	0.263	0.212	0.118
N	472	472	472
Share unmarried	0.797	0.797	0.797

Notes: Endline sample in Blattman and Dercon (2018), women only (=81% of sample). Robust standard errors in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table A.3: First-stage effects on employment, split by baseline motherhood status.

	(1) Months worked	(2) Months worked	(3) Months worked	(4) Months worked	(5) Any month worked	(6) Any month worked	(7) Any month worked	(8) Any month worked
Treatment	7.73***	6.95**	8.89***	6.95**	0.24***	0.17***	0.24***	0.17***
	(2.09)	(2.92)	(2.19)	(3.09)	(0.036)	(0.050)	(0.037)	(0.053)
Control mean	16.69	23.29	16.69	23.29	0.41	0.64	0.41	0.64
N	714	326	714	326	714	326	714	326
Sample	Mothers	Non-mothers	Mothers	Non-mothers	Mothers	Non-mothers	Mothers	Non-mothers
Controls	Lasso	Lasso	Necessary	Necessary	Lasso	Lasso	Necessary	Necessary

Notes: All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table A.4: Main fertility outcomes: alternative specifications

	(1) Any children	(2) Any children	(3) Any children	(4) Number of children	(5) Number of children	(6) Number of children
Treatment	0.041***	0.031*	0.038**	0.12**	-0.021	0.10*
	(0.016)	(0.018)	(0.017)	(0.057)	(0.088)	(0.059)
Control mean	0.90	0.90	0.90	2.51	2.51	2.51
N	1130	1130	1130	1130	1130	1130
Controls	Lasso	Necessary	Full	Lasso	Necessary	Full

Notes: All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table A.5: Effects on fertility outcomes over years.

	Panel A: Any child								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1 year	2 years	3 years	4 years	5 years	6 years	7 years	8 years	9 years
Treatment	-0.00026	0.0080	0.026	0.038**	0.049***	0.041**	0.052***	0.041**	0.042***
	(0.012)	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)
Control mean	0.76	0.80	0.82	0.84	0.86	0.87	0.89	0.90	0.90
N	1123	1123	1123	1123	1123	1123	1123	1123	1123
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso
	Panel B: Number of births								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1 year	2 years	3 years	4 years	5 years	6 years	7 years	8 years	9 years
Treatment	-0.056*	-0.026	0.0028	0.026	0.032	0.036	0.090*	0.095*	0.096*
	(0.031)	(0.036)	(0.040)	(0.043)	(0.045)	(0.048)	(0.052)	(0.055)	(0.057)
Control mean	1.56	1.70	1.84	1.99	2.11	2.23	2.35	2.46	2.49
N	1123	1123	1123	1123	1123	1123	1123	1123	1123
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso

Notes: All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table A.6: Any child. Sample of non-mothers at baseline.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1 year	2 years	3 years	4 years	5 years	6 years	7 years	8 years	9 years
Treatment	0.017	0.031	0.093*	0.14***	0.16***	0.14***	0.18***	0.15***	0.15***
	(0.036)	(0.046)	(0.050)	(0.049)	(0.050)	(0.050)	(0.050)	(0.049)	(0.048)
Control mean	0.16	0.28	0.38	0.45	0.50	0.56	0.60	0.64	0.65
N	350	350	350	350	350	350	350	350	350
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso

Notes: All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table A.7: Effect on having had a wage job last 6 months. Sample of non-mothers at baseline.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.040	0.28***	0.11*	-0.0014	0.034	-0.11	-0.071
	(0.039)	(0.055)	(0.060)	(0.062)	(0.066)	(0.067)	(0.051)
Control mean	0.48	0.44	0.50	0.59	0.53	0.52	0.35
N	1669	305	283	259	233	241	348
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso
Wave	2-7	2	3	4	5	6	7

Notes: The sample is restricted to non-mothers at baseline. All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table A.8: Time use effects in Wave 2.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Leisure	Eating	Soc. rel.	School	Sleeping	Travel	Unpaid	HH work	Personal care	Paid work
Treatment	-2.85***	-0.090	-0.54*	0.039	-0.52	0.60**	-0.56	-1.96***	-0.29**	10.2***
	(1.09)	(0.15)	(0.28)	(0.27)	(0.44)	(0.23)	(0.34)	(0.62)	(0.12)	(1.32)
Control mean	28.50	6.27	5.97	1.26	57.88	4.65	1.46	28.04	4.27	18.30
N	1261	1262	1262	1262	1262	1262	1262	1262	1262	1262
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso

Notes: The 9 columns cover the full W2 time-use module for the woman: leisure, eating, social/religious, school, sleeping, travel, unpaid work outside home, household work inside home, and personal care. The last column reports hours of paid work. All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table A.9: Heterogeneity in first-stage effect by baseline income quintile.

	(1) Wage job last 6 m.
Treatment (Q1 ref.)	0.15***
	(0.042)
T x Q2	-0.013
	(0.057)
T x Q3	0.014
	(0.060)
T x Q4	-0.038
	(0.058)
T x Q5 (richest)	0.025
	(0.058)
Q2	0.033
	(0.040)
Q3	0.017
	(0.040)
Q4	-0.024
	(0.040)
Q5 (richest)	-0.069*
	(0.039)
N	6669
Controls	Lasso

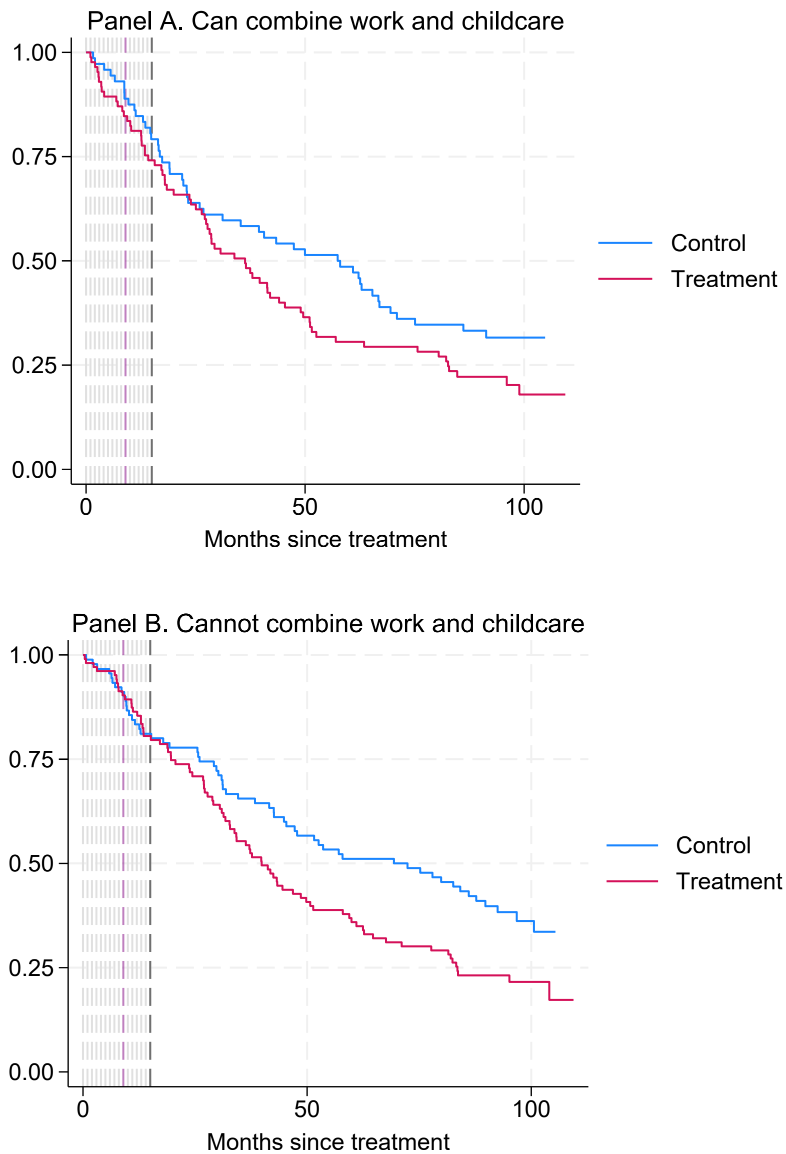
Reference quintile is Q1 (poorest). The interaction term "T $\times$ Q5 (richest)" tests whether the first-stage effect is smaller (or larger) in the top income quintile. Sample: all waves >1 (W2–W7). Clustered SE at the individual level.

Table A.10: Heterogeneity by baseline desired fertility.

	(1) Any child (young)	(2) Any child (all ages)
Treatment ($\leq$2 ref.)	-0.032	-0.018
	(0.048)	(0.036)
Wants 3–4	-0.064*	-0.048*
	(0.036)	(0.028)
Wants $\geq$5	-0.0098	-0.017
	(0.060)	(0.035)
T $\times$ Wants 3–4	0.10**	0.073*
	(0.053)	(0.039)
T $\times$ Wants $\geq$5	0.0070	0.018
	(0.072)	(0.046)
N	793	1130
Sample	Young (base_age<=25)	All ages
Controls	Lasso	Lasso

The "Young" column restricts to women who are in their early reproductive life at baseline (younger than 25 years old). Roughly 70% of the sample wants 3 or 4 children at baseline; 15% wants 2 or less; 15% wants 5 or more.

Figure A.5: Time to first birth among baseline non-mothers, by ability to combine work and child-care



Notes: Kaplan–Meier survival curves for time to first birth among women who were not mothers at baseline, by treatment arm. Panel A restricts to women who reported at endline that they could continue factory work with a newborn; Panel B to those who reported they could not (endline question `K1_53`). The dashed purple line at 9 months marks the gestation window: births before it were conceived before the job offer and serve as a balance check. The grey band (through 15 months) marks the period over which a detectable first-stage effect on employment persists for baseline non-mothers; the black dashed line at 15 months marks its end. Log-rank tests of the equality of the survival curves give $\chi^2(1) = 2.90$ ($p = 0.089$) among women who can combine work and childcare (Panel A, $N = 117$ births) and $\chi^2(1) = 5.69$ ($p = 0.017$) among those who cannot (Panel B, $N = 138$ births).

Table A.11: No effects on contraceptive use, Waves 2–6

	(1) Uses any method	(2) Uses any method	(3) Uses any method	(4) Uses any method	(5) Uses any method
Treatment	0.011	0.013	-0.016	-0.032	-0.010
	(0.025)	(0.027)	(0.029)	(0.031)	(0.030)
Control mean	0.69	0.64	0.65	0.62	0.57
N	1262	1174	1074	997	975
Controls	Lasso	Lasso	Lasso	Lasso	Lasso
Wave	W2	W3	W4	W5	W6

Notes: The outcome is an indicator for using any method to avoid pregnancy, collected in Waves 2–6 (not collected in Wave 7). All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table A.12: Aspirations for children

	(1) Education years	(2) Age of marriage	(3) Importance of own earnings	(4) Number of children
Panel A. Aspirations for daughters				
Treatment	0.19	0.17	0.11	0.027
	(0.13)	(0.19)	(0.086)	(0.10)
Control mean	16.72	23.41	9.01	4.34
N	1130	1129	1130	1129
Panel B. Aspirations for sons				
Treatment	0.073			0.055
	(0.12)			(0.10)
Control mean	16.77			4.46
N	1130			1129

Notes: Intent-to-treat effects, estimated with the paper's main specification (post-double-selection LASSO with block fixed effects partialled out and robust standard errors), on mothers' stated aspirations for their children, measured at endline. Panel A reports aspirations for daughters and Panel B for sons. The outcomes are the desired years of education, the desired age at marriage, how important the mother thinks it is that the daughter earns her own income, and the desired number of children. The age-at-marriage and own-earnings questions were asked only about daughters, so the corresponding cells are blank in Panel B. The control-group mean of each outcome is reported below the estimate. Standard errors are in parentheses; $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

B Data Appendix

This section is taken from the pre-analysis plan (Table 1). We describe the data collection and measurement of key variables in the Wave 7 survey.

Outcome variables

Our main fertility outcomes come from the fertility module, in which we collected the dates of all live births. *Any child* is an indicator equal to one if the woman has ever given birth to a living child, and *Number of children* is her reported number of live births. We capture fertility preferences with three further variables: the *appropriate age of first birth*, i.e. the age the woman considers a good age for a woman to have her first child; the *wanted number of children*, i.e. the number she would choose to have over her whole life, asked as if she could go back to having no children; and an indicator for the *partner wanting a higher number of children* than the woman herself.

On the employment and income side, *any current wage job* is an indicator for positive income from factory or other wage employment over the last six months, and *earnings from wage job* is the corresponding total in Birr. We construct the woman's *share of earnings* as her wage earnings over the last six months divided by her husband's, bounded between zero and one, and *she earns more than him* as an indicator for that share exceeding 0.5. For each of these we also construct the corresponding variables restricted to factory jobs and, separately, to earnings and income from any source.

We also use two time-use outcomes, each measured as hours over the last seven days: time spent on *social and religious activities* and time spent on *leisure* (watching TV, reading, playing, exercising, recreation, and the like).

Baseline covariates

From the baseline survey we use the woman's age in years; her religion (Orthodox Christian, Catholic, or other; Muslim; or Protestant); her education, grouped as low (fewer than 10 years), medium (10 years), or high (more than 10 years); and an indicator for whether she had ever held a formal wage job. We also retain the baseline values of the fertility outcomes themselves — parity, the wanted number of children, the appropriate age of first birth, and whether the partner wanted more children — together with baseline leisure and social and religious time use.

Control sets

We always include list (block) fixed effects, since job offers were randomized within lists. Beyond these, we report three control specifications. The *necessary* specification includes the list fixed effects only. The *full* specification adds the complete set of baseline covariates described above, each entered together with a missing-value indicator: the covariate is set to zero when the baseline value is missing and a dummy flags the imputation, so that no observation is dropped on account of an unanswered baseline question. Continuous covariates — age, leisure, and social and religious time — enter as quintile dummies. Our main specification instead selects controls from this same pool using a post-double LASSO procedure (Belloni et al. 2014), always partialling out the list fixed effects. LASSO retains the covariates that predict either treatment or the outcome (the union of the two selected sets), which improves precision and helps correct for the modest baseline imbalances in parity, religion, and education.

C Attrition

Table C.1: Attrition in Wave 7

	Respondent missing in Wave 7		
	(1)	(2)	(3)
Treatment	-0.030	-0.033	-0.029
	(0.023)	(0.032)	(0.023)
Number of children		-0.045***	
		(0.013)	
Number of children*Treatment		-0.0021	
		(0.016)	
Age (std)			-0.076***
			(0.016)
Age (std)*Treatment			0.020
			(0.020)
Control mean	0.24	0.24	0.24
N	1463	1463	1463
Controls	Necessary	Necessary	Necessary

Notes: The fertility variables are based on the variable "how many children do you have" since we did not have the fertility module at baseline. Necessary controls are list fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Even though attrition is not correlated with treatment, we still face 23% of our baseline sample that could not be reached at endline. Even in the absence of differential attrition by treatment status, attrition can bias estimates if treatment effects are heterogeneous and those who attrit differ systematically from those who remain. In our case, these women are likely more mobile, more ambitious, or have better employment networks—traits that may be correlated with fertility outcomes.

We implement two robustness exercises: (i) a sample-restriction comparison that varies how easily respondents were tracked, mirroring Molina-Millán and Macours (2025)'s logic that harder-to-track respondents are informative about the still-missing non-respondents; (ii) an inverse-probability weighting (IPW) correction following the same paper, implemented within the 1,130 Wave-7 endline sample using phone vs in-person interview as the harder-to-track signal.

Results for "any children" are reported in Table C.2. Column 1 reproduces the main estimate for comparison. Column 2 restricts the sample to respondents interviewed in person, while column 3 restricts to the 127 respondents whom we were only able to interview by phone. Because phone interviews were only conducted when an in-person meeting was not feasible, this group represents respondents who were more difficult to reach. The point estimate for the phone-interview sample is larger, though imprecisely estimated. This suggests that, if anything, the coefficient in column 1 might underestimate the treatment effect we would obtain if we had been able to track all respondents. In column 4, we implement the IPW correction: the point estimate slightly increases, and remains significant at 10%. The results for "number of children" in Table C.2 show a similar pattern. Taken together, these results make a substantial upward attrition bias unlikely.

Table C.2: Attrition: robustness tests

Panel A. Any children				
	(1) Full Wave 7	(2) In person	(3) Phone	(4) IPW
Treatment	0.041***	0.037**	0.066	0.043*
	(0.016)	(0.016)	(0.062)	(0.026)
Control mean	0.90	0.91	0.80	0.87
N	1130	1003	127	1130
Controls	Lasso	Lasso	Lasso	Lasso
Sample	Wave 7	W7 not phone	W7 phone	Wave 7
Panel B. Number of children				
Treatment	0.12**	0.12**	0.19	0.15**
	(0.057)	(0.061)	(0.16)	(0.076)
Control mean	2.51	2.57	2.02	2.41
N	1130	1003	127	1130
Controls	Lasso	Lasso	Lasso	Lasso
Sample	Wave 7	W7 not phone	W7 phone	Wave 7

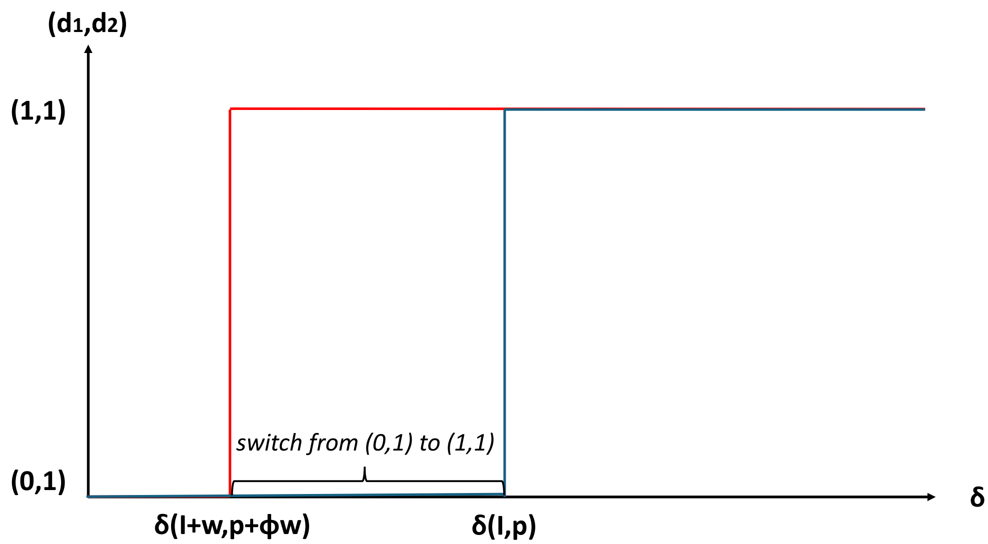
Notes: Panel A reports effects on having any children, Panel B on the number of children. Column 1 is the full Wave-7 sample; column 2 restricts to in-person interviews; column 3 to phone interviews; column 4 applies the inverse-probability weighting correction. All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

D Dynamic model

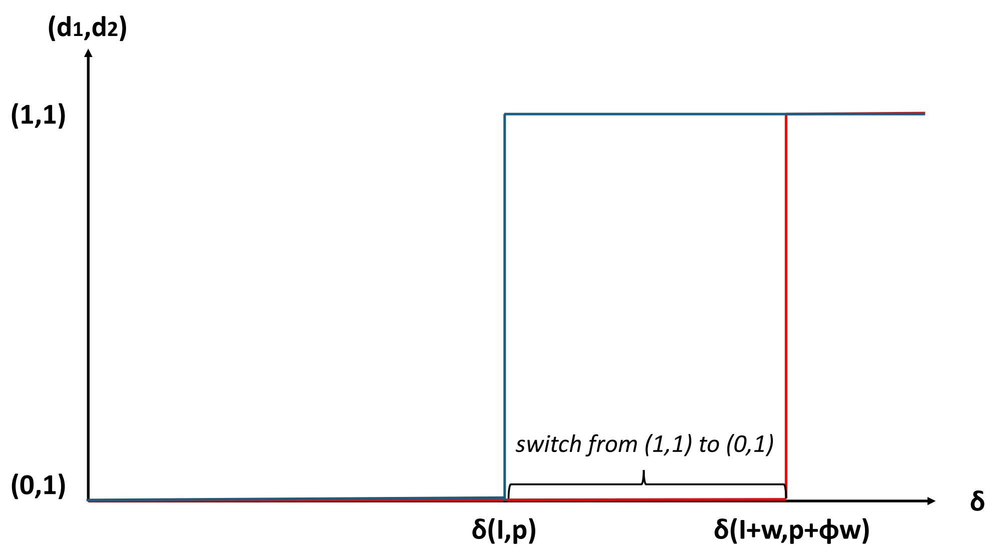
Figure D.1 illustrates how our experiment changes the cutoff in the dynamic model. The lines represent the optimal (d_1, d_2) as a function of δ, separately in the control group ($w = 0$, in blue) and in the treatment group ($w > 0$, in red). When the income effect dominates, the cutoff is lower in the treatment group and more couples have a child in period 1 ($d_1 = 1$). The opposite is true when the substitution effect dominates. The magnitude of the treatment effect increases with w.

Figure D.1: Illustration of the effect on the timing of first births

(a) Income effect dominates: $\frac{p}{I} > \phi$



(b) Substitution effect dominates: $\frac{p}{I} < \phi$



To understand why we observe sequential effects – first on labor supply, then on fertility – let us focus on period 1 and model the optimal birth timing within this period among couples who have decided to have a child. We consider a continuous time framework with a spell length normalized to 1. We write the optimization as follows:

$$\max_b \int_0^1 log(c(t))dt$$

$$\text{s.t. } \dot{a}(t) = I(t) + a(t) - c(t)$$
$$\text{and } \text{a(t)} \geq 0$$
$$\text{and } \text{I(t)} = \{ {}^{I \text{ if } t\in[b;b+\phi]}_{I+w \text{ otherwise}}$$
$$\text{and } \text{b} \leq 1-\phi$$

Couples choose the optimal birth timing b in order to maximize their total consumption. Suppose that women cannot work and care for the child simultaneously. When a child is born, the mother has to stop working for ϕ units of time. The total household income earned at the end of the period is equal to $I + (1-\phi)w$. Couples can save but not borrow: $a(t)$, their accumulated assets at date t, cannot be negative. .

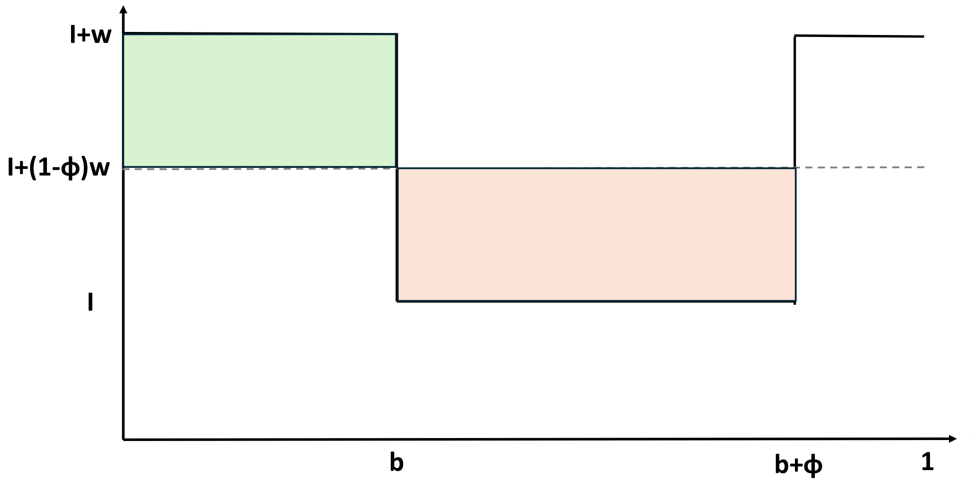
The utility is maximized when the instantaneous consumption is constant over time, that is $c(t) = I + (1-\phi)w$. To smooth consumption perfectly, couples have to save at least as much between $t = 0$ and $t = b$ as they dissave between $t = b$ and $t = b + \phi$. Formally:

$$((I + w) - (I + (1-\phi)w)) \times b \geq ((I + (1-\phi)w) - I) \times \phi$$

$$\Leftrightarrow b \geq 1-\phi$$

Together with the feasibility constraint, this implies that $b^* = 1-\phi$. Figure D.2 provides a graphical illustration. The solid lines represent the flows of income and the dashed lines represent the flows of consumption as a function of time. Income flows depend on the birth timing b. In order to keep consumption constant, the green area has to be at least as large as the orange area. This is only feasible if women work first during $(1-\phi)$ unit of time, in order to accumulate savings, and then exit the labor market for the remaining ϕ unit of time in order to have children.

Figure D.2: Illustration of the sequential effect



E Heterogeneity: high- versus low-wage firms

Correlation between firm wage, other firm characteristics and treatment

Based on survey data with managers collected in 2017 we also have other characteristics of the factories. The higher paying factories are more likely to be Ethiopian, less likely to have conflicts at the workplace, have slightly fewer share of female workers, and have operated for longer. Running a regression of

$$\text{Treatment} \times \text{High wage firm}_{if}$$

on all the baseline characteristics while controlling for block fixed effects shows that the samples seem to be balanced with respect to worker characteristics.

Table E.1: Other characteristics of high- and low-wage factories

	All		High wage firm		Low wage firm	
	(1)		(2)		(3)	
	Mean	SD	Mean	SD	Mean	SD
Ownership						
Asian	0.35	(0.48)	0.14	(0.34)	0.56	(0.50)
American	0.03	(0.16)	0.00	(0.00)	0.05	(0.22)
Ethiopian	0.63	(0.48)	0.86	(0.34)	0.39	(0.49)
Other factory characteristics						
No conflicts	0.72	(0.45)	0.79	(0.41)	0.66	(0.47)
Share female workers	0.72	(0.18)	0.70	(0.19)	0.73	(0.17)
Production start year	2008.77	(8.13)	2004.44	(8.02)	2013.18	(5.43)
N	1399		698		701	

Notes: The sample contains women working in factories with more than 10 observations. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table E.2: Balance of worker characteristics

	(1)
	Treatment*High wage firm
Any child	-0.029
	(0.033)
Number of children	0.011
	(0.0067)
Partner wants more kids	-0.0067
	(0.027)
Wanted number of children	-0.0015
	(0.0044)
Appropriate age of first birth	0.0021
	(0.0038)
Any wage job ever	0.024
	(0.025)
Age	-0.0034**[*]
	(0.0017)
Muslim	-0.041
	(0.040)
Protestant	-0.0057
	(0.034)
Years of education	0.0027
	(0.0038)
Leisure time last 7 days (hours)	0.000064
	(0.00057)
Social and religious activities last 7 days (hours)	-0.0024*[*]
	(0.0015)
N	1102
F-value	1.04
P-value F-test	0.40

Notes: All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$. Stars in [] correspond to differences when the variables are run one by one, only including block fixed effects.

E.1 Heterogeneous treatment effects on various outcomes

Table E.3: Main fertility outcomes Wave 7. Heterogeneity by high or low paying firms

	(1) Any children	(2) Any children	(3) Number of children	(4) Number of children	(5) Number of children	(6) Wanted nr children	(7) Appropriate age first child
Treatment	-0.014	-0.056	0.012	0.020	0.080	-0.026	0.22
	(0.014)	(0.086)	(0.086)	(0.094)	(0.24)	(0.18)	(0.28)
Treatment*High wage firm	0.11***	0.26**	0.22*	0.12	0.28	-0.19	-0.090
	(0.031)	(0.10)	(0.11)	(0.14)	(0.27)	(0.22)	(0.37)
Control mean	0.90	0.65	2.50	3.07	1.11	4.84	23.40
Sample	All	Non-mothers	All	Mothers	Non-mothers	All	All
N	1113	345	1113	761	345	1113	1113
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso

Notes: All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$. Treatment $\times$ High wage firm is defined as treatment interacted with an indicator for factories with above-median mean basic monthly wages (median computed among factories with more than 10 observations), where wages are measured in wave 2 using the question "How much is the basic payment (salary/wage) per month from this job?"

Table E.4: Effects on having had a wage job last 6 months. High-paying firms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.12***	0.37***	0.24***	0.14***	0.12**	-0.016	-0.11***
	(0.031)	(0.042)	(0.045)	(0.047)	(0.050)	(0.052)	(0.041)
Control mean	0.43	0.39	0.44	0.49	0.47	0.47	0.38
N	2754	504	470	440	400	394	546
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso
Wave	2-7	2	3	4	5	6	7

Notes: All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table E.5: Effects on having had a wage job last 6 months. Low-paying firms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.18***	0.40***	0.26***	0.19***	0.12***	0.072*	0.092**
	(0.028)	(0.038)	(0.042)	(0.045)	(0.046)	(0.043)	(0.039)
Control mean	0.22	0.19	0.20	0.24	0.23	0.22	0.25
N	2967	517	505	471	449	465	560
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso
Wave	2-7	2	3	4	5	6	7

Notes: All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table E.6: Effects on cumulative income and savings: high vs. low-wage paying firms.

	(1) Savings (High wage)	(2) Savings (Low wage)	(3) HH income (High wage)	(4) HH income (Low wage)
Treatment	3896.1**	-1142.4	6417.1*	1740.6
	(1939.8)	(4627.2)	(3735.9)	(5868.8)
Control mean	11416	19687	73037	99719
N	673	658	673	658
Controls	Lasso	Lasso	Lasso	Lasso
Sample	High wage	Low wage	High wage	Low wage

Notes: All regressions control for block fixed effects. Standard errors are in parentheses. Outcomes are cumulative W2–W7 sums (six 6-month windows). P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table E.7: Heterogeneity by within-region high-wage status.

	(1) Any children	(2) Any children	(3) Number of children	(4) Number of children	(5) Number of children	(6) Wanted nr children	(7) Appropriate age first child
Treatment	-0.023	-0.082	0.0057	0.094	-0.13	0.023	0.14
	(0.022)	(0.11)	(0.098)	(0.11)	(0.27)	(0.19)	(0.35)
Treatment*High wage firm (within region)	0.087***	0.26**	0.15	-0.048	0.50*	-0.20	0.052
	(0.030)	(0.12)	(0.12)	(0.14)	(0.29)	(0.23)	(0.41)
Control mean	0.90	0.65	2.51	3.07	1.11	4.83	23.41
Sample	All	Non-mothers	All	Mothers	Non-mothers	All	All
N	1129	350	1129	772	350	1129	1129
Controls	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Lasso

Notes: "High-wage WITHIN region" is an indicator that factory mean wage exceeds the median wage of factories in the same baseline region. The main effect of the within-region high-wage indicator is omitted from the table; only Treatment and Treatment×High-wage(within) are reported. All regressions control for block fixed effects. Standard errors are in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

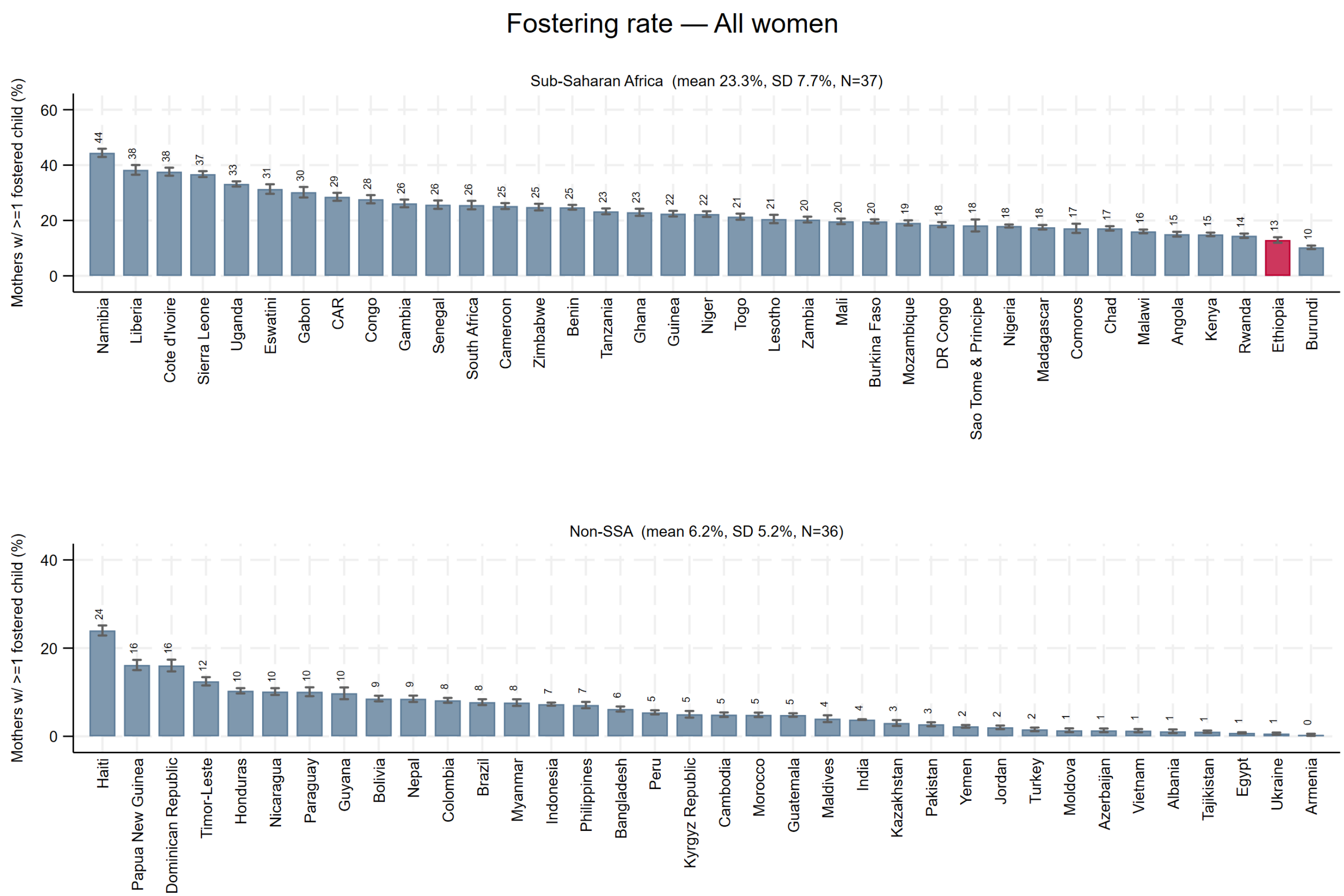
F External Validity

Table F.1: Relative income gains and fertility responses in related experiments in sub-Saharan Africa.

Study (setting)	Sampling frame	Income gains	Fertility effect (control mean; parity heterogeneity)
Panel A. Income effect dominates: positive fertility response			
Berge et al. 2026 (Tanzania, entrepreneurship)	Very young women, secondary-school leavers; unmarried at baseline; high-fertility, self-employment context	Earned (self-employment); total income +27–42% (long-run ≈+27%)	+9.5pp early pregnancy ($p < 0.001$); +7.1pp started childbearing long-term (CM ≈0.24). Low-parity entry margin; heterogeneity by baseline SES
Donald et al. 2025 – DOT [Alibhai et al. 2019] (Ethiopia)	Female micro/small entrepreneurs, urban Mekelle; mean age ≈32; ≈1.4 employees	Earned business profits, +30%	+0.11 children (CM 2.33), +5%; +6pp additional child (CM 0.27), +22%. Stronger for women without a son / poorer
Donald et al. 2025 – PI [Campos et al. 2017] (Togo)	Entrepreneurs (≈half male) in business ≥12 mo, non-agricultural, <50 employees; Lomé	Earned business profits, +30%	+0.39 children (CM 0.57, under-5 proxy); +14pp additional child (CM 0.14), ≈ 2×. Strongest for women without a son
Panel B. No fertility response			
Walker et al. 2026 (Kenya, GiveDirectly UCT)	Poor rural households (10,500+ HHs, 653 villages), western Kenya; general population (already mothers, higher parity)	Unearned one-time UCT; USD 1,000 lump sum (≈75% of annual HH expenditure)	Transient +10% births during transfer period only; completed and total fertility unchanged. Child mortality ↓ −48% infant, −45% under-5 (timing/maternal-health channel, not selection)
Carneiro et al. 2025 (N. Nigeria, CDGP)	Extremely poor women, pregnant/non-pregnant at baseline; high-fertility	Unearned transfer to pregnant mothers; $22/mo PPP (=12% of HH monthly earnings)	Weak acceleration of birth timing only; no clear completed-fertility effect. Population already mothers (higher parity)
Donald et al. 2025 – WALN [Bakhtiar et al. 2022] (Ethiopia)	Established female business owners (≥1 employee; ≈3.6 employees, ≈72,000 birr/mo revenue); most educated sample	Earned (business/mentoring); +0.20 SD profit/sales index	Null (−0.16/−0.19, n.s.; CM 3.59). High baseline parity (≈1.3 more children than DOT controls)
Duflo, Dupas, Spelke & Walsh 2024 (Ghana, secondary scholarships)	Adolescents passing entrance exam but unable to afford fees; 4 yrs free secondary	Earned (post-education wages); women: small earnings gains, mostly via public-sector job access	Women −6.9pp ever pregnant (CM 48%); −0.15 children; teen pregnancy ↓ 17%; total fertility catches up later. Intergenerational: −45% under-3 mortality, +0.25 SD child cognition. Schooling channel
Buehren et al. 2024 (Tanzania, ELA replication)	Adolescent girls (3,100+), unmarried; same ELA program	Earned (intended) – none materialized; no impact on any economic outcome	None (no fertility/health/social effect)
Panel C. Negative fertility response			
Bandiera et al. 2020 (Uganda, BRAC ELA)	Adolescent girls (≈14–20), unmarried; vocational + reproductive-health/life-skills	Earned (self-employment), but tiny: +$50/yr earnings; 2× as likely to run an income-generating activity	Teen pregnancy ↓ ≈34%; early marriage ↓ ≈62% (relative). Low-parity adolescents
Baird et al. 2011 (Malawi)	Never-married girls 13–22, in/out of school	Unearned (CCT school-conditional + UCT); ≈$10/mo × 10 mo/yr (≈+10% HH consumption)	Ever-pregnant ↓ ≈27–30%; marriage ↓ ≈40% (relative; largest in UCT, out-of-school). Low-parity adolescents

Notes: CM = control-group mean of the fertility outcome; pp = percentage points. For the adolescent-girl programs, fertility effects are reported as the relative reductions stated in the source papers.

Figure F.1: Share of mothers with at least one fostered child, by DHS country



Notes: Mother-level fostering rates (means and 95% confidence intervals) computed from the latest DHS wave of each of the 73 countries in the sample of Zipfel (2025). The denominator is mothers (women aged 15–49) with at least one living child aged 0–14 whose residence status is known; mothers without such a child (e.g., those whose children are all 15 or older, or who are childless) are excluded. The numerator is mothers with at least one such child who does not reside in their household, whether the child lives with their father, other relatives, or elsewhere.