Children and Relationship Quality*

Belén Rodríguez Moro† Olatz Román Blanco‡

August 14, 2023

WORK IN PROGRESS – DO NOT CIRCULATE

Abstract

We examine the impact of having a child on couples' relationship quality (RQ), defined as the non-pecuniary gains from being in a relationship. Adopting a pseudo-experimental approach, we perform an event study analysis around first child birth and find a sharp and persistent decrease in RQ for both fathers and mothers immediately after birth. Individuals ranking in the 90th percentile of RQ before child birth are pushed down to the median. We attribute this effect primarily to changes in time use. First, a decrease in time spent together as a couple can explain half of the decrease in RQ. Second, we document a substantial increase in unpaid housework absorbed by women. We uncover heterogeneity in the impact of first child birth on RQ based on the division of work before birth, with women experiencing larger increases in unpaid housework also suffering a larger decrease in RQ after first child birth. Using a state-funded childcare expansion, we estimate the causal effect of changes in time use on RQ. Preliminary estimates show that households benefiting the most from this expansion suffered the largest drops in RQ. We also find that mothers increase their labor supply in a larger magnitude that their decrease in housework time. Upon former analyses, this evidence could suggest a link between over-load mothers and larger drops in RQ.

Keywords: Fertility, Marital decisions, Time allocation, Relationship quality

JEL Codes: J12, J13, J22, D13

*We are grateful to our supervisors Sule Alan, Thomas F. Crossley and David K. Levine for their invaluable guidance and support. We would also like to thank Johanna Reuter (discussant) and seminar and conference participants at the European University Institute, Universitá di Firenze, Universitat Pompeu i Fabra, the 18th Doctorissimes Conference, Universitat de Barcelona, Universidad Carlos III de Madrid, Universidad de Salamanca, the BSE Jamboree, the EDP Jamboree and the Understanding Society Scientific Conference for helpful comments and suggestions.

†European University Institute, belen.rodriguez@eui.eu.

‡European University Institute, olatz.roman@eui.eu.

1 Introduction

Fertility rates have experienced a structural change over the last few decades. Fewer individuals have children and the ones who do have fewer of them. This demographic transition has led governments to introduce a number of policies to reverse the fertility decline. However, the impact that having children may have on couple's lifestyle and couple decision-making remains unclear, despite being key to assessing the effectiveness of these policies.

This paper delves into one aspect of the impact of children on couples, examining the causal effect of having a child on relationship quality. Relationship quality refers to the non-pecuniary benefits individuals experience from being in a relationship, which strongly influence marital decisions. If relationship quality is negatively impacted, separation becomes more likely. A priori, the link between having children and relationship quality is ambiguous. On one hand, having a child may increase general happiness, leading to higher relationship quality. However, having a child may also create challenges such as having less time to spend together, financial stress and increased domestic work, affecting relationship quality negatively.

To pin this effect down, we construct a novel measure of relationship quality (RQ). We use a questionnaire periodically asked in Understanding Society, the UK household longitudinal panel. Compared to other data sources, Understanding Society combines three key advantages: it collects rich information on different qualitative aspects of individuals' relationships; it interviews each member of the couple individually, allowing for within-couple comparisons; and it follows individuals at different stages of their relationships (unlike other data-sets that rely on recalled data), being able to validate this measure with other observed outcomes. We categorize the items of this questionnaire into two blocks, depending on the information they convey: (i) subjective assessments of the quality of the relationship, such as considering divorce or happiness with the relationship and (ii) couple time use, such as engaging in outside activities together or having stimulating exchanges of ideas. We construct a unified measure of RQ combining all this information through factor analysis and use it for our main analysis. We similarly construct two sub-measures, Subjective RQ and Couple time RQ, using the information of each block separately. We leverage variation in the timing of first child birth and use an event study analysis as our main specification to evaluate the dynamic effect of children on RQ.

We find a sharp decrease in RQ immediately after birth. Illustratively, individuals who ranked in the 90^{th} percentile of RQ before having their first child are pushed down to the median RQ within the first three years after birth. This negative impact persists over the observation period, never recovering the initial values, and it is consistent for both mothers and fathers. The results are robust to using alternative samples, specifications and measures of relationship quality.

We explore the mechanisms driving this effect, which operate through differences in home production after birth. Children increase the workload of parents, creating childcare needs and increasing domestic work. We interpret this as a time shock and argue that impacts couple behaviors in non-negligible ways and consequently influences RQ. First, we document a sizeable increase in unpaid domestic work after child birth, excluding childcare. We find that this increase is almost fully borne by women, regardless of the division of paid and unpaid work before child birth. Women in couples that shared tasks equally before child birth report the greatest increase in domestic work. We exploit the variation in the redistribution of tasks after first child birth to uncover heterogeneity in the impact of fertility on RQ: women with larger increases in the share of unpaid housework also experience larger decreases in RQ.

Finally, we use the expansion of free childcare from 15 to 30-hours a week introduced in England in 2017 to establish a causal relationship between changes in household specialization and RQ. This policy is targeted to parents of three- and four-years-olds and aims to alleviate the cost of childcare for working parents, explicitly requiring that both parents work at least the equivalent of sixteen weekly hours at the minimum pay. We use two methodology strategies to leverage quasi-random variation in the availability of free additional childcare. First, we exploit geographic variation in the effective implementation of the policy and compare parents living in fast-adopter regions with parents living in slow-adopter regions. Preliminary results show that the RQ of couples living in the former regions deteriorates significantly more. This result is robust to the inclusion of individual, couple and regional controls. Our results also confirm that fast-adopter regions had higher take-up rates and that mothers living there significantly increase their labor supply by 9-7 hours a week. Although there is suggestive evidence of a decrease in the number of housework weekly hours provided by mothers, the magnitude of this effect is substantially smaller and non-significant. Altogether, our results suggests that mothers may have become over-loaded, increasing their time spent on paid work without reducing housework responsabilities. However, we are still cautious when interpreting these results. As next step, we plan to carry a Regression Discontinuity Desing that exploits natural discontinuities formed around the school starting terms and children's month of birth to test the robustness of our results.

Related literature. The first contribution of this paper is introducing a novel measure of relationship quality into the economics literature. Psychologists and sociologists have already studied similar measures (see, for example, Carlson and VanOrman, 2017 in sociology and Joel et al., 2020 in psychology). However, the larger sample and longitudinal dimension of our data enable us to use causal identification methods that were not feasible before. We are able to lay out a newly discovered fact with large consequences on household decision making.

This paper also adds to the study of the consequences of fertility. This literature, largely led by Claudia Goldin, has mostly focused on documenting the disparity in the impact of children on mothers' and fathers' labor market outcomes (see Goldin, 2021, among many others). For instance, Kleven et al. (2019) find sizeable effects of first child birth on mothers' labor force participation and earnings, while fathers' outcomes remain unchanged. Cortés and Pan (2020) show that this disparity can explain a large share of the remaining gender gaps in the labor market. Other authors have studied the impact of children on more subjective outcomes, such as general happiness (Korsgren and van Lent, 2020). This paper studies the effect of fertility on the subjective component of coupled individuals' welfare.

Furthermore, this paper speaks to the relatively recent literature on household time allocation. This issue has received great attention during and after the COVID-19 pandemic, which was an unprecedented shock to childcare (Sevilla and Smith, 2020; Hupkau and Petrongolo, 2020; Alon et al., 2020). The empirical findings of these papers support the recent explanations of household specialization in which gender identity play a central role (Akerlof and Kranton, 2000). Our results show that the arrival of children reinforces traditional views of gender identity, inducing household specialization even among couples that had an equal division of tasks before child birth. This result is independent of the gender attitudes reported by the couples.

Finally, the empirical observation of this measure provides relevant insights to the literature on the welfare gains of family formation. Standard household models acknowledge the relevance of relationship quality in the decision-making process of households (see Greenwood et al., 2017 or Chiappori and Mazzocco, 2017 for a recent survey of the literature). However, these models are undecided on whether this measure follows a learning or a stochastic process, and in many cases they assume it is uncorrelated with past events. Our main result shows that this is not the case. The introduction of this measure opens a new empirical research line in the family economics literature that can guide future advances in family economics modelling.

Roadmap. The rest of the paper is organized as follows. Section 2 describes the dataset used and presents the measure of relationship quality. Section 3 describes the event study approach. Section 4 presents the main results. Section 5 explores the potential mechanisms at play. Section 6 concludes.

2 Data

2.1 Dataset

We combine data from the British Household Panel Survey (BHPS) and Understanding Society (University of Essex, Institute for Social and Economic Research, 2022). The BHPS is a longitudinal household panel containing around 10,000 households and covering the period 1991-2008. In 2009 it was replaced by Understanding Society, which includes 8,000 voluntary BHPS households and 40,000 new households. The survey is still running and 12 waves have been released, until the year 2021.

This dataset is particularly valuable to answer the question at hand. First, it contains a questionnaire with a rich set of questions about individuals' relationships, which allows us to pin down the two mechanisms considered. Second, it consists of a longitudinal panel, which allows us to follow individuals at different stages of their relationship, and relate the different measures of relationship quality to observed outcomes such as marriage and divorce.

2.2 Measure of Relationship Quality

Every other wave of Understanding Society includes an individual 10-item questionnaire asking about the relationship with their partners. Most items refer to behaviors such as "How often do you and your partner quarrel?", which have answers ranging from "All of the time" to "Never" on a six-point Likert scale.[1] The module also includes a question about the degree of happiness with the couple and about shared outside interests. Table 1 contains the full set of items. These questions are asked individually to all respondents who are cohabiting with their partners, whether they are married or not. This information is available for the period 2009-2021. Which waves?

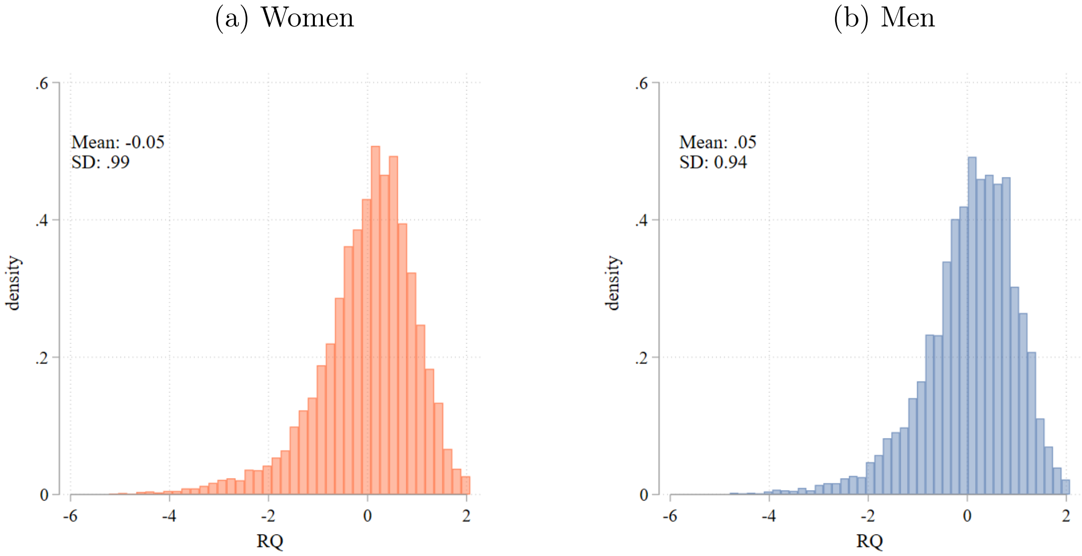
Table 1: Questions in the Understanding Society Partner module.

(a) Subjective assessment		(b) Couple time use	
How often do you... ?		*How often do you... ?*	
consider splitting	6pt, freq (-)	work together on a project	6pt, freq (+)
regret getting married	6pt, freq (-)	stimulating exchange of ideas	6pt, freq (+)
quarrel	6pt, freq (-)	calmly discuss something	6pt, freq (+)
get on each others nerves	6pt, freq (-)	kiss partner	6pt, freq (+)
What is the... ?		*Do you and your partner... ?*	
degree of happiness	7pt, degree (+)	engage in outside interests	5pt, amount (-)

We divide these items in two blocks, based on the information they convey. Table 1 (a) lists the *subjective assessment* items, which are related to the degree of happiness and

[1] These items are used in the psychology literature.

Figure 1: Distribution of RQ in the sample.



Notes: This figure plots the distribution of RQ in the sample of individuals who become parents for (a) women and (b) men separately. The mean RQ in the full data is 0 and its standard deviation is 1.

conflict in the relationship. Table 1 (b) contains the *couple time use* items, which inform of the way in which the couple members use their time together.

To construct the main outcome, we first transform all the items such that lower values correspond to worse couple behaviors. With the responses to these questions in the full dataset, we carry out a factor analysis and use the first factor to construct a unified measure of relationship quality (RQ). All items have positive loadings and the factor explains 40.49% of the variation in the data.[2] The resulting variable is centered at zero and has a unit standard deviation. Higher values indicate a better relationship.

Figure 1 displays the distribution of RQ in the sample of individuals who become parents, separately for (a) women and (b) men. In both cases, the distribution is skewed towards the left, indicating a higher frequency of high-quality relationships. RQ is somewhat more dispersed and is lower on average for women than for men.

We follow the same factor analysis procedure to construct separate measures of RQ per item block in Table 1. We construct *subjective RQ* using the items in (a) and *time RQ* with the items in (b). Qualitatively, they summarize the separate pieces of information contained in the RQ measure. We plot the distribution of these measures in Figure A.5.

Validity of the measure. Subsection A.2 reports a number of tests to verify that RQ provides valuable information about the quality of a relationship. Following the life satisfaction literature, we first verify that the measure is informative. We do so by investigating how RQ performs in predicting individual behaviour. We find that marital

[2]The retained factor has eigenvalue 4.05, while the next one has eigenvalue 1. All the factor loadings are reported in Table A.1.

transitions and fertility decisions are precluded by significant RQ deviations, especially when it comes to couple dissolution. This evidences that RQ is an informative proxy of relationship quality. Second, we evaluate the interpersonal comparability of the measure. We study the correlation of responses across couple members. We find a high level of correlation across responses, concluding that there is a degree of commonality and, thus, objectivity in the measured concept.

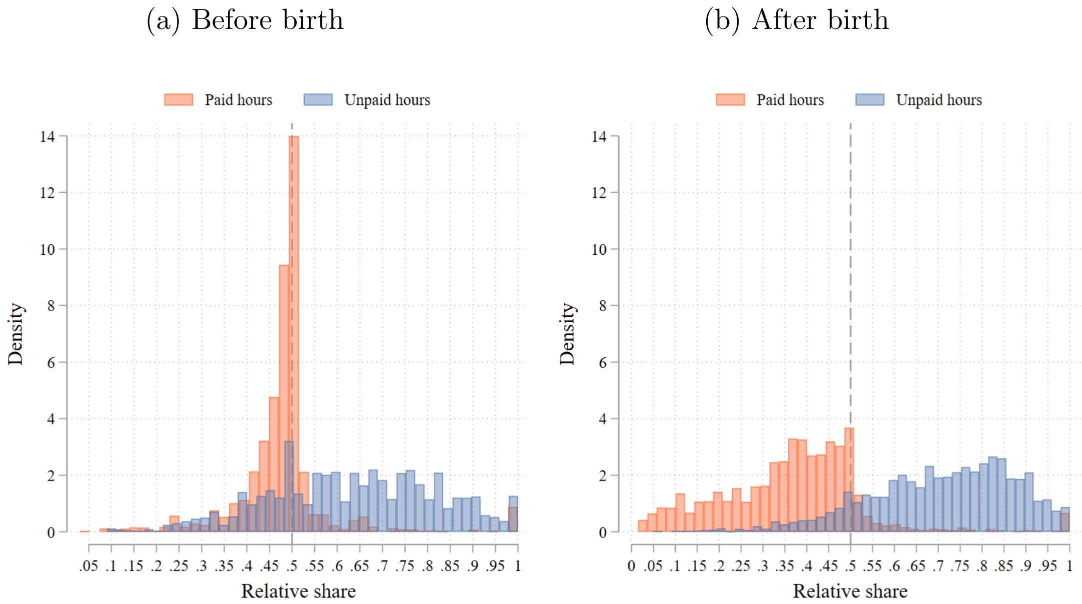
2.3 Household Specialization

We study household specialization making use of the time use variables provided by Understanding Society. The data provides two time-use variables at the individual level: hours spent on housework and number of hours normally worked per week.[3] These correspond to unpaid routinely housework and paid labor market work, respectively.

We are interested in the share of each type of work done by the couple members and how this evolves after child birth. We look into the share of the total housework hours and the total paid hours done by women. We refer to this as the *female share* of unpaid and paid hours. A 50% share of both types of work indicates no specialization, and different splits indicate specialization. Figure 2 plots the distribution of the female share of paid and unpaid hours (a) before and (b) after first child birth. There is large variation in household specialization before birth: the distribution of paid hours share is concentrated around 50%, whereas the distribution of unpaid work is uniformly distributed above the 50% threshold. Both distributions become largely polarized after the first child birth. The mass of paid hours is shifted under the 50% threshold and the share of unpaid hours becomes more concentrated above this mask.

[3]The specific questions are "About how many hours do you spend on housework in an average week, such as time spent cooking, cleaning and doing the laundry?" and "Thinking about your (main) job, how many hours, excluding overtime and meal breaks, are you expected to work in a normal week?", respectively. Note that, following Borra et al. (2021) we do not consider childcare to be part of routinely housework.

Figure 2: Distribution of the female share of paid and unpaid hours



Notes: These graphs plot the distribution of the share of the household total housework and labor market hours carried out by women (a) before first child birth and (b) after.

We classify couples according to the female share of paid and unpaid hours *before* first child birth. We distinguish four types of couples: (i) traditional couples, where women contribute mostly to housework and men to paid work; (ii) burdened women couples, where women take the largest share of work in both paid and unpaid labor, (iii) egalitarian couples, where the split of both types of work is equal for both couple members; and (iv) counter-traditional couples, where men take the largest share of housework. Table B.1 in Appendix B summarizes the characteristics of the different types of couples before birth. Traditional couples are formed by less educated partners, where men have full time jobs and women have part-time jobs plus around 12 hours of housework. In unbalanced couples men and women have full time jobs, but women spend 6 more hours weekly on housework. Egalitarian couples work on average more hours than the previous ones and times are the same for both couple members. Finally, in counter-traditional couples men spend more time on both the labor market and housework. They are the richest, on average.

2.4 Controls

Throughout the analysis we control for age and relationship tenure, period (wave), gender, education and area of residence (urban or rural). Table 2 summarizes these characteristics in the sample, as well as employment status, log monthly gross personal income and marital status.[4] Individuals are on average 32 years old. We observe slightly more women than men. They are mostly employed and living in urban areas. About 14% of them are

[4] We do not include this last set of variables as controls in our specifications since they are likely to change with first child birth.

college educated. All individuals are in cohabiting relationships and cohabitation spells are on average 8 years long. Finally, around half of the individuals in the sample are married.

Table 2: Summary statistics.

	(1) Mothers	(2) Fathers
Panel A: Individual characteristics		
Age	35.60	32.62
	(7.363)	(6.384)
College educated (%)	48.83	59.01
	(49.99)	(49.19)
Employed (%)	88.85	81.88
	(31.47)	(38.52)
Gross monthly income	2902.6	1915.7
	(1925.4)	(1397.3)
Observations	7087	7516
Panel B: Couple characteristics		
Tenure	7.592	
	(5.169)	
Married (%)	66.71	
	(47.13)	
In urban areas (%)	77.27	
	(41.91)	
Observations	14603	

Notes: This table presents mean values of the set of controls considered for the considered sample. All the values are reported at the individual level. Standard errors in parentheses. Do we want to show the summary by sex? The analysis is at the individual level and not separate for men and women.

2.5 Samples

The population of interest consists of individuals in cohabiting relationships, married or not, who become parents. We use a sample of individuals who become parents for the first time during the observation period. We only consider observations corresponding to the couple in which they have their first natural child. We restrict to individuals who were older than 18 and younger than 45 (women) or 50 (men), when they first became parents. We also exclude couples living with children from previous relationships.

The main sample is composed of individuals meeting those criteria and with available information about RQ, age, relationship tenure and sex. We further restrict to individuals reporting RQ at least once before and after having the first child. The resulting sample is an unbalanced panel formed by 1,600 individuals observed up to 5 times.

We use two different samples for the analysis on household specialization. The first sample is the one used when paid and unpaid work hours are used as the outcome. It is composed of individuals meeting the usual criteria and with available information about paid and unpaid work hours, age, relationship tenure and sex, where work hours are observed at least once before and after the birth of the first child. The resulting sample is an unbalanced panel formed by 3,255 individuals observed up to 23 times.

The second sample is the one used when RQ is used as the outcome. The sample consists of individuals from the main sample with available information about work hours for both couple members at least once before the birth of the first child. The resulting sample is an unbalanced panel formed by 940 individuals observed up to 5 times.

3 Empirical Strategy

3.1 Event Study design

We take an event study approach to study the causal impact of children on RQ. This methodology exploits sharp changes in outcomes of parents after the birth of their first child. The dynamic nature of this approach allows for treatment heterogeneity with time relative to the event.

We denote as G_i the year in which the first child was born to individual i. Thus, $t - G_i$ denotes time since i's first child was born (event-time). The sample consists of an unbalanced panel of *new-parents* in which we irregularly observe individual RQ at different stages of the fertility process. The available information allows us to look at 4 periods before first child birth and up to 7 periods after.We denote the RQ of individual i at time t by $y_{i,t}$ and we estimate the following regression:

$$y_{i,t} = \alpha_i + \mu_t + \sum_j \mathbb{1}\{j = t - G_i\}\delta_j + u_{i,t} \tag{1}$$

where we include the full set of dummies for time relative to first child birth ($\mathbb{1}\{j = t - G_i\}$), as well as individual and period fixed effects. Given that RQ is standardized, the coefficients are interpreted in terms of standard deviations. We cluster the standard errors at the couple level. We implement the estimator proposed by Callaway and Sant'Anna (2021) using only not yet treated individuals as controls. Through this method we avoid the negative weighting in the two-way fixed effects estimation in the case of treatment heterogeneity. This method is reviewed in Appendix C.

The causal interpretation of the obtained estimates relies on two assumptions. First, changes in RQ do not predict when individuals have their first child (*no anticipation*). This assumption would be violated if couples decided to have children in response to some periods of high RQ, for example. Second, in absence of treatment (birth of the first child) individuals' RQ would have evolved similarly regardless of the period when they had their first child (*parallel trends*). Although we cannot directly test whether these assumptions are satisfied, we provide some evidence about their plausibility.

First, we test whether the evolution of RQ is flat in the periods preceding the birth of the first child. However, flat pre-trends could be the result of averaging out couples who are not doing well and have a child to try to solve it and couples that decide to have a child because they are very happy. We show that this is not the case illustrating that the variation in RQ does not change with time relative to first child birth. Second, we show that the evolution of RQ with age and tenure is not lumpy. Third, we show that parents that experienced interruptions in fertility (used a fertility treatment or had an involuntary pregnancy interruption before their first child was born) do not differ from those who did not.

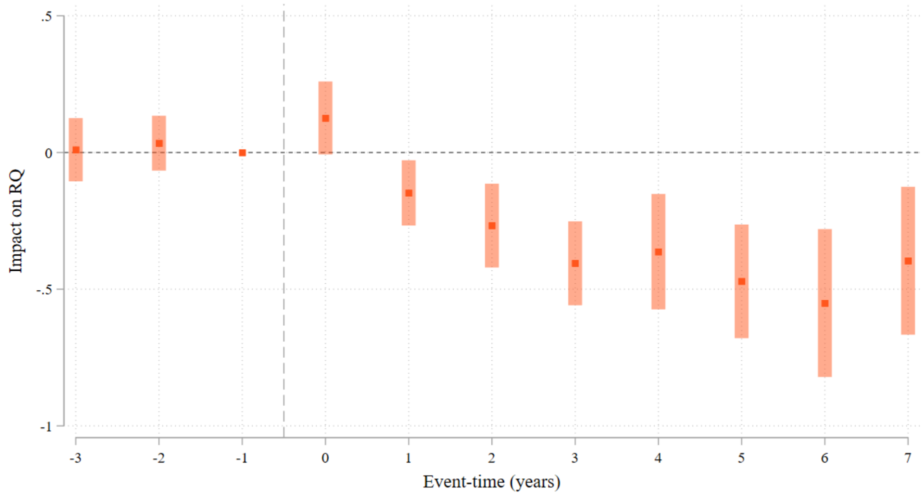
4 Impact of children on RQ

Figure 3 depicts the estimated effect of first child birth on RQ at each period relative to birth. The coefficients corresponding to periods before birth are not significantly different from zero. This confirms that the decision to have a child is not endogenous to the evolution of RQ after controlling for age and tenure. There is a significant decrease in RQ during the first three years after child birth. RQ stabilizes four years after having the child. The resulting value of RQ is on average half a standard deviation below the baseline. This is a remarkable finding. Having a child significantly shifts average RQ downwards, but it does not alter the trajectory of its evolution over the relationship.

Child birth is known to have different consequences on women and men (Kleven et al., 2019; Goldin, 2021). We check for gender differences in the effect of first child birth on RQ, interacting the full set of event-time dummies with gender in Equation 1. Figure 4 (a) plots the marginal effects of the years around birth on RQ by gender. There are no gender differences in baseline levels of RQ, the divergence starts the period after birth. Although the event impacts both parents' RQ significantly negatively, the impact on mothers is steeper and more sustained, stabilizing at a lower value than fathers'. Figure 4 (b) tests whether the impact on mothers is significantly different from the impact on fathers. We do so by fully interacting Equation 1 with the gender dummy. Mothers' point estimates are consistently below those of fathers, but they are never significantly different.[5]

[5] One potential concern could be that those couples with larger gender differences are the ones separating. We address this concern by repeating the analysis on the sub-sample of parents that never split. We find similar marginal effects by gender to the pooled sample.

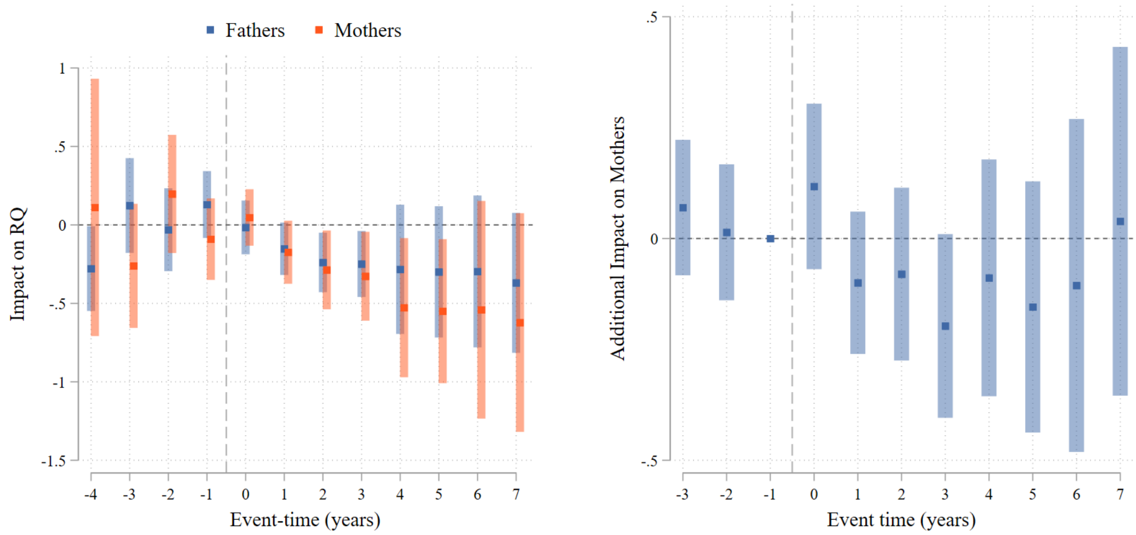
Figure 3: Dynamic effect of first child birth on RQ.



Notes: This graph plots the results of an event study of first child birth on RQ. The period prior to birth is taken as baseline. The plotted coefficients are the effects on RQ of leads and lags around the event. Confidence intervals are estimated at 95% level.

Figure 4: Effect of first child birth on RQ by gender

(a) Marginal effects by sex (b) Differences in impact



Notes: This graph plots the estimates of an event study of first child birth on RQ. The period prior to birth is taken as baseline. The plotted coefficients are the effects on RQ of leads and lags around the event. Confidence intervals are estimated at 95% level. (a) plots the marginal effects separately by gender, from estimating Equation 1 interacting the full set of event-time dummies with gender. (b) tests for significant gender differences plotting the interacted event-time dummies, from estimating Equation 1 fully interacted with gender.

4.1 Robustness

We test for the robustness of these results in a number of ways. First, we address any endogeneity issue that could arise from unobserved individual heterogeneity. We estimate the dynamic impact of fertility through Two-Way Fixed Effects, removing any unmeasured and time-invariant individual variation from our analysis. Our most parsimonious specification includes the full set of individual and time-fixed effects, exploiting within-individual, age, relationship tenure, and period deviations from parent-age-tenure-period means.

The TWFE estimated results in Figure D.1 are even more negative than the ones found using our main specification.[6] The decrease in RQ is more sustained with time relative to first child birth. One potential explanation relates to the OLS problem in dynamic TWFE specifications that has been raised recently by the literature of differences-in-differences (see Callaway and Sant'Anna (2021); Borusyak et al. (2022)). This literature acknowledges the threats to the identification of the ATTs parameters that derive from the types of comparisons made by OLS in settings with staggered treatment. We address this concern and perform the estimation of the Group-Average ATT estimator proposed by Callaway and Sant'Anna (2021). The results are reported in **??**. The impact is largely sustained although more imprecisely estimated.[7]

Next, we use alternative measures of relationship quality. In Figure D.3 we repeat the analysis constructing RQ separately for each item block in Table 1, in (a) for the subjective assessment items and in (b) for the couple time use ones. This would rule out that the impact is coming from a specific subset of items. We see that the impact is given in both blocks. The relevance of each item in Table 1 when constructing RQ might change after birth. We repeat the factor analysis using observations *after* child birth and reconstruct RQ. Figure D.4 indicates no differences in the estimated impact using this measure. Last, we use some similar measures from the psychology literature in Figure D.5 and observe similar effects. Furthermore, as we argue in Section 4, having a child significantly reduces quality time together. In such case, the time use items in Table 1 could be driving the results. We construct the measure again excluding these items and, thus, only using the subjective assessment items. As can be seen in Figure D.3, the results are slightly more modest in magnitude, but are still strongly significant and persistent.

Finally, we use different subsamples to address potential sample selection issues. First, in Figure D.6 we repeat this exercise using only couples that do not break up to remove any potential selection bias. The results do not change. Second, the results provided so far correspond to all parents, regardless of the total fertility of the couple. Figure D.7 repeats this analysis using subsamples of parents depending on their total lifetime fertility.

[6] We repeat the estimation of the TWFE specification using age and relationship tenure as time-varying variables. Results are displayed in panels (a) and (b) of **??**, respectively.

[7] The computation of this estimator is based on multiple aggregations of 2×2 differences-in-differences estimates. This requires observing individuals during the periods right before and after they receive the treatment. This restriction reduces significantly our main sample and explains the loss of precision.

The initial impact of the first-born is equally sharp for all parents, but the decrease in RQ is sustained for couples with more children. Figure D.8 repeats the analysis separately for individuals whose first child was a boy and a girl, finding no differences in the evolution of RQ. Finally, we repeat the analysis separating individuals according to the (a) age and (b) tenure when they had their first child in Figure C.5. We find that the dynamic effects of first child birth on RQ do not vary with age or tenure at birth.

5 Mechanism: Children as a time shock

Child birth is an unprecedented shock to time use: new tasks related to childcare arise and routinely housework greatly increases. This requires couples to adapt to the tightening of time constraints. First, leisure time of both men and women is reduced after first child birth (e.g., Aguiar and Hurst, 2007). Second, there is a need to redistribute time between paid (labor market) and unpaid (house) work. As shown, for example, by Kleven et al. (2019) men's labor outcomes do not react to child birth, whereas women greatly reduce their labor force participation and hours. Simultaneously, women take on most of the new housework tasks.

We argue that this structural change in the distribution of tasks and time between couple members may be mediating the impact of first child birth on RQ. We test this in two ways. First, we look at changes in the amount of quality time spent together as a couple. This mechanically impacts RQ, being part of the measure. We quantify the strength of this component in explaining changes in RQ. Second, we look into household specialization in paid and unpaid work. We classify couples depending on their baseline distribution of tasks and verify if there are differential effects of first child birth by couple type.

5.1 Increase in housework and time rearrangement

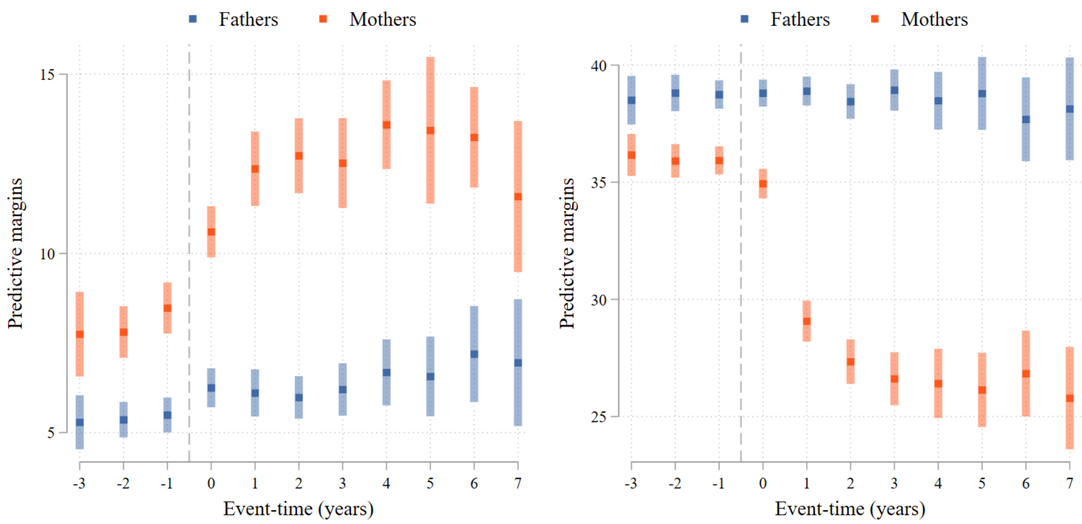
We start by documenting the increase in housework induced by first child birth. We perform an event study on weekly housework hours using the specification in Equation 1 and interacting the full set of event-time dummies with sex.[8] Figure 5 (a) plots the predictive margins from this estimation. Before child birth, women spend on average 2.5 weekly hours more than men in housework. After birth, mothers' housework hours slowly increase, more than doubling the baseline time by four years after birth. There is a small increase for men amounting to 1 additional weekly hour. This is evidence that the increase in housework induced by children is almost fully absorbed by women.

[8]Note that this measure corresponds to routinely housework, which does not include childcare (Borra et al., 2021).

[Figure 5](#) (b) plots the equivalent exercise for weekly hours worked in the labor market. We observe no change for men, but a strong decrease for women, who largely substitute full-time work for part-time work after first child birth. Thus, women decrease their paid work time to accommodate for the increasing demand for housework after having a child. This is evidence of household specialization induced by the presence of children.

Figure 5: Impact of first child birth on paid and unpaid hours, predictive margins

(a) Weekly housework hours (b) Weekly paid hours



Notes: This figure plots the impact of first child birth on weekly (a) housework and (b) paid work hours separately for men and women. We estimate Equation 1 using unpaid and paid hours as outcomes and interacting the full set of event-time dummies with sex. The period prior to birth is taken as baseline. The plotted coefficients are the effects on each item of leads and lags around the event. Confidence intervals are estimated at the 95% level. We plot the predictive margins derived from this estimation.

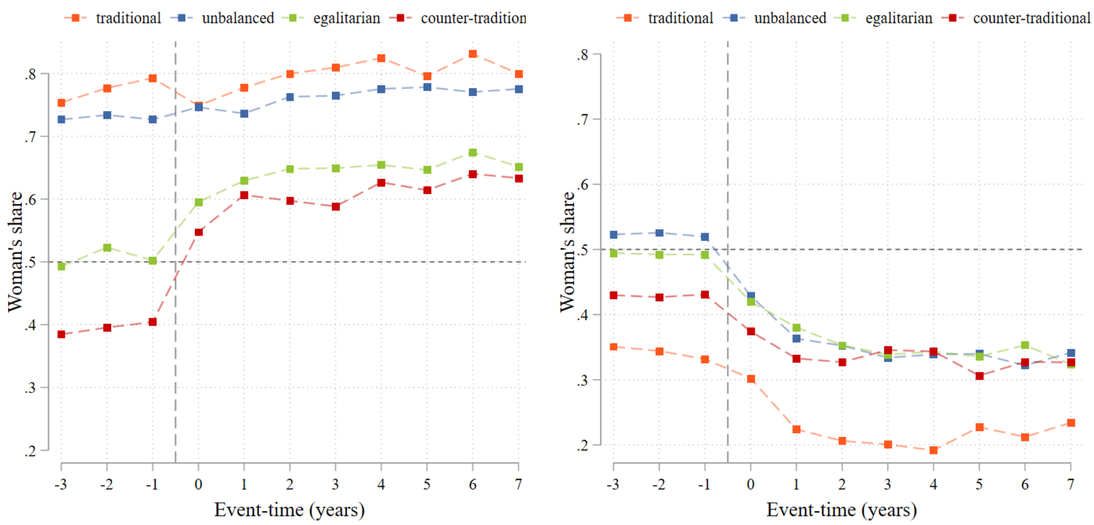
To quantify the relative importance, if any, of this time rearrangement, we separately study couples that experienced different shocks to household specialization. We divide couples in four groups depending on the female share of paid and unpaid work before first child birth, as explained in Subsection 2.3. We verify whether the female share evolves differently across couple types.

In Figure 6 we investigate the correlation between timing around first child birth and (a) the female share of housework time and (b) the female share of labor market time, for the different types of couples. Women who were doing a larger share of housework before having a child (traditional and unbalanced) increase their housework share less. The increase is larger for those who had a more egalitarian split (egalitarian and counter-traditional), but the share of this last group stays lower than the former.

In terms of labor market time share, Figure 6 (b) shows that traditional women were doing a smaller share on the baseline and they decrease it even further after birth. The other three groups have differently sized decreases, but they all converge to similar shares. This indicates convergence to a situation where fathers are in a full-time job and mothers in a part-time job. Overall, these graphs suggest the presence of variation in the magnitude of the relative time arrangements experienced by different types of couples.

Figure 6: Impact of first child birth on female time shares

(a) Housework share (b) Paid work share



Notes: These graphs plot the correlation between first child birth and the female shares of (a) housework and (b) paid work time by couple type. We plot the average share for each type of couple at each time around first child birth.

Next, we check whether larger changes in time shares mediate the impact of first child birth on RQ. The first row in Table 3 displays the average RQ per couple type before first child birth. All couple types' average is above 0, given that they are all part of the subsample of individuals who will eventually have a child. There is some heterogeneity in the baseline RQ. Individuals in more egalitarian couples report higher levels of RQ on the baseline, whereas traditional couples report the lowest values.

Due to data limitations, we cannot study the dynamics of the impact by groups, but we can estimate the static impact through a difference-in-differences design. We define $D_{i,t}$ to be a dummy equal to one if individual i already had a child at time t and C_i^j, $j \in \{1,2,3,4\}$, to be a set of dummies equal to one for each couple type. We estimate the following model:

$$\begin{aligned} y_{i,t} = & \sum_j C_i^j \rho_j + \sum_j D_{i,t} C_i^j \delta_j + \sum_a \mathbb{1}\{a = \text{age}_{i,t}\}\alpha_a + \sum_d \mathbb{1}\{d = \text{tenure}_{i,t}\}\gamma_d + \\ & + \sum_w \mathbb{1}\{\text{period}_w = t\}\psi_w + \mathbf{X}_{i,t}\boldsymbol{\beta} + v_{i,t} \end{aligned} \tag{2}$$

The parameters $_j$'s account for any existing differences in the level of RQ across the different types of couples. The parameters δ_j's capture the impact of first child birth for each couple.

The second row of Table 3 reports the marginal effects by couple type from estimating Equation 2.[9] All the coefficients are negative, meaning that all couple types are negatively impacted by first child birth. However, this impact is only significantly different from zero for individuals in unbalanced and egalitarian couples. This is the product both of a precision loss when dividing the sample and of heterogeneity in the impact. The smallest coefficient corresponds to individuals in traditional couples. In fact, these are the ones experiencing smallest changes in the share of housework and labor market hours. The largest impact is that of unbalanced couples. These individuals do not experience large changes in housework shares, since women were already doing most before birth. However, they are the ones suffering the largest decrease in paid work hours.

Table 3: Regression analysis by couple type

	Traditional	Unbalanced	Egalitarian	Counter-traditional
Baseline RQ	0.300	0.428	0.513	0.489
	(1.018)	(0.788)	(0.635)	(0.777)
Marginal Effects	-0.138	-0.204**	-0.184*	-0.169
	(0.187)	(0.076)	(0.074)	(0.090)
Observations	1363	3456	2098	1668

Notes: This table reports the baseline RQ (first row) and the marginal effects from estimating Equation 2 by couple type (second row). Standard errors clustered at the couple level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.2 Impact of time rearrangements on RQ: evidence from an expansion of state-funded childcare

Authors' note: this section contains preliminary results. Waiting for data access to run a different type of analysis.

We provide causal evidence of the impact of time rearrangements on RQ exploiting an expansion of state-funded childcare for three- and four-year-olds from 15 to 30 hours a week. The U.K. government introduced this reform in September 2017 in England, with the aim of helping with the provision of childcare to working parents. However, the effective implementation of this policy was not immediate, and it varied across educational administrative regions (Local Educational Authorities) depending on their response capacity to the new local demand for childcare.[10] We deliver preliminary results that exploit

[9] Note that the estimation of the static regression instead of the dynamic regression used in the main analysis requires an extra assumption: homogeneity of treatment effects with time relative to treatment. The main results suggest that this assumption is not satisfied. Thus, the estimates are a weighted average of the treatment effects at different points in time. The weights given to the treatment effects for each relative time are increasing in the number of observations at that event-time (Goodman-Bacon, 2021). In this case, relative times closer to first child birth have higher weights.

[10] The provision of early childcare is administered by Local Authorities (LAs) or Local Educational

Table 4: Regression analysis by couple type and gender

	RQ		Subjective RQ	
	Women	Men	Women	Men
Traditional	-0.356*	-0.188	-0.158	-0.0708
	(0.164)	(0.166)	(0.135)	(0.148)
Unbalanced	-0.316***	-0.215*	-0.228**	-0.124
	(0.088)	(0.088)	(0.074)	(0.076)
Egalitarian	-0.304***	-0.101	-0.203**	-0.0564
	(0.078)	(0.086)	(0.067)	(0.079)
Counter-traditional	-0.232*	-0.161	-0.186*	-0.0978
	(0.093)	(0.095)	(0.089)	(0.093)
Observations	2160	2090	2686	2606

Notes: This table reports the marginal effects from estimating Equation 2 by couple type and gender. Standard errors clustered at the couple level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

this geographic variation, comparing time changes in RQ between parents of three- and four-years olds living in LAs with different policy adoption speeds. [11]

Institutional context: Since 1999, all parents of three- and four-year-olds could claim 15 hours of free childcare a week for 38 weeks of the year in England, and similar schemes were applicable in the rest of the country.[12] The 15-hour entitlement is universal, and a child becomes eligible on the day of her third birthday. After this day, parents should apply for a code via the Department of Education. Once this code is validated, they can make use of the 15-hour entitlement from the beginning of the next starting term *conditional on the availability of school providers* in their educational area.[13] [14]

In April 2016, the U.K. government announced the extension of 15 additional hours of free childcare (so-called 30-hour entitlement) for households in which the two parents worked the equivalent to at least sixteen weekly hours at the minimum pay, and earn less than 100.000 pounds a year.[15] The application procedure is the same as for the 15-hour

Autorithies (LEAs) in U.K. LEAs are the local councils responsible for education within their jurisdictions. Although its terminology is no longer mentioned in legal documents since 2010, they are still used to distinguish local authorities with education functions from those without them.

[11] There are 153 LA, and in some cases, they cover several local authority districts (LAD). We have individual geographic information at the LAD level, and hence, use information from the U.K. Postcode office in 2016 to construct a cross-walk between LADs and LAs.

[12] This law is called the *Childcare Act 1991* add reference of the law

[13] The application process can be made online through the Department of Education's website in approximately twenty minutes. Parents need to provide some personal detail, and the National Insurance Number or the Unique Taxpayer Reference if they are self-employed.

[14] The effective implementation of the policy varied across different regions, depending on the number of school providers and the capacity that each region had to adapt to the new demand for childcare.

[15] See *Childcare Act 2016*

entitlement: once a child turns three years old, parents can claim the entitlement, and it becomes effective in the next starting school term once it is validated by the Department of Education.

To prevent the underprovision of childcare supply, the government tested a pilot program in eight local authorities in September 2016.[16] In September 2017, the policy becomes effective in all of England and Wales. However, despite the state's provision efforts, the 30-hour entitlement was not delivered uniformly across all regions.

Measures and geographic variation in the roll-out of the policy: We use two measures to assess the effectiveness of the policy. The first measure is the share of three- and four-year-olds registered in 2018 under the 30-hours entitlement in each local authority, which with we proxy the take-up rate or speed of policy adoption in each region. Unfortunately, information on the number of eligible children (with parents meeting the working and income requirements) is not available at the local authority level. Instead, we use Census data on the number of three- and four-year-olds living in each LA in 2018, defining the eligible population based only on child-age requirements. We address later the potential issues that this measure may introduce into the analysis. We rank LAs according to this measure and classify them into slow, medium, and fast adopters of the policy.[17]

We use the ratio of children per childcare provider by local authority as a proxy of the capacity of each local authority to accommodate to the potential rise in childcare demand.[18] As before, we use this measure to classify LAs according to their provision capacity into crowded, medium, and available.

Figure 7 shows geographic variation in the implementation of the policy in the year 2018 using our take-up (Panel (a)) and capacity (Panel (b)) measures. These two measures are positively correlated, meaning that the implementation of the policy was more effective in those regions with a less crowded childcare system prior to its introduction. We exploit this geographic variation to leverage quasi-random variation in the availability of (free) childcare across households living in different LAs.

[16]The eight local authorities were: Wigan, Staffordshire, Swindon, Portsmouth, Northumberland, York, Newham and Hertfordshire. Four additional local authorities – Dorset, Leicestershire, North Yorkshire and Tower Hamlets – were added to the pilot program in April 2017.

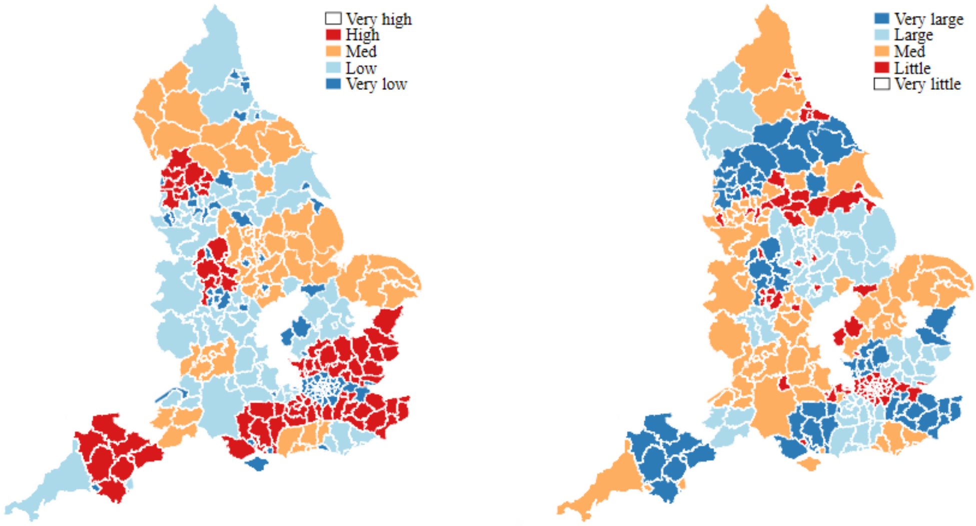
[17]There is an average of 0.18, 1.41, and 2.55 registered children per eligible population in slow, medium, and fast LAs, respectively. Figures in which the number of registered children outnumbers the population may due to errors in the population estimates provided by the Office for National Statistics at such granular levels as LAs. Moreover, these estimates do not account for short-term migrants or persons who change their residence in less than a year.

[18]We construct this measure using Census data on the number of children aged between 3-4 years old living at each LEA in 2017 as the numerator. We use data provided by the Department of Education on the total number of state-funded, Private, Independent, and Voluntary childcare providers existing in each local authority in 2018.

Figure 7: Geographic variation in the expansion of childcare, 2018

(a) Eligible children per centre

(b) Registered children per eligible population



Notes:

Empirical strategy The implementation of this policy leads to several empirical strategies to estimate the causal impact on the outcome of interest. We first detail the main strategy followed so far, and then briefly comment on the methodology we plan to apply upon receiving an additional data set. Importantly, given the nature of the policy, we can only aim to estimate an intention-to-treatment (IIT) type of causal effect.

Our first approach is similar to Blanden et al. (2014); Brewer et al. (2014), and uses the regional variation in the roll-out of the policy. However, differently from them, we do not use a continuous measure of treatment intensity, but rather our categorical measure of fast and slow adopters. Equation 3 specifies the main regression:

$$y_{i,t,l} = \beta_0 + \beta_1 C_{i,l} + \beta_2 Post_{i,t} + \beta_3 C_{i,l} \times Post_{i,t} + \Gamma X_{i,t} + \delta Z_{l,t} + \psi_w + \epsilon_{i,t,l} \qquad (3)$$

The term $y_{i,t,l}$ refers to the outcome of interest of individual i, in period t at local authority l. $C_{i,l}$ is a categorical variable classifying LAs into slow, medium, and fast adopters of the policy. The baseline category is slow adopters, and hence, the parameter β_1 measures any time-invariant differences between medium and fast LAs relative to slow LA adopters. $Post_{i,t}$ is an indicator variable that takes value one whenever an observation is after the implementation of the policy in September 2017.

We are interested in the parameter β_3, which captures the differential effect of the introduction of the policy across LAs with different speeds of adoption. In other words, it measures the impact of the policy by comparing *less* and *more* treated regions. The key assumption for the causal interpretation of this parameter is that differences in the speed of adoption across LAs are orthogonal to any other unboserved difference among them. To prevent confounding effects with the speed adoption of the policy, we introduce a measure of local male employment at the LA level at each period, $Z_{l,t}$. $X_{i,t}$ includes the usual set of individual and couple characteristics. We finally include period fixed effects to account for any general time trend. We estimate this equation by OLS on the subsample of parents with 3-4 years old aged children during the periods 2016-2019. We use robust standard errors and cluster them at the individual level.

A final issue concerning the causal effect of the policy is the presence of no-anticipation effects. There was more than one year between the announcement of the policy (April 2016) and its introduction. Given the working eligibility requirements, parents and especially mothers of three- and four-years olds in 2017 may have started to look for a job before the introduction of the policy. This could bias downwards the estimated effect of the policy on mothers' labor supply. Nevertheless, our setup compares time changes across LAs according to their speed of policy adoption. Our main identifying threat would be that anticipatory behaviours were correlated with the roll-out of the policy at the local level. This could happen if, for instance, information or parental incentives to apply for the 30-hours entitlement differ across LAs in a systematic way. To check whether this was the case, we use data from the Department of Education to check that the share of claims issued by parents of three- and four-years olds is similar across all LAs.

Authors' note: Upon access to sensitive data on children's month of birth, we plan to carry a Regression Discontinuity Design using the natural discontinuities in the starting date of entitlement to childcare that arise from the child's month of birth. The paragraphs below present a preliminary empirical strategy that can be carried out using the data we have access to now. In the final version of the paper, this strategy is likely to be a robustness exercise.

Impact of the reform on time use and RQ: We first analyze the impact of the policy (intention-to-treatment effect) on the measures of RQ and Subjective RQ, as well as individual well-being. Table 5 presents the corresponding results. Columns (1)-(4) use the measures of RQ and Subjective RQ, for the unconditional regressions (without controls) and the full specification. Colum (5) uses as outcome variable a measure of individual well-being, and estimates the main equation using only the subsample of mothers.

We can see that the policy had a negative effect on the RQ of parents living in fast and medium-adopters LAs relative to those living in slow-adopter regions. This result is even bigger in magnitude for the outcome of Subjective RQ. Importantly, there are no significant baseline differences in the values across differently classified regions. The well-being of mothers is also significantly worse off by the expansion of childcare, being half of a standard deviation lower for those one living in fast-adopter regions.

Table 5: Estimates of the expansion of childcare on RQ and well-being measures

	RQ		Subjective RQ		Well-being
	(1)	(2)	(3)	(4)	(5)
Medium	0.274*	0.172	0.174	0.103	0.114
	(0.157)	(0.156)	(0.156)	(0.158)	(0.180)
Fast	-0.027	0.097	-0.043	0.061	-0.042
	(0.195)	(0.187)	(0.190)	(0.181)	(0.205)
Post	0.410**	0.178	0.283	0.108	0.338
	(0.199)	(0.250)	(0.219)	(0.261)	(0.220)
Medium $\times Post$	-0.635	-0.605	-0.217	-0.262	-0.163
	(0.401)	(0.407)	(0.360)	(0.395)	(0.301)
Fast $\times Post$	-0.980*	-1.085*	-1.047*	-1.170*	-0.541*
	(0.573)	(0.626)	(0.598)	(0.651)	(0.317)
Controls	-	Yes	-	Yes	Yes
Observations	227	210	229	212	344

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports the marginal effects from estimating Equation 2 by couple type and gender. Standard errors clustered at the couple level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We now try to analyze the above results through the lens of our main mechanism. In the section above, we provide evidence that the first child birth leads couples to traditional types of time arrangements, no matter the level of household specialization prior to the child. We also show that the first child birth had a large negative impact on the RQ couples in which women experience large changes in time use. The policy that we analyze aims to alleviate this since it entitles eligible parents to 30-hours of free childcare. Indeed, the policy explicitly incentives mothers' labor supply, since it requires both parents to be working the equivalent of at least 16 hours a week at the minimum pay.

We provide causal evidence of that result in columns (2)-(3) of Table 6. Mothers in fast-adopter regions increased their unconditional labor supply by 9 hours a week, although the estimate loose significance and magnitude once we include all the set of controls. Similarly, columns (4)-(5) provide suggestive evidence that the impact of the policy decreased the number of weekly hours that mothers spent on housework, although the magnitude of such reduction is wide smaller than the corresponding one in paid hours.

Finally, we show in Column (1) that the introduction of the policy increases the use of childcare in fast and medium-adopter regions relative to slow-adopter ones (38. and 22 p.p., respectively).

Although our results may suggest that mothers of fast-adopter regions become overloaded, we are still cautious when interpreting these results. Once we have access to children's month of birth, we will contrast the estimates of this analysis with the ones provided by a Regression Discontinuity Desing.

Table 6: Estimates of the expansion of childcare on take-up rates and mothers' time use

	Childcare use	Paid hours		Housework hours	
	(1)	(2)	(3)	(4)	(5)
Medium	0.055	1.114	-0.888	-3.997**	-3.763**
	(0.076)	(3.458)	(3.170)	(1.695)	(1.752)
Fast	0.034	-1.754	-2.101	0.395	0.313
	(0.096)	(3.336)	(3.053)	(2.496)	(2.454)
Post	-0.146	-2.632	3.009	1.085	2.238
	(0.102)	(2.397)	(3.329)	(1.920)	(4.012)
Medium $\times Post$	0.220*	1.502	2.979	-2.123	-1.547
	(0.113)	(4.875)	(4.791)	(2.328)	(2.413)
Fast $\times Post$	0.389***	9.602**	7.258	-1.434	-0.957
	(0.115)	(4.684)	(4.501)	(3.916)	(3.745)
Controls	Yes	-	Yes	-	Yes
Observations	345	338	330	201	195

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports the marginal effects from estimating Equation 2 by couple type and gender. Standard errors clustered at the couple level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.3 The role of income

Economic resources can help relieve household specialization by means of externalizing housework. To study this, we divide couples into quartiles according to gross household income before first child birth. **??** in Appendix B summarizes the average values of the usual set of characteristics before child birth across income quartiles. The main difference between the highest and the lowest income couples is the level of education: 60% of individuals in the top quartile are college educated, whereas only 5-10% of those in the lowest quartile are.

To assess the different impact of first child birth on couples depending on their income quartile, we repeat the analysis in Equation 2 using the quartiles as groups and display the results in Table 7. The first row contains the baseline differences in RQ across income quartiles. Couples above the median income level report much higher RQ before the birth of their first child. The highest quartile is on average one standard deviation of RQ above the bottom quartile.

The second row of Table 7 reports the estimated marginal effects by income group. The impact of first child birth on RQ is larger for individuals in richer households. Nevertheless, the resulting level of RQ is still above that of poorer households after first child birth.

We need to check how these types of couples differ in the other mechanisms, too.

Table 7: Impact of first child birth on RQ measures, by income quartile

	(1) bottom	(2) second	(3) third	(4) top
Baseline RQ	-0.034	0.176	0.421	0.471
	(1.113)	(1.104)	(0.761)	(0.746)
Marginal Effects	-0.130	-0.116	-0.256***	-0.235***
	(0.180)	(0.082)	(0.049)	(0.041)
Observations	433	1836	3438	4502

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6 Concluding Remarks

To be completed.

References

Aguiar, M. and Hurst, E. (2007). Measuring trends in leisure: The allocation of time over five decades. *The Quarterly Journal of Economics*, 122(3):969–1006.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The quarterly journal of economics*, 115(3):715–753.

Alon, T., Doepke, M., Olmstead-Rumsey, J., and Tertilt, M. (2020). The impact of covid-19 on gender equality. Technical report, National Bureau of economic research.

Blanden, J., Del Bono, E., Hansen, K., Mcnally, S., and Rabe, B. (2014). Evaluating a demand-side approach to expanding free preschool education. *Unpublished Manuscript, University of Surrey.*

Borra, C., Browning, M., and Sevilla, A. (2021). Marriage and housework. *Oxford Economic Papers*, 73(2):479–508.

Borusyak, K., Jaravel, X., and Spiess, J. (2022). Revisiting event study designs: Robust and efficient estimation. *Available at SSRN 2826228.*

Brewer, M., Cattan, S., Crawford, C., and Rabe, B. (2014). The impact of free, universal pre-school education on maternal labour supply. *Institute for Fiscal Studies, London: IFS.*

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Carlson, M. J. and VanOrman, A. G. (2017). Trajectories of relationship supportiveness after childbirth: Does marriage matter? *Social Science Research*, 66:102–117.

Chiappori, P.-A. and Mazzocco, M. (2017). Static and intertemporal household decisions. *Journal of Economic Literature*, 55(3):985–1045.

Cortés, P. and Pan, J. (2020). Children and the remaining gender gaps in the labor market. Technical report, National Bureau of Economic Research.

Goldin, C. (2021). *Career and Family: Women's Century-Long Journey toward Equity.* Princeton University Press.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Greenwood, J., Guner, N., and Vandenbroucke, G. (2017). Family economics writ large. *Journal of Economic Literature*, 55(4):1346–1434.

Hupkau, C. and Petrongolo, B. (2020). Work, care and gender during the covid-19 crisis. *Fiscal studies*, 41(3):623–651.

Joel, S., Eastwick, P. W., Allison, C. J., Arriaga, X. B., Baker, Z. G., Bar-Kalifa, E., Bergeron, S., Birnbaum, G. E., Brock, R. L., Brumbaugh, C. C., et al. (2020). Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies. *Proceedings of the National Academy of Sciences*, 117(32):19061–19071.

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

Korsgren, P. and van Lent, M. (2020). Earmarked paternity leave and well-being. *Available at SSRN 3753475*.

Sevilla, A. and Smith, S. (2020). Baby steps: The gender division of childcare during the covid-19 pandemic. *Oxford Review of Economic Policy*, 36(Supplement_1):S169–S186.

University of Essex, Institute for Social and Economic Research (2022). Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009. Technical Report 15th Edition, UK Data Service. http://doi.org/10.5255/UKDA-SN-6614-16.

A RQ measure

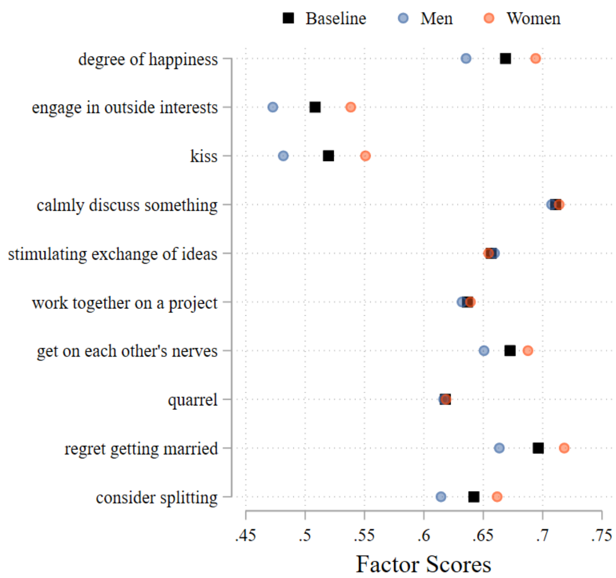
A.1 Factor Analysis

Table A.1: Factor loadings of RQ

(a) Subjective assessment		(b) Couple time use	
How often do you... ?		*How often do you... ?*	
consider splitting	0.642	work together on a project	0.636
regret getting married	0.697	stimulating exchange of ideas	0.657
quarrel	0.618	calmly discuss something	0.711
get on each others nerves	0.672	kiss partner	0.520
What is the... ?		*Do you and your partner... ?*	
degree of happiness w/ relationship	0.508	engage in outside interests	0.669

Notes: This table reports the factor loadings of the factor analysis on the 10 items in the Understanding Society Partner module. The first factor, which we call RQ, is the measure of relationship quality used in the analysis. It has eigenvalue 4.05 and explains 40.49% of the variation in the data. The left panel shows the subjective assessment items and the right panel displays the couple time use items.

Figure A.1: Factor loadings by sex



Notes: This graph displays the factor loadings from computing the factor analysis on the entire sample, only on women and only on men. explanation power in each case

A.2 Validity

To confirm the validity of RQ we follow the life satisfaction literature and verify that the measure fulfils two criteria: informativeness and interpersonal comparability.

Informativeness. First, we verify that the information provided by RQ is meaningful. We do so by assessing the predictive capacity of this measure for couple decisions: (a) marriage and separation and (b) fertility decisions. Marriage increments separation costs, acting as a commitment mechanism. We hypothesize that couples transitioning into marriage should report higher than average RQ. Separation, instead, is the result of bad quality relationships. Thus, we should observe lower than average RQ on those couples about to dissolve. Finally, we hypothesise that couples deciding to have a child have a higher than average RQ.

To assess the predictive power of RQ on these decisions, we first partial it out of the controls listed in Subsection 2.4 and obtain the residuals. Figure A.2 plots the empirical cumulative distribution function (cdf) of these residuals for different samples. Panel (a) compares the overall distribution of the RQ residuals in the full data with the residuals one period before marriage and one period before dissolution. As expected, the distribution before marriage is shifted to the right, indicating that before getting married individuals report higher RQ at any point of the distribution. In contrast, the distribution before dissolution is largely shifted to the left. Individuals report lower RQ before dissolution at any point of the distribution. Interestingly, the pre-dissolution distribution is much further from the full sample distribution than the pre-marriage is. This indicates that negative RQ deviations have a larger impact on marital decisions than positive deviations do.

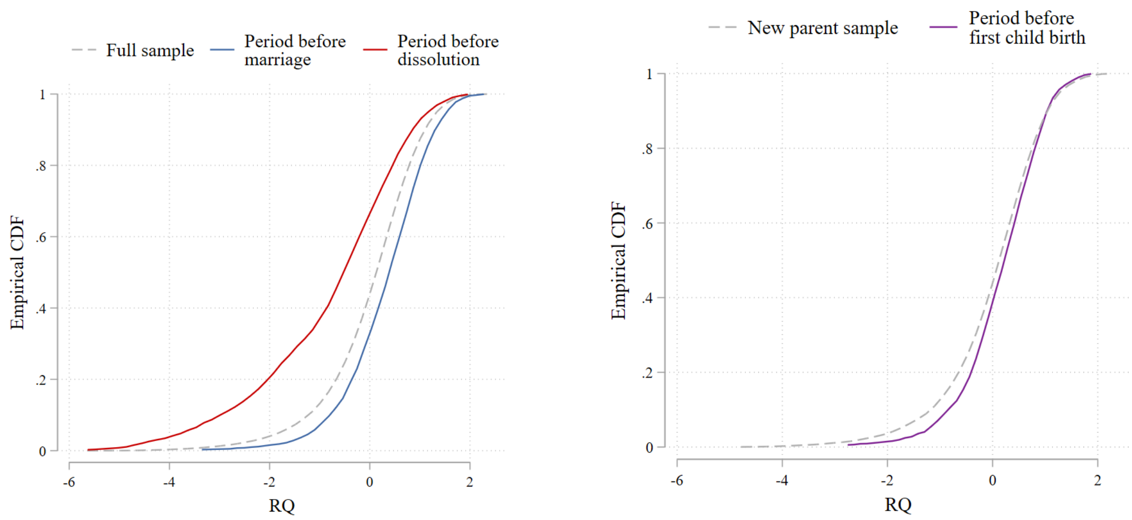
Figure A.2 (b) compares the distribution of the RQ residuals between the new parent sample and the observations one period before having the first child, that is, at the time of conception. This distribution is slightly shifted to the right in comparison to the benchmark. However, the empirical distribution of this sample does not seem to be significantly different from the benchmark.

We formally test the differences between these empirical distributions through a two-sample Kolmogorov-Smirnov equality-of-distributions test. This tests whether two samples are derived from the same population and, thus, follow the same distribution. Table A.2 displays the D-statistics and p-values obtained from this test for the samples considered. We find that the pre-divorce and pre-marital samples contain respectively significantly smaller and significantly larger values than the full sample. Additionally, the pre-child sample contains significantly larger values than the new parent samples, indicating that the differences observed in Figure A.2 (b) are sufficiently large. The combined test indicates that all three samples come from different distributions in comparison to

Figure A.2: Informativeness of RQ: behavior prediction.

(a) Marital transitions

(b) First child birth



Notes: This figure displays the empirical cdf of the residual obtained from regressing RQ on the set of controls listed in Subsection 2.4. Panel (a) presents the residual for the full data, observations one period before marriage (1,150 instances) and observations one period before dissolution (923 instances). Panel (b) displays the residual for the new parent sample and observations one period before the birth of the first child (821 instances).

the benchmarks.

Table A.2: Two-sample Kolmogorov-Smirnov test.

	d_0 = Full sample		d_0 = New parent sample
	d_1 = Before marriage	d_1 = Before divorce	d_1 = Before first child
$d_0 > d_1$	0.000	0.1257	0.0841
	(1.000)	(0.000)	(0.000)
$d_0 < d_1$	-0.2752	-0.0003	-0.0174
	(0.000)	(1.000)	(0.696)
Combined	0.2752	0.1257	0.0841
	(0.000)	(0.000)	(0.000)

Notes: This table displays the results of two-sample Kolmogorov-Smirnov tests on different samples. The reported coefficients are the resulting D-statistics and p-values (in parentheses).

The periods precluding marital transitions and fertility decisions are characterized by significant deviations from the average RQ. We conclude that RQ provides valuable information about couple behaviour, which is largely dictated by the quality of the relationship. This argues in favour of the validity of this measure.

Interpersonal comparability. Second, there should be some degree of commonality in the concept that RQ contain shared across individuals. We test this by assessing the level of correlation of RQ across the members of a couple. Table A.3 displays the descriptive results from regressing women's RQ on their (male) partners' RQ and the usual set of controls. Man's RQ is a highly significant predictor of woman's RQ. In fact, it is the largest in magnitude, almost quadrupling the second largest: being married. The coefficient indicates that a unit increase in man RQ is correlated with an increase in woman RQ of around 0.6.[19]

Table A.3: Regression of women's RQ on man's RQ.

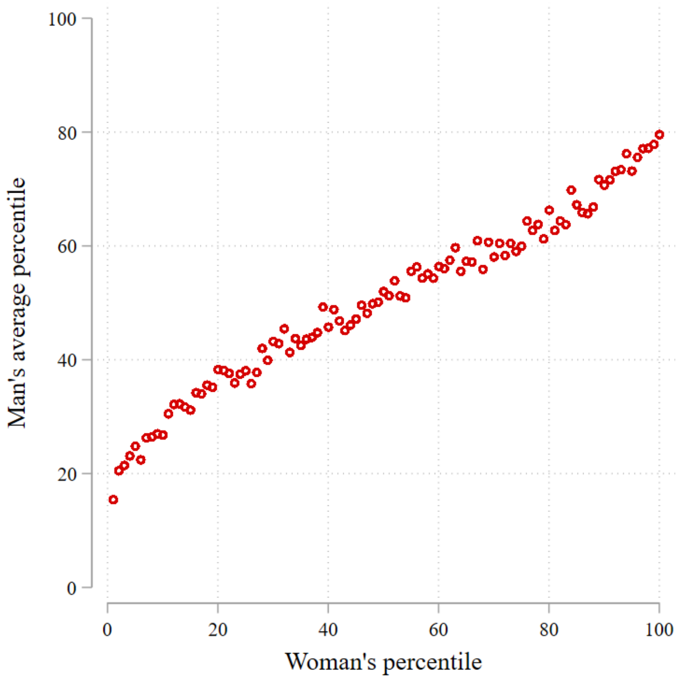
	Woman's RQ
Man's RQ	0.587***
	(0.009)
College Degree	0.049**
	(0.016)
Employed	0.048*
	(0.019)
Log Personal Income	-0.002
	(0.006)
Urban	-0.032
	(0.016)
Married	0.155***
	(0.024)
At least one child	-0.111***
	(0.018)
Constant	0.134
	(0.075)
Age	✓
Tenure	✓
Wave	✓
Observations	25884
R^2	0.3243

Notes: This table displays the descriptive results from regressing women's RQ on their (male) partners' RQ and the usual set of controls. Standard errors clustered at the couple level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We look at the non-linear relation between the RQ of both couple members through a rank-rank plot. We first partial it out of the regressors in Table A.3. Figure A.3 displays the average RQ residual percentile rank of the man per woman's percentile rank. Although there is no perfect correlation between the two, there is a clear positive relation. Perfect correlation would result in a 45 degree line. The slope is steepest for the top and bottom percentiles, being of around one point. It flattens out at the center of the distribution by almost half. This indicates that extreme assessments of the quality of the relationship are shared much more intensely than intermediate ones.

[19]Note that the standard deviation of RQ is one, so we can interpret this coefficient in RQ units.

Figure A.3: Rank-rank correlation of RQ residual across couple members.



Notes: This figure plots the average husband RQ residual percentile rank per wife RQ residual percentile rank.

A.3 Evolution of RQ

We study the evolution of RQ over time by looking at the evolution of this measure with age and relation tenure. We estimate the following regression:

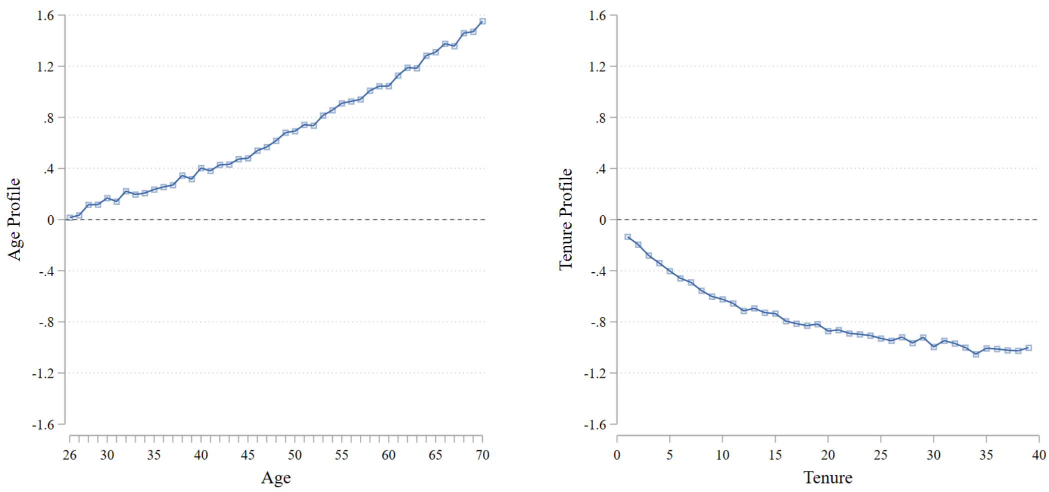
$$y_{it} = \sum_{a} \mathbb{1}\{a = \text{age}_{it}\}\alpha_a + \sum_{d} \mathbb{1}\{d = \text{tenure}_{it}\}\gamma_d + \sum_{w} \mathbb{1}\{w = \text{wave}_t\}\psi_t + \mathbf{X}_{it}\boldsymbol{\beta} + u_{it}$$

where y_{it} denotes RQ of individual i at time t, we include full sets of age, tenure and wave dummies and $\mathbf{X}_{it}$ includes the rest of the controls. We use a fixed effects approach to eliminate unobservable individual heterogeneity, which contain cohort effects. Doing so, we abstract from this type of variation and preserve only the variation that can be attributed to an additional year of age or tenure. Since we include both variables non-parametrically, the estimated coefficients provide the age and tenure profiles of RQ.

Figure A.4 (a) plots the age profile of RQ, in comparison to the baseline of 25 years. Whilst this is clearly observational, aging has a positive effect on RQ. Additional years of age induce increasingly larger levels of RQ. These increments are highly smooth and almost linear. Figure A.4 (b) does the same for tenure, taking one-year-old relationships as a baseline. RQ steeply decreases with tenure during the first ten to fifteen years. It stabilizes for sufficiently long relationships. As with age, additional years of tenure reduce RQ smoothly, without significant jumps.

Figure A.4: Age and tenure effects on RQ.

(a) Age effects (b) Tenure effects

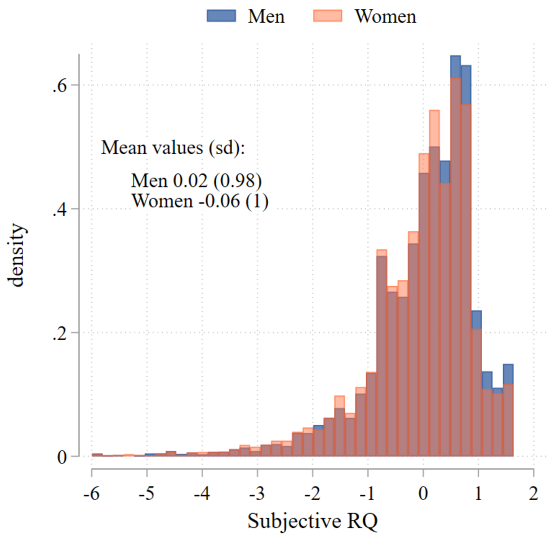


Notes: This figure plots the age and relationship tenure profiles of RQ. These are obtained estimating a non-parametric regression of age and tenure on RQ through fixed effects. Panel (a) takes 25 as the baseline age and panel (b) takes 1 as the baseline tenure.

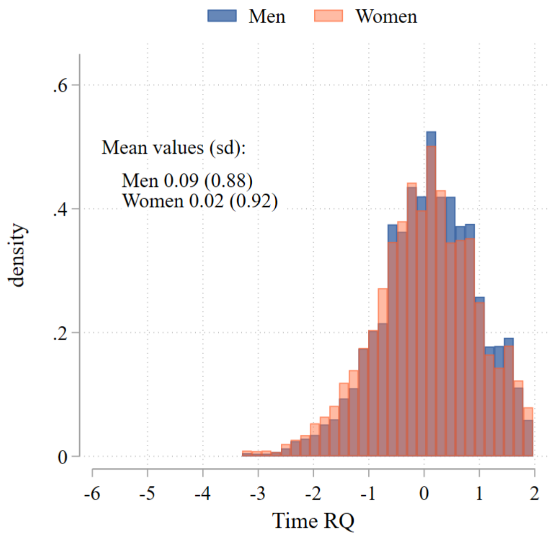
A.4 Other measures

Figure A.5: Distribution of RQ by item block

(a) Subjective RQ



(b) Time RQ



Notes: This figure plots the distribution of (a) subjective RQ and (b) time RQ in the sample by gender. The mean of both measures in the full data is 0 and their standard deviation is 1.

B Summary Statistics

Table B.1: Summary statistics for differently specialized couples

	Traditional		Unbalanced		Egalitarian		Counter-traditional	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women	Men	Women	Men	Women	Men	Women	Men
Age	28.57	30.84	28.90	31.34	29.03	31.07	28.45	31.30
	(5.213)	(5.729)	(4.562)	(5.677)	(4.075)	(5.360)	(4.135)	(5.595)
College	34.28	30.42	42.20	37.61	52.05	41.77	53.52	41.40
	(47.50)	(46.04)	(49.40)	(48.45)	(49.98)	(49.34)	(49.91)	(49.29)
Weekly paid hours	25.85	41.16	36.83	36.01	37.35	38.42	33.29	39.45
	(14.02)	(10.53)	(6.163)	(9.921)	(5.150)	(4.251)	(9.940)	(10.41)
Weekly housework hours	12.12	3.403	10.01	3.829	7.072	6.894	5.810	8.203
	(6.617)	(2.792)	(5.402)	(3.061)	(3.978)	(3.637)	(3.988)	(4.127)
Gross monthly income	1129.8	2125.9	1696.6	2077.2	1843.0	2293.3	1865.4	2377.1
	(1029.5)	(1448.9)	(1001.3)	(1405.6)	(988.7)	(1256.2)	(1169.3)	(1479.9)
Gender norm attitudes	0.237	-0.0168	0.558	0.330	0.514	0.282	0.534	0.410
	(0.927)	(0.796)	(0.827)	(0.790)	(0.762)	(0.818)	(0.928)	(0.919)
Observations	676	687	1721	1735	1054	1044	832	836
Panel B: Couple characteristics								
Tenure	4.163		4.006		4.119		3.956	
	(3.544)		(3.281)		(2.842)		(2.698)	
Married (%)	61.54		59.44		49.91		53.49	
	(48.69)		(49.11)		(50.02)		(49.91)	
At least one child (%)	2.515		0.0581		0.190		1.563	
	(15.67)		(2.411)		(4.354)		(12.41)	
Female share of paid hours (%)	36.78		52.63		49.25		46.19	
	(8.705)		(9.552)		(1.823)		(11.28)	
Female share of unpaid hours (%)	77.29		72.47		50.47		39.87	
	(11.04)		(11.39)		(5.409)		(11.62)	
Gross monthly income	3530.6		3937.1		4250.4		4405.9	
	(2071.3)		(2217.6)		(2030.1)		(2267.3)	
Urban (%)	81.42		78.45		78.48		78.33	
	(38.98)		(41.15)		(41.14)		(41.24)	
Observations	696		1770		1064		861	

Standard errors in parentheses.

C Empirical Strategy

We use a sample of individuals who become parents to study the impact of the birth of the first child (treatment variable) on RQ (outcome variable). The treatment is staggered because individuals have their first child birth in different periods. This divides the sample into different cohorts of individuals depending on the calendar year when they become parents for the first time. Having a sample of new parents implies that everyone is treated at some point. The control group at each period is formed by individuals who have not become parents yet. Add something about comparability of parents with never parents.

Most of the analysis focuses on the estimation of dynamic effects. This allows for heterogeneity in the treatment effect with time since the treatment started.

An essential condition for the estimation of the specification in Equation 1 is that the treatment is absorbing. We feel comfortable making this assumption in our context, since the treatment is having a child. The only way in which this would not be absorbing is if the child dies, which would bring a bunch of other things along. Look at variable lchdy4. There are not many instances of natural child death in the dataset.

C.1 Callaway and Sant'Anna (2021) method

The goal of using this method is to overcome the problems derived from two way fixed effects estimation in a context of staggered treatment adoption (See for instance Goodman-Bacon, 2021, for a review.). There are two main issues of this estimation. First, it requires homogeneity of treatment effects across treatment cohorts. The violation of this assumption induces negative weights when computing the estimates of the average treatment effects, resulting in biased estimates. Second, it carries out *forbidden* comparisons between individuals changing status from control to treated and already treated cohorts.

The method proposed by Callaway and Sant'Anna (2021) overcomes these issues by clearly separating identification, aggregation and estimation in the procedure. As mentioned above, it requires a panel dataset where the binary and absorbing treatment is adopted in a staggered manner. It requires two assumptions:

A1. Limited anticipation. If a unit is untreated in period t, its outcome in that period does not depend on when it will be treated in the future. In our context, we need to assume that changes in RQ before the birth of the first child do not predict when individuals have their first child. They allow for limited anticipation, that is, there can be some preceding reaction but we need to know (or make an assumption on) how many periods before treatment δ.

A2. (Conditional) parallel trends based on "not-yet treated". All treatment cohorts would have evolved in parallel in absence of treatment. In our context, individuals' RQ would have evolved in parallel regardless of the period when they had their first child. They allow for parallel trends conditional on covariates.

Under those two assumptions, they can identify the average treatment effect for the treated (ATT) for each treatment cohort g and at each period t, which they denote as $ATT(g,t)$. The assumptions allow them to construct the counterfactual in each of these cases using (i) the period before treatment as the baseline period, and (ii) all treatment cohorts that have not been treated by t. Therefore, the control group at each t for the same cohort g varies because at each subsequent period new cohorts enter treatment status. Each $ATT(g,t)$ is estimated as a 2×2 difference-in-differences coefficient using the control group described. The estimation can be done in 3 ways: using outcome regression, inverse probability weighting, or doubly robust estimands. We use the last method in the estimation.

$ATT(g,t)$-s are the building blocks used to summarize the treatment effects across treatment cohorts, period or time relative to treatment. They propose different aggregation methods to allow for heterogeneity in those three dimensions. The main results of this paper are the product of aggregating $ATT(g,t)$ at the event-time level.

C.2 Identification

Although the assumptions stated above cannot be directly tested, we can provide some evidence in favor of their plausibility.

The no anticipation assumption would be violated, for instance, if individuals decided to have children in response to a negative shock to RQ. Figure 3 indicates that, on average, there is no evidence that the event (the birth of the first child) is preceded by deviations in RQ, since all the point estimates before the event are not statistically different from zero. However, this could be masking heterogeneity in the evolution of RQ before first child birth, averaging individuals that decide to have children after a positive or a negative RQ deviation. If that were the case, the standard deviation of RQ would be larger over the periods before the event. Figure C.1 plots the mean and the standard deviation of RQ at each event time. The standard deviation of RQ increases after the birth of the first child, supporting the evidence against anticipation. To do: Divide into groups depending on the pre-trend and see if the evolution of RQ is the same for both types.

The flat pre-trends shown in Figure 3 are also suggestive evidence in favor of the parallel trends assumption. To be continued...

C.3 Choice of controls

The Callaway and Sant'Anna (2021) method allows including covariates at pre-treatment levels. In the main specification we do not include any covariates. We verify that the

Figure C.1: Sample mean and standard deviation of RQ around first child birth

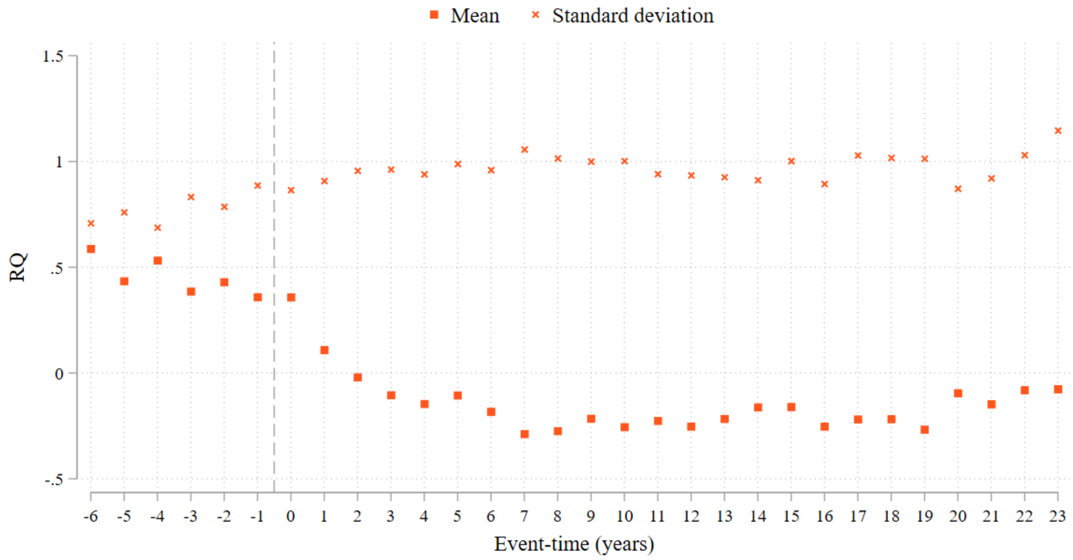
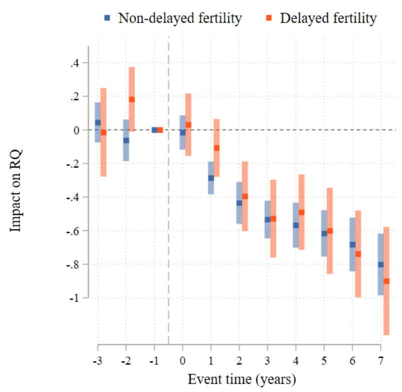
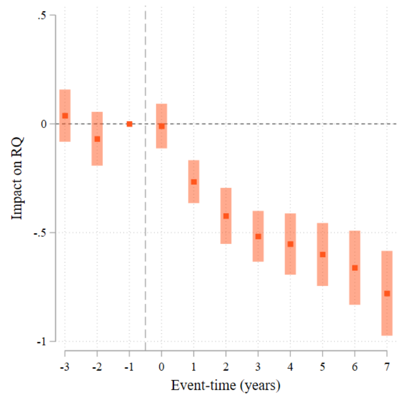


Figure C.2: Delayed fertility vs non-delayed fertility

(a) Marginal effects by delayed fertility



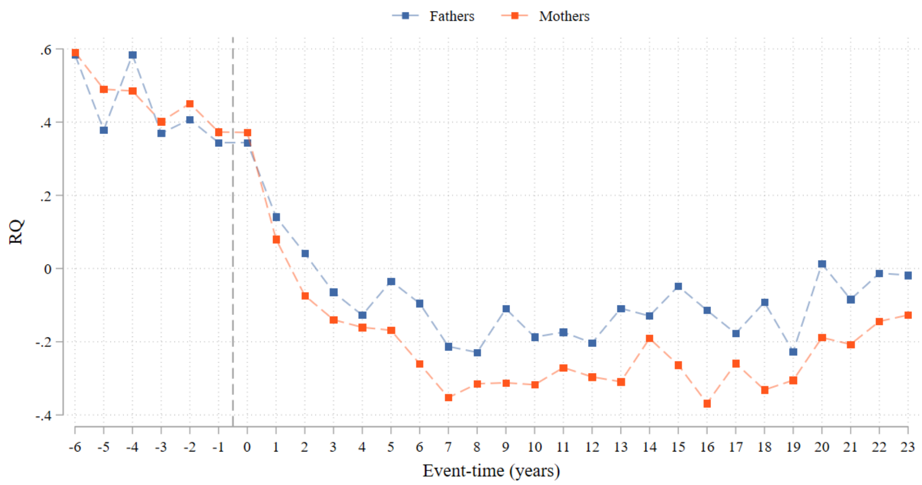
(b) Effects for non-delayed fertility



results are not driven by heterogeneity in omitted covariates.

Sex. As shown in Subsection A.2, there is large correlation between wife and husband's RQ. We verify that the impact of having a child is also parallel for both couple members. In Figure C.3 we plot the average RQ at each period relative to the birth of the first child for fathers and mothers separately. The decrease in RQ after birth is similar for both, being always about 0.1 standard deviations larger for women. In Figure C.4 we plot the results from carrying out the main analysis separately for men and women. In line with the correlation results, we find that the impact of having a first child on RQ becomes slightly larger for mothers than for fathers four years after birth.

Figure C.3: Average RQ per event-time period, by sex



Notes: This figure plots the average RQ at each event-time period by sex.

Figure C.4: Effects of first child birth on RQ, by sex

Notes: This figure plots the impact of first child birth on RQ at each event-time period by sex.

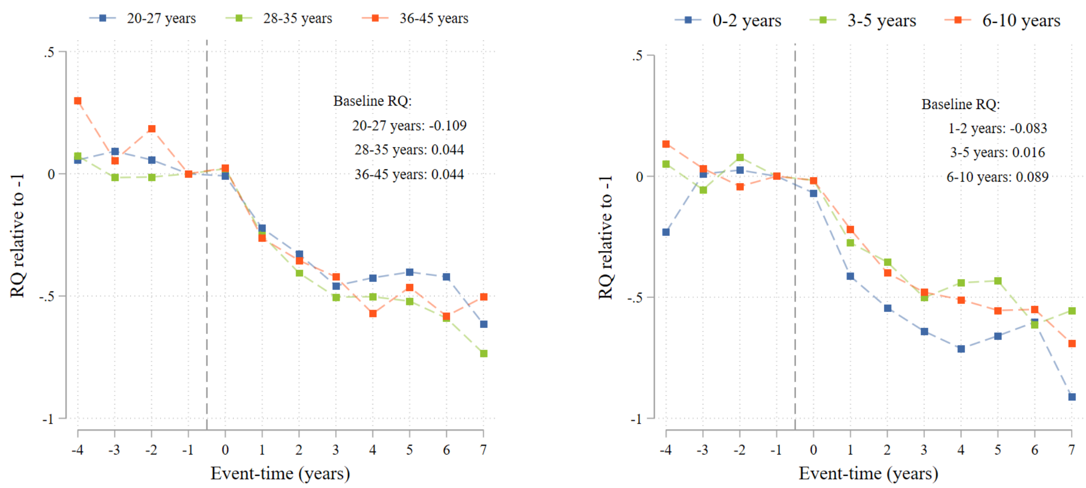
Age and tenure. Arguably, individuals having children at different stages of their lives or relationships might have different experiences from it. We study whether individuals having children at different ages and relationship tenures experience different changes in RQ. We divide individuals into groups based on the age and the relationship tenure at the time when they had their first child. Figure C.5 plots the average RQ at each event-time period by (a) age bin and (b) tenure bin, normalizing RQ the period before birth to zero. Although individuals having children at different ages have different baseline levels of RQ, the evolution of this variable after the birth of the first child is indistinguishable across groups.

Baseline levels also differ across tenure groups. In this case, it does look like individuals having their first child at the very beginning of their relationship experience a greater decrease in RQ. As seen in Figure A.4 (b) the decrease in RQ is largest over the first few periods of the relationship. Thus, the greater decrease depicted in Figure C.5 (b) could be partially capturing this decreasing general trend. There is no noticeable difference among the other two groups.

Table C.1 displays the results from a static estimation using the Callaway and Sant'Anna (2021) method. The analysis is done separately for each age and tenure bin. This analysis confirms the conclusions drawn from the descriptive study above. Whereas there are no relevant differences in the impact of the birth of the first child for individuals becoming parents at different ages, the relationship tenure at which the first child is born is relevant. Namely, the impact is much larger for individuals becoming parents shortly after their relationship started.

Figure C.5: Average RQ per event-time period

(a) Age bins (b) Tenure bins



Notes: This figure plots the average RQ at each event-time period, after normalizing the baseline period (-1) to zero, by (a) age and (b) tenure bin.

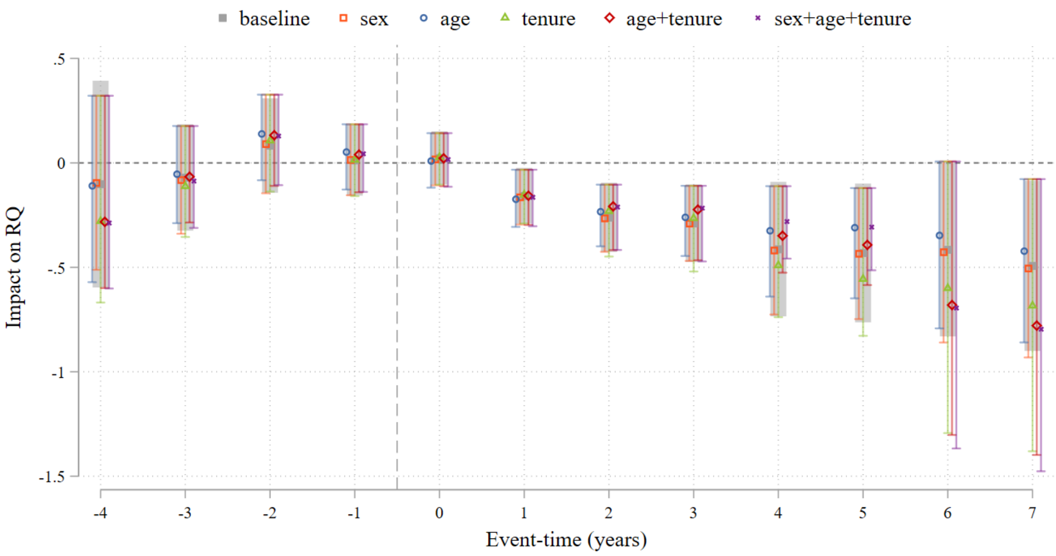
Table C.1: Static ATT by age and tenure bin

	Age bins			Tenure bins		
	20-27 years	28-35 years	36-45 years	0-2 years	3-5 years	6-10 years
ATT	-0.251	-0.276	-0.234	-0.585	-0.123	-0.136
(sd)	(0.258)	(0.095)	(0.084)	(0.156)	(0.173)	(0.110)

Notes:

Analysis including controls. Lastly, we verify that the results from our main analysis do not vary significantly after including controls. Figure C.6 displays the results from the main analysis using different sets of controls. The differences for all sets of controls are negligible until six years after the birth of the first child. At that point the coefficients of the specifications including tenure become larger. Thus, our main results could be interpreted as a lower bound.

Figure C.6: Impact of RQ using different sets of controls



D Robustness

D.1 Different estimation strategies

Figure D.1: Effects of first child birth on RQ, using TWFE

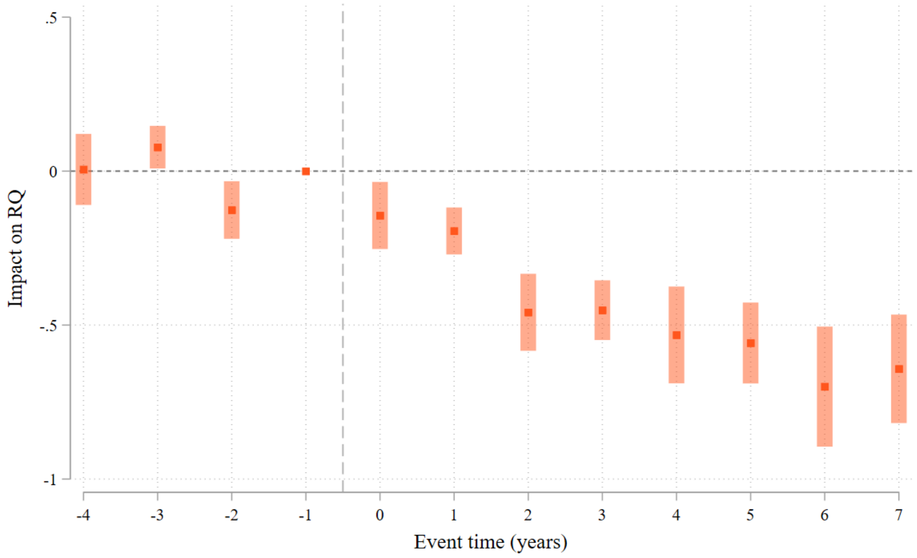
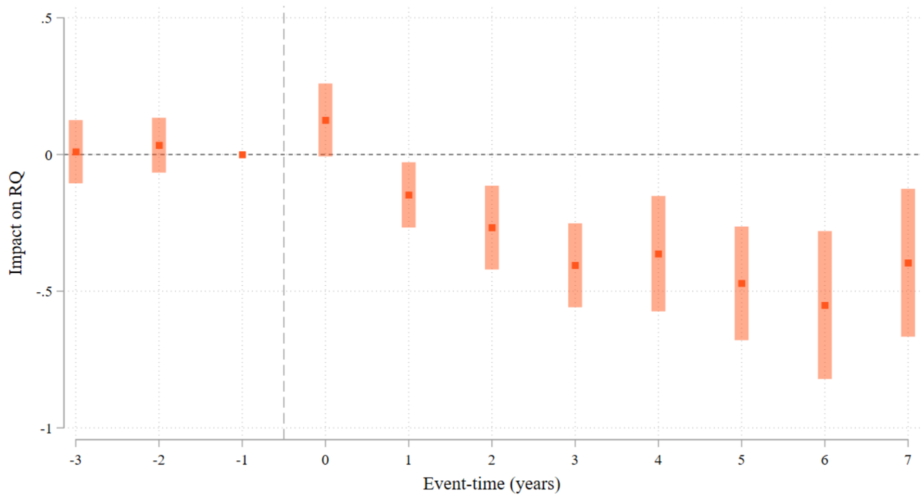


Figure D.2: Effects of first child birth on RQ, using Kleven et al. (2019)



D.2 Alternative measures of RQ

Figure D.3: RQ by item block

(a) Subjective assessment (b) Couple time use

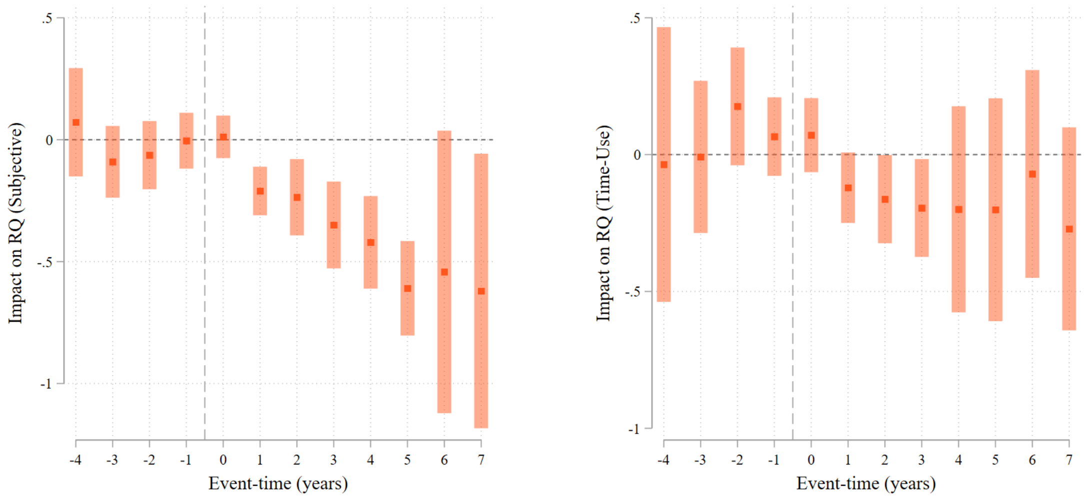


Figure D.4: RQ using factor scores after birth

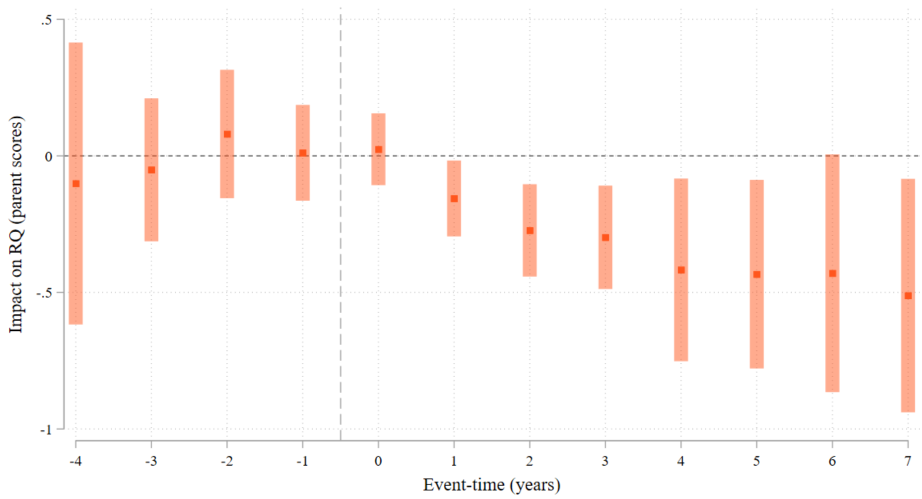
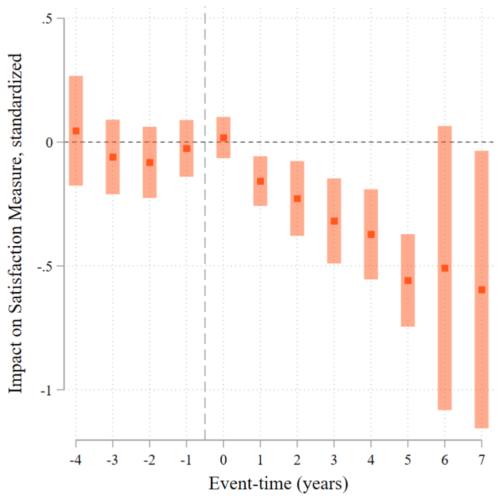
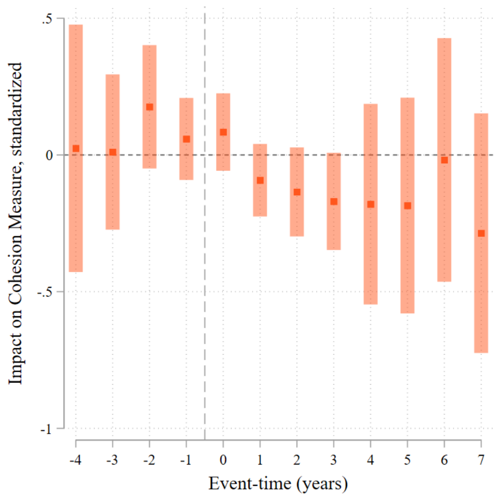


Figure D.5: Psychology measures

(a) RDAS satisfaction



(b) RDAS cohesion



D.3 Different subsamples

Figure D.6: Effects of first child birth on non-separating couples.

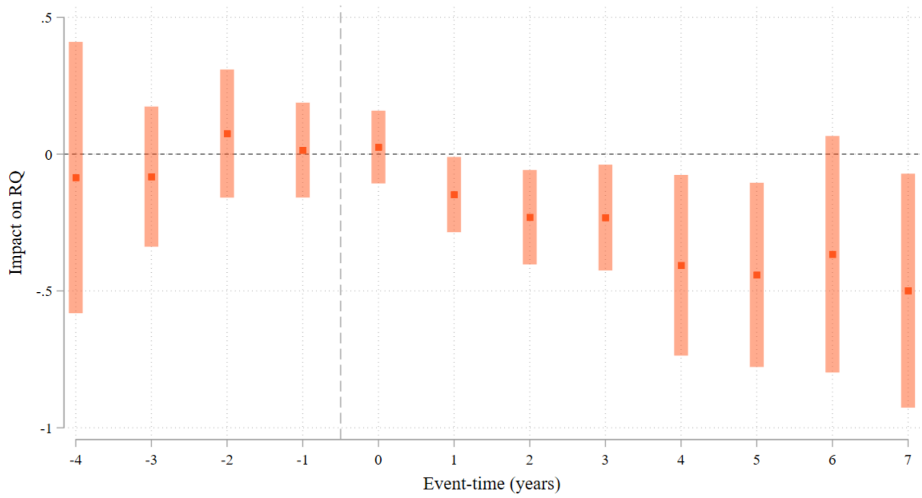


Figure D.7: Impact by final number of children.

(a) One child (b) Two children (c) Three + children

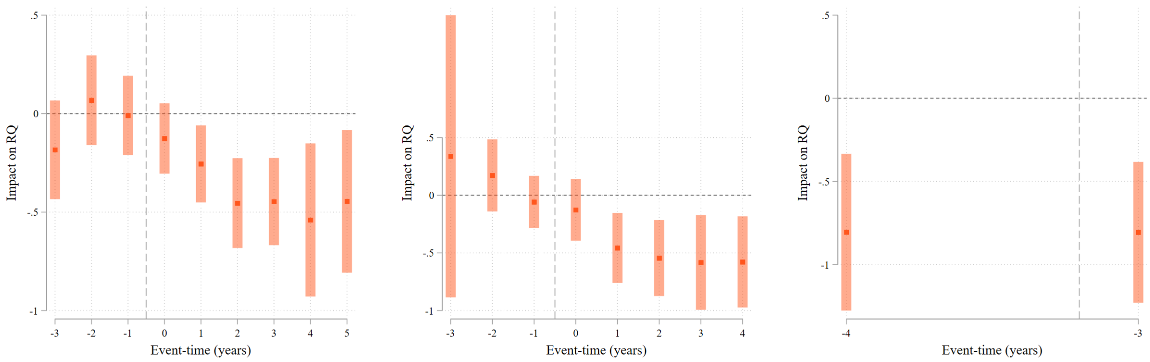
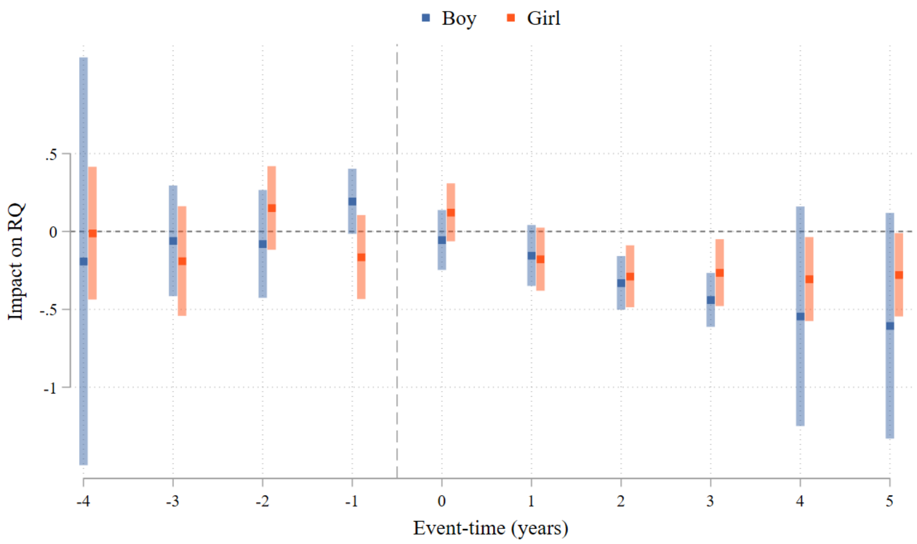


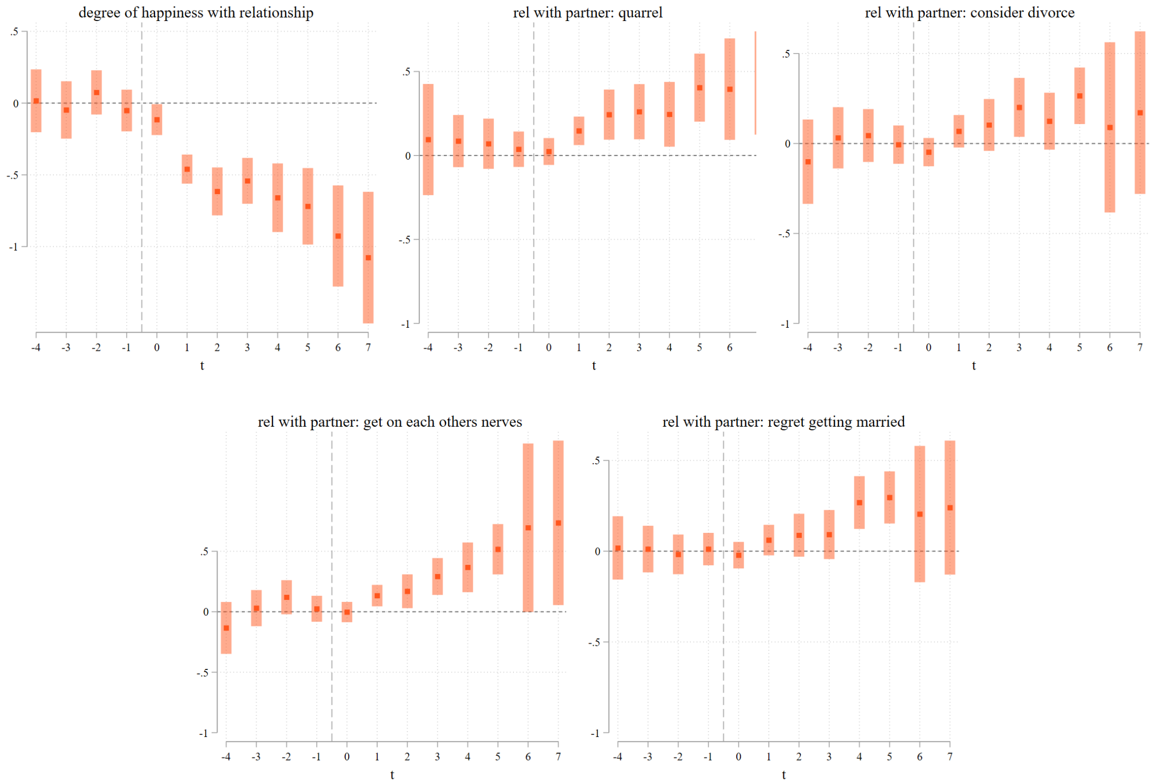
Figure D.8: Effects of first child birth on RQ, boys vs. girls.



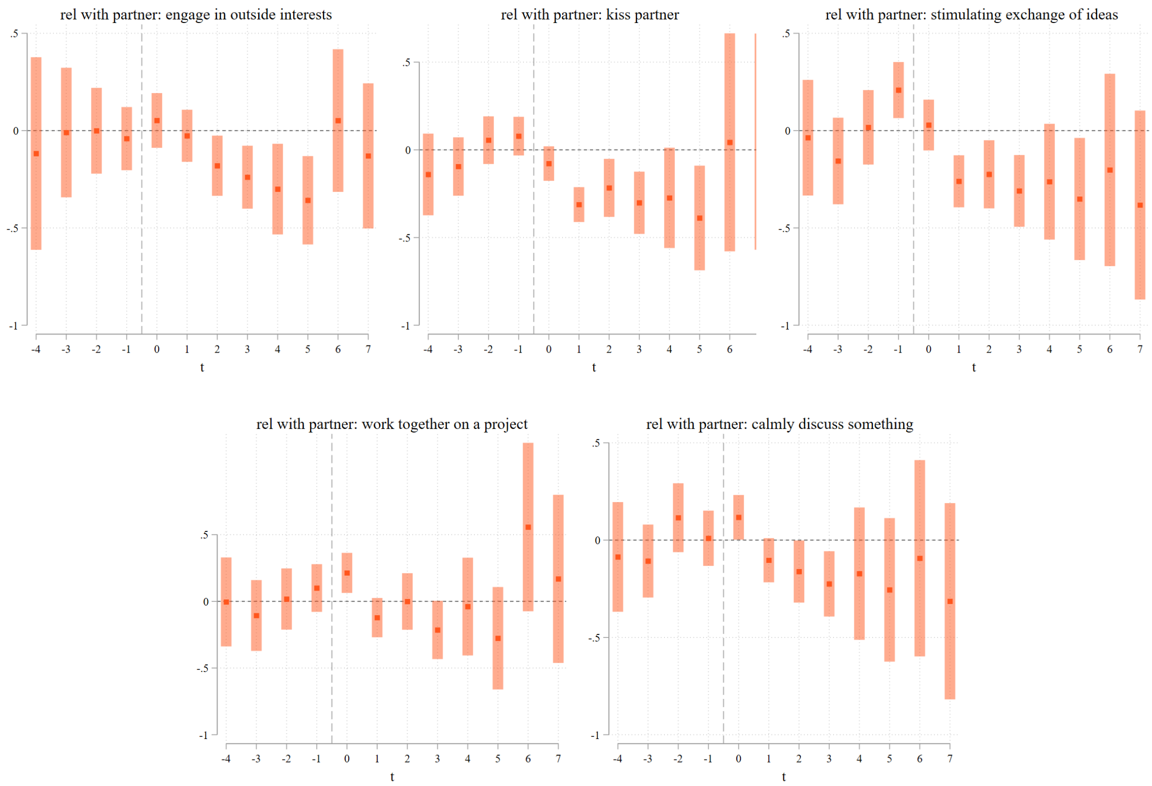
E Event Study per item

Figure E.1: Impact of first child on each item.

(a) Subjective assessment



(b) Couple time use



F Expansion of childcare

Table F.1: Estimates of childcare expansion on mothers' time use

	(1) Paid hours	(2) Paid hours	(3) Housework hours	(4) Housework hours	(5) RQ	(6) RQ
Available	-0.439	-2.191	-0.465	-0.485	-0.151	-0.070
	(3.704)	(3.425)	(2.440)	(2.504)	(0.166)	(0.176)
Medium	1.498	0.005	-2.765	-3.031	0.247	0.274^{*}
	(3.260)	(3.000)	(1.795)	(1.836)	(0.171)	(0.158)
Post	-1.441	4.316	0.807	2.152	0.443^{**}	0.301
	(2.498)	(3.437)	(1.860)	(4.011)	(0.203)	(0.236)
Available $\times Post$	6.530	5.705	0.459	1.376	-0.727	-0.717
	(5.429)	(5.363)	(3.907)	(3.915)	(0.563)	(0.583)
Medium $\times Post$	-0.166	-0.165	-1.866	-0.780	-1.079^{**}	-1.116^{**}
	(4.431)	(4.068)	(2.641)	(2.692)	(0.427)	(0.459)
Controls	-	Yes	-	Yes	-	Yes
Observations	338	330	201	195	227	222

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports the marginal effects from estimating Equation 2 by couple type and gender. Standard errors clustered at the couple level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.