Children and Relationship Quality*

Belén Rodríguez Moro† Olatz Román Blanco‡

June 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 90^{th} 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.

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, BSE Jamboree 2023, Universidad de Salamanca and EDP Jamboree 2023 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. We explore two channels. First, there is a decrease in time spent together as a couple, which we observe in Couple time RQ. We distinguish couples based on the values of this measure before child birth and estimate the differential impact of fertility on RQ. We find that couples that used to spend more time together are more negatively impacted, both in Couple time and Subjective RQ.

Second, 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.

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 and sample

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.

The population of interest consists of individuals in cohabiting relationships, either married or not, who become parents. Due to the nature of our empirical strategy, we use the sample of individuals who have their first child (become *new-parents*) during the observation period. We restrict to couples in which the mother was between sixteen to fifty years old by the time their first child is born. We also exclude couples that live with children from previous relationships to ensure that couples had no previous time arrangements involving childcare. The resulting sample is an unbalanced panel of 9,269 individuals observed up to five times.

2.2 Measures 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 calmly discuss something?" or "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. 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.

Table 1: Questions in the Understanding Society Partner module.

(a) Subjective assessment

How often do you... ?	
consider splitting	6pt, freq (-)
regret getting married	6pt, freq (-)
quarrel	6pt, freq (-)
get on each others nerves	6pt, freq (-)
What is the... ?	
degree of happiness	7pt, degree (+)

(b) Couple time use

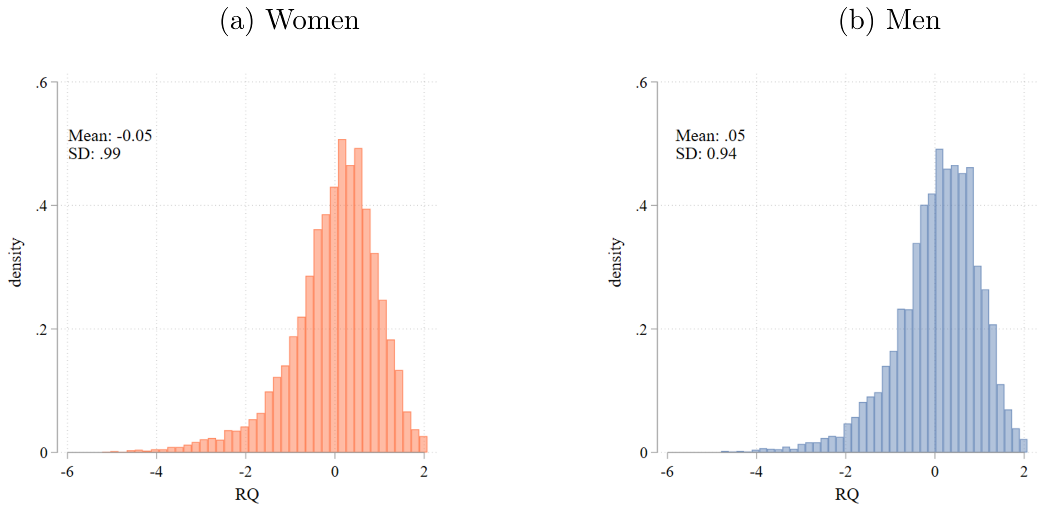
How often do you... ?	
work together on a project	6pt, freq (+)
stimulating exchange of ideas	6pt, freq (+)
calmly discuss something	6pt, freq (+)
kiss partner	6pt, freq (+)
Do you and your partner... ?	
engage in outside interests	5pt, amount (-)

We distinguish between two types of items in the questionnaire, based on the information they convey. Table 1 (a) lists the items that refer to as *subjective assessment* items, which are related to the degree of happiness and conflict in the relationship. The items in Table 1 (b) are referred to as *couple time use* items, since they inform of the way in which the members of a couple use their time together.

We 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.[1] The resulting variable is centered at zero and has a unit standard deviation. Higher values indicate a better relationship.

[1] All the factor loadings are reported in Table A.1.

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.

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 right, 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.1.

Validity of the measure. Appendix B 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 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

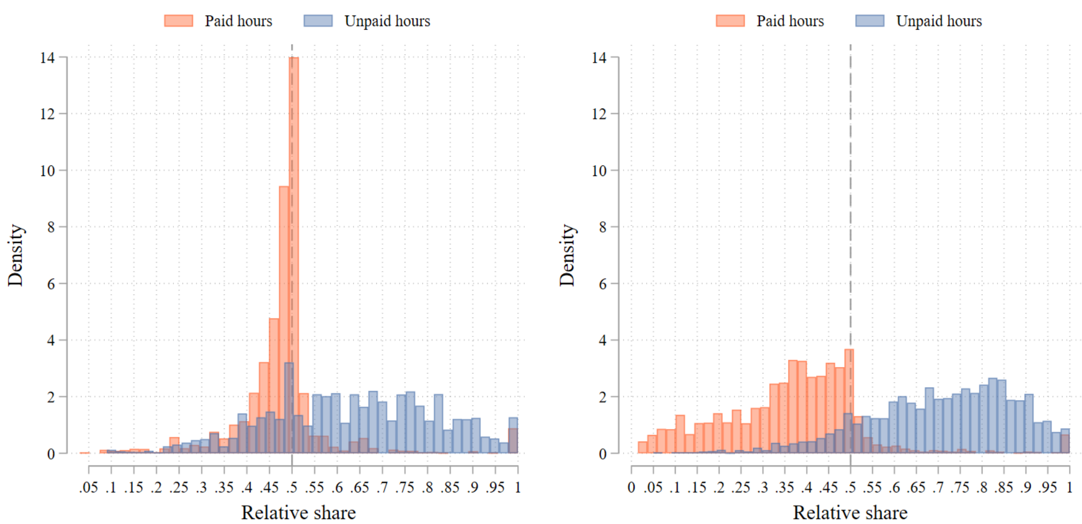
We study household specialization making use of the time use variables provided by Understanding Society. The data provides two time-use variables at the individual level:

hours spent on housework and number of hours normally worked per week.[2] 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.

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

(a) Before birth (b) After birth



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)

[2]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.

counter-traditional couples, where men take the largest share of housework. Table D.1 in Appendix D 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.[3] 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 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.

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

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.

3 Empirical Strategy

3.1 Event Study design

We follow the literature on child penalties led by Kleven et al. (2019) and 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 event (treatment) of first child birth. Below we discuss the assumptions under which this approach leverages quasi-random variation in the outcome. The main benefit of this methodology is that it allows for heterogeneity of treatment effects with time relative to the event. It provides the dynamic impact of children and its potential persistence, once individual and couple decisions accommodate to child birth.

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. 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 3 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:

$$\begin{aligned} y_{i,t} = & \sum_{j\neq -1} \mathbb{1}\{j = t - G_i\}\delta_j + \sum_a \mathbb{1}\{a = \text{age}_{i,t}\}\alpha_a + \sum_d \mathbb{1}\{d = \text{tenure}_{i,t}\}\gamma_d + \\ & + \sum_w \mathbb{1}\{w = t\}\psi_w + \mathbf{X}_{i,t}\boldsymbol{\beta} + v_{i,t} \end{aligned} \tag{1}$$

where we include the full set of dummies for time relative to first child birth ($\mathbb{1}\{j = t - G_i\}$), age dummies ($\mathbb{1}\{a = \text{age}_{i,t}\}$), relationship tenure dummies ($\mathbb{1}\{d = \text{tenure}_{i,t}\}$) and period dummies ($\mathbb{1}\{w = t\}$). The inclusion of age, tenure and period fixed effects accommodates trends in RQ along these three time-varying dimensions.[4] Finally, $\mathbf{X}_{i,t}$ includes gender, college education, and area of residence. Given that RQ is standardized, the coefficients are interpreted in terms of standard deviations. We cluster the standard errors at the couple level. We estimate this equation by standard OLS estimation.

An important consideration is that all individuals in the sample are observed becoming parents and individuals who do not have children in the observation period are not considered. Conditional on the set of controls, we leverage variation in the period in which individuals become parents to estimate the causal effect of first child birth on RQ. Each time event dummy $\mathbb{1}\{j = t - G_i\}$ provides changes in RQ relative to the period before child birth (baseline). At each period t we pool comparisons between all treated individuals, with not-yet-treated and already-treated individuals. In the next section we discuss the concerns associated with establishing this type of comparisons in the identification of the causal effect.

3.2 Identifying Assumptions

The parameters of interest are δ_j, which provide the dynamic impact of having the first child on RQ. The causal interpretation of these as the average treatment effect on the treated (ATT) relies on a number of assumptions.

First, we restrict treatment anticipation. This means that we assume that the birth of the first child is not preceded by changes in RQ. We verify this by looking at the evolution of RQ during the periods before child birth, in Figure 3. All the leads to child birth are precisely estimated zeros, which strongly suggests no anticipation effects in RQ at least three periods before child birth.[5]

[4] We show the evolution of RQ across age and relationship tenure in Appendix C for the full population of cohabiting couples.

[5] We plot the standard deviation of RQ around first child birth to verify whether flat pre-trends are driven by extreme values of our measure cancel out when averaging them. Where?

Second, we need to impose some restrictions on the evolution of untreated potential outcomes. We do so through a conditional parallel trends assumption: conditional on covariates, the evolution of the average outcomes of treated and untreated units at any point would have been the same in the absence of treatment. In our setting, this implies that for the subset of new parents at period G, the evolution of RQ would have been the same as that of new parents at period $G + \epsilon$, had fertility never occurred. Verifying this assumption would imply observing the counterfactual scenario of unvoluntary non-fertility. We proxy this scenario using the group of parents that either declare having used a fertility treatment or had an involuntary pregnancy interruption prior to their first child being born.[6] Figure E.2 panel (a) shows the marginal effects of first child birth for the two different groups of parents, using our main specification. Differences are not statistically significant between the two groups, suggesting the plausability of the conditional parallel trends assumption.[7] Quote Lidia's paper about IFT?

Finally, the causal interpretation of $\hat{\delta}_j$ requires homogeneous treatment effects across treatment groups G_i. Namely, it relies on the assumption that the evolution of RQ after first child birth is the same for individuals that had their first child in different calendar years. We relax this assumption by implementing the methodology proposed by Callaway and Sant'Anna (2021). This method is reviewed in Appendix H.1.

4 Impact of children on RQ

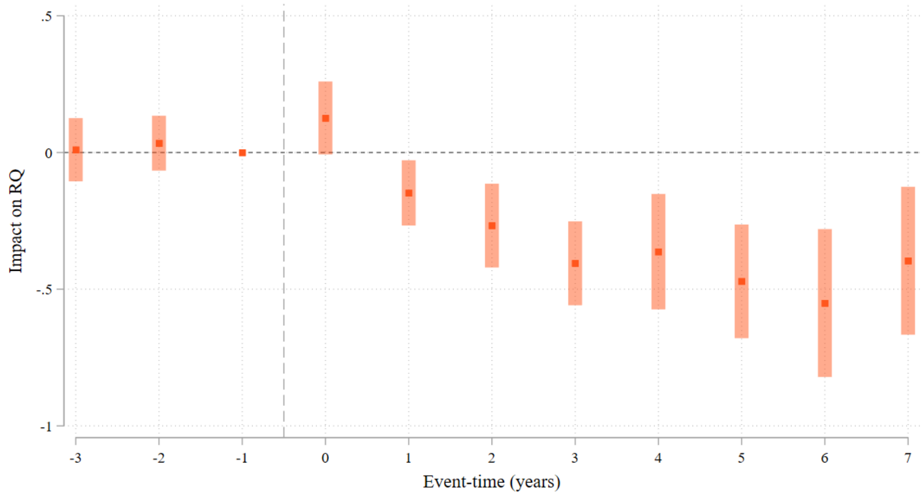
Figure 3 depicts the estimated effect of first child birth on RQ at each period relative to birth. The coefficients corresponding to periods before birth are not significantly different from zero. This confirms that the decision to have a child is not endogenous to the evolution of RQ after controlling 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)

[6] We use women respondents to the fertility module provided by UKHLS during waves 2-12. We input the answers to the male partners of these women and classify couples that receive any fertility treatment as the control group.

[7] Although differences are not statistically significant, parents that experience a delay in fertility seem to have slightly higher values of RQ two periods before child birth. We repeat the main analysis on the sub-sample of parents that do not suffer delayed fertility under our classification (See Figure E.2 panel (b)). Our main result is robust to this exercise, suggesting that the slightly positive value in RQ preceding the conception of the first child among parents with delayed fertility is not biasing our estimates.

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.

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.[8]

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.

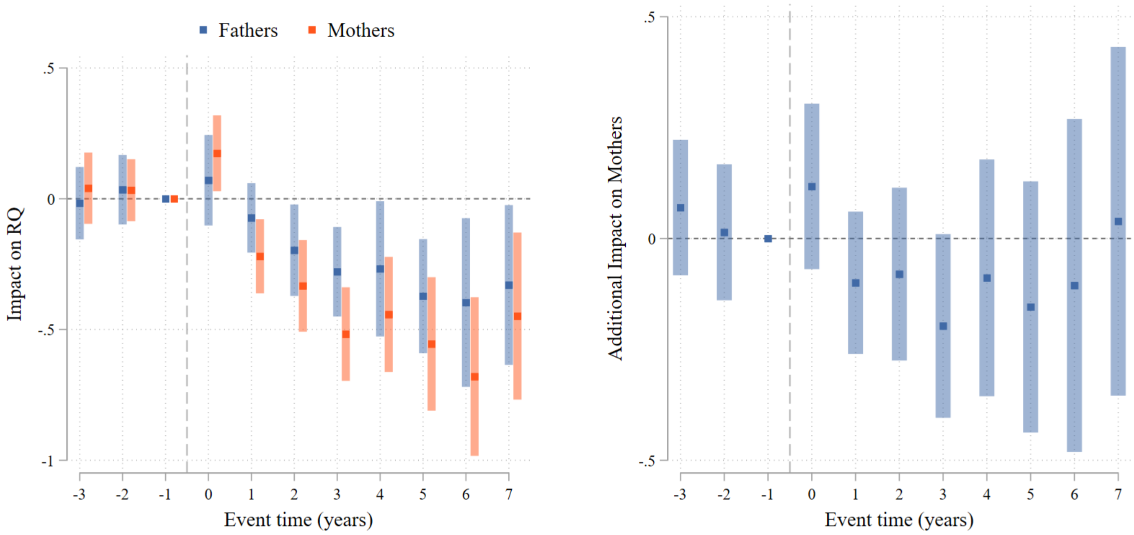
The TWFE estimated results in Figure F.1 are even more negative than the ones found using our main specification.[9] The decrease in RQ is more sustained with time relative to first child birth. One potential explanation relates to the OLS problem in dynamic

[8]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.

[9]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 Figure F.2, respectively.

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.

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 Figure F.3. The impact is largely sustained although more imprecisely estimated.[10]

Next, we use alternative measures of relationship quality. In Figure F.4 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 F.5 indicates no differences in the estimated impact using this measure. Last, we use some similar measures from the psychology literature in Figure F.6 and observe similar effects.

[10]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.

Finally, we use different subsamples to address potential sample selection issues. First, in Figure F.7 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 F.8 repeats this analysis using subsamples of parents depending on their total lifetime fertility. 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 F.9 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 F.10. 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.

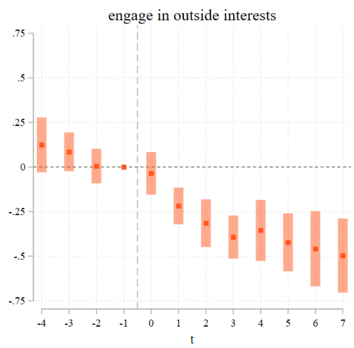
5.1 Quality time together

First child birth has a direct impact on the amount of available time. Thus, we should observe individuals reduce the amount of quality time together with their partner, which would mechanically reduce RQ. We quantify the relative relevance of quality time together by looking into the couple time use items in Table 1. We repeat the event study on each item separately using the usual set of controls. We plot the results in Figure 5. Lower levels of these variables indicate lower frequency.

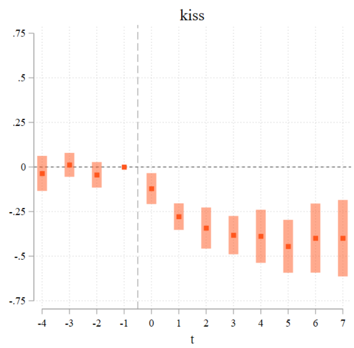
There is a significant drop in the frequency with which couples engage in outside interests, kiss and have stimulating exchanges of ideas. The largest impact is on kissing, indicating a decrease in intimacy after the birth of the first child. There is a small and insignificant decrease in working together on a project and calmly discussing something. These activities could be related to the child. In Figure G.2 in Appendix G we verify that the changes observed are symmetric across genders.

Figure 5: Impact of first child on couple time use items

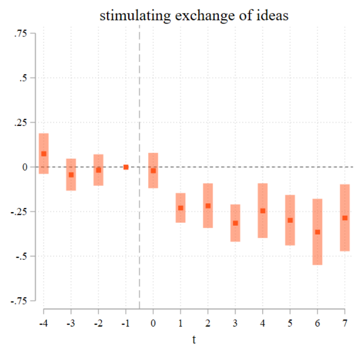
(a) Engage in interests



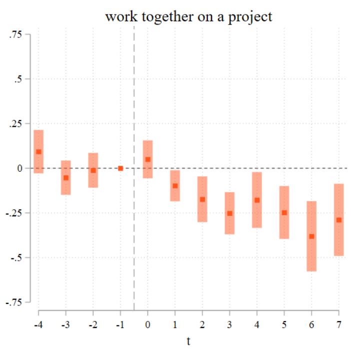
(b) Kiss



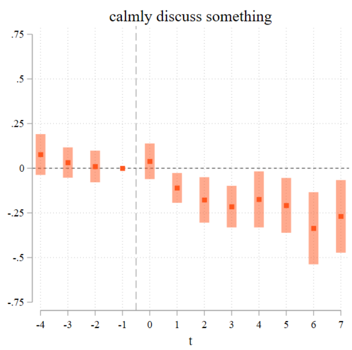
(c) Exchange of ideas



(d) Work on a project



(e) Discuss



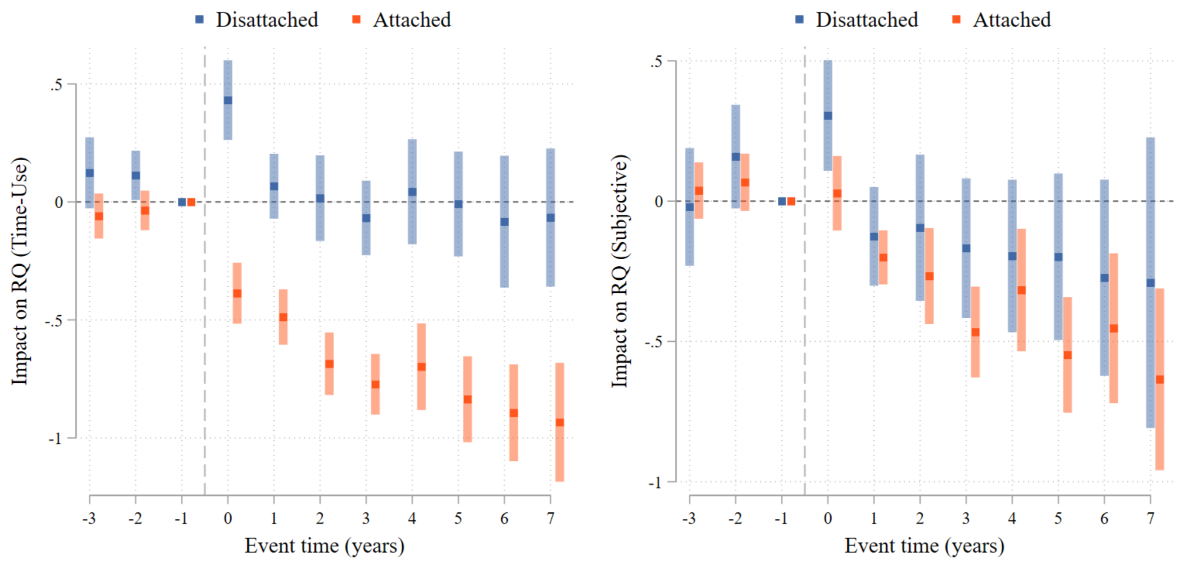
Notes: This graph plots the estimates of an event study of first child birth on the couple time use items in Table 1. The period prior to birth is taken as baseline. The plotted coefficients are the effects on each item of leads and lags around the event. Confidence intervals are estimated at the 95% level. All variables are standardized and increasing in frequency.

To assess the share in the decrease of RQ that can be explained by the reduction in quality time together, we first construct two new RQ measures: one using only the time use block in Table 1 and another one on the subjective assessment block. We call the resulting measures *time RQ* and *subjective RQ*, respectively.[11] As displayed in Figure 6, first child birth has a significant impact on both measures separately. Thus, despite the documented decrease in time use together, this cannot explain the entire drop in RQ.

However, the responses to both blocks may be correlated. We exploit the heterogeneity in couple time use to quantify the share of the impact on subjective RQ explained by

[11] Show factor loadings for this too.

Figure 6: Impact of first child birth by time RQ group



Notes: This graph plots the estimates of an event study around first child birth on RQ (Subjective) and on RQ (time-use), by time RQ group. We divide couples depending on their average time RQ *before* the birth of their first child. We refer to couples in the first two quintiles as disattached and to those in the last two quintiles as attached. 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.

differences in time use. We divide couples into quintiles depending on their average time RQ *before* the birth of their first child. We refer to couples in the first two quintiles as disattached and to those in the last two quintiles as attached. Table D.3 in Appendix D shows that the two types of couples are similar in the usual set of individual and couple characteristics.

In Figure 6 (a) we check whether time use was differently impacted by the birth of the first child for disattached and attached couples. Strikingly, we see that disattached couples spend more time together the period when their first child is born and then return to the baseline level. This stands in stark difference with what we observe for attached couples, who reduce their time together at increasing rates.

In Figure 6 (b) we see if there is also a differential impact on subjective RQ for these two groups of people. The impact is worse for the attached although differences between the two groups are not statistically significant. The observed pattern is similar to that of the time RQ measure, indicating that there is a correlation in the changes between these two measures.

In Table I.1 in Appendix I, we show that the impact of first child birth on the other time outcomes considered (paid and housework time) is similar for both types of couples. This result shuts down the possibility that the observed differential effects across attached and unattached couples are induced by household changes in specialization.

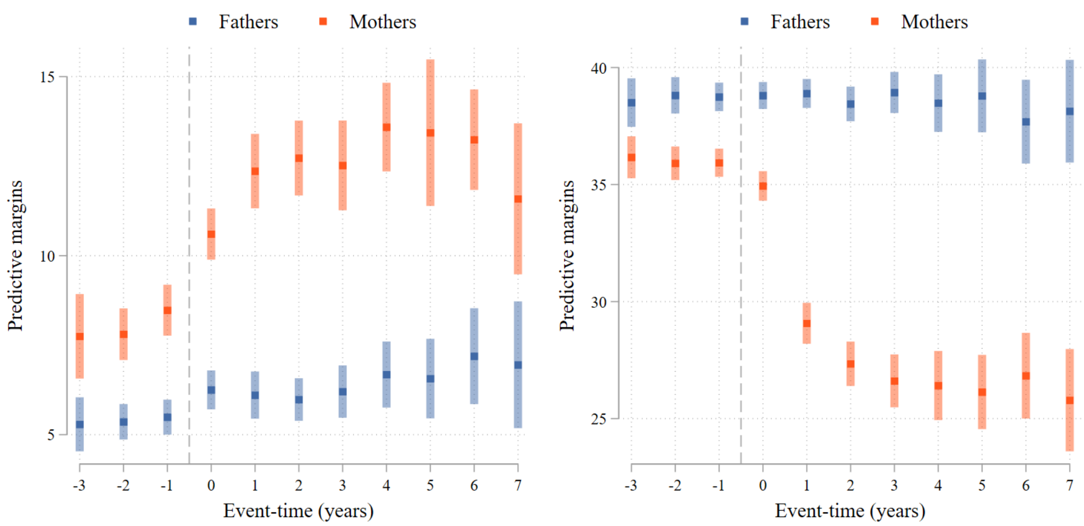
5.2 Increase in housework and time rearrangement

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.[12] Figure 7 (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.

Figure 7 (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 7: 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

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

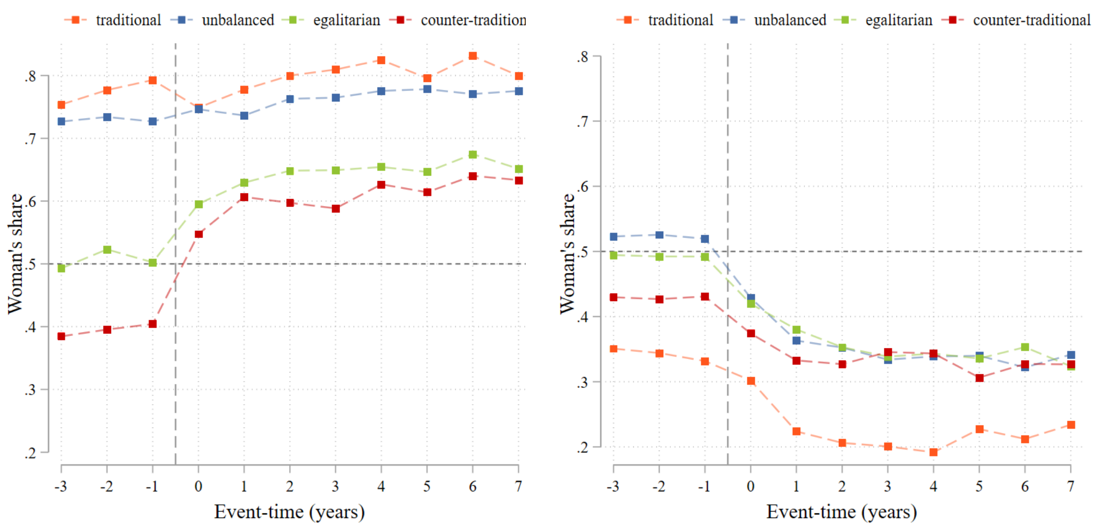
child birth, as explained in Subsection 2.3. We verify whether the female share evolves differently across couple types.

In Figure 8 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 8 (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 8: 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 second row of Table 3 reports the marginal effects by couple type from estimating Equation 2.[13] 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. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.3 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. Table D.2 in Appendix D summarizes the average values

[13] 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.

Table 4: Regression analysis by couple type and gender

	Housework hours		Paid hours	
	Women	Men	Women	Men
Traditional	5.979***	0.833***	-10.75***	-2.344***
	(0.507)	(0.237)	(0.936)	(0.647)
Unbalanced	5.038***	0.760***	-12.52***	1.465**
	(0.431)	(0.184)	(0.616)	(0.445)
Egalitarian	5.456***	-0.0395	-12.17***	-0.467
	(0.464)	(0.265)	(0.755)	(0.360)
Counter-traditional	5.558***	-0.914**	-10.30***	-0.637
	(0.544)	(0.349)	(0.788)	(0.691)
Observations	7758	7632	10165	9717

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

Table 5: 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. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

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 6. 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 6 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.

Table 6: Impact of first child birth on RQ measures, by income quartile

	(1)	(2)	(3)	(4)
	bottom	second	third	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.

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*.

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 Factor Analysis

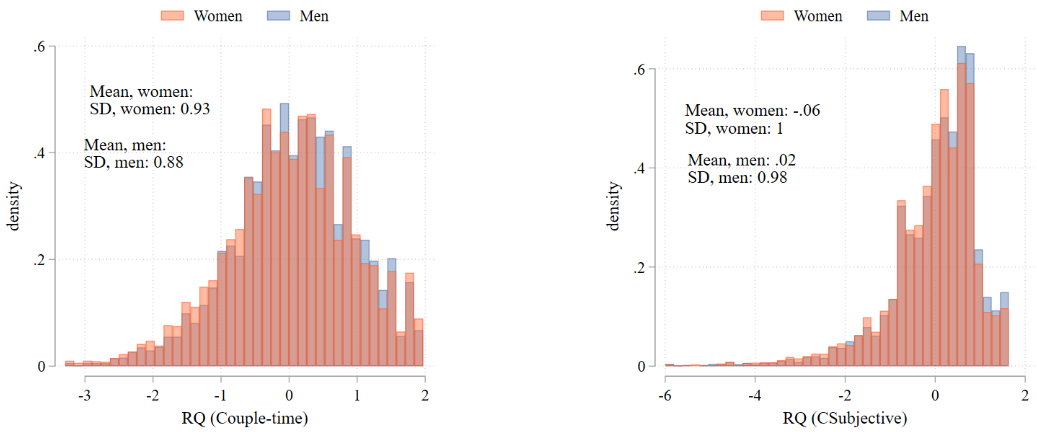
Table A.1: Factor loadings of RQ.

(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

Note: 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: Distribution of RQ by item block

(a) Time RQ (b) Subjective RQ



Notes: This figure plots the distribution of (a) time RQ and (b) subjective RQ in the sample. The mean of both measures in the full data is 0 and their standard deviation is 1. Correct the labels. We should always call these measures the same> subjective RQ and time RQ.

B Validity of RQ

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 B.1 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.

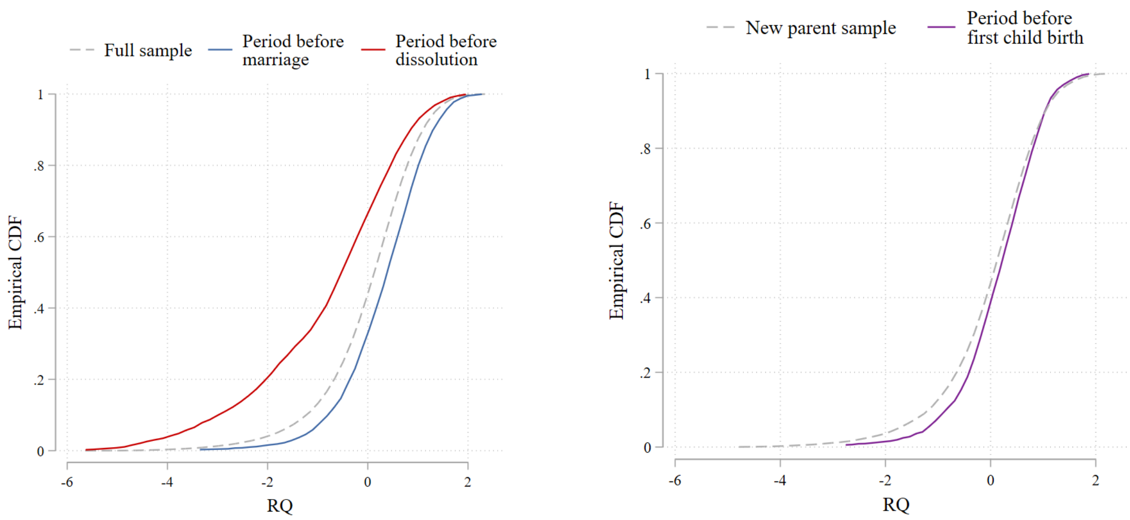
Figure B.1 (b) compares the distribution of the RQ residuals between the new parent sample and the observations one period before having the first child, that is, at the time of conception. This distribution is slightly shifted to the right in comparison to the benchmark. However, the empirical distribution 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 B.1 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 B.1 (b) are sufficiently large. The combined test indicates that all three samples come from different distributions in comparison to

Figure B.1: 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 B.1: 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 B.2 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.[14]

Table B.2: Regression of women's RQ on man's RQ.

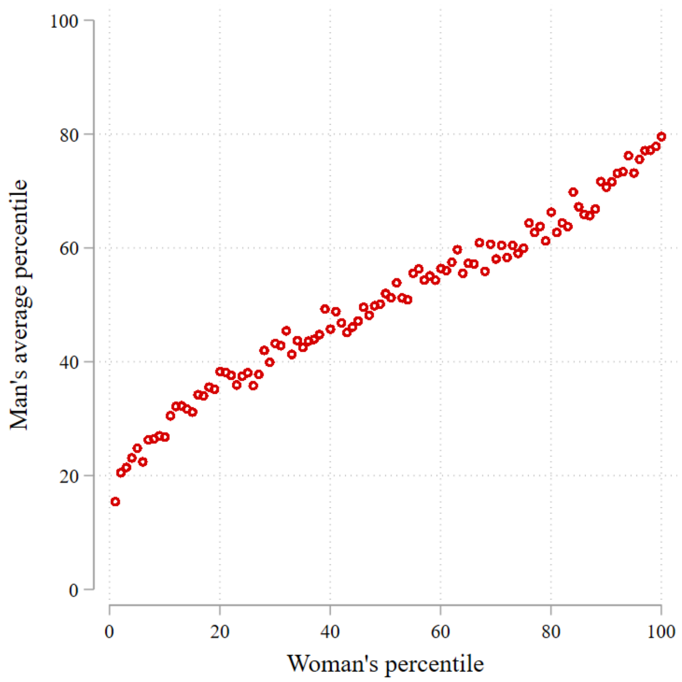
	Woman's RQ
Man's RQ	0.587***
	(0.009)
College Degree	0.049**
	(0.016)
Employed	0.048*
	(0.019)
Log Personal Income	-0.002
	(0.006)
Urban	-0.032
	(0.016)
Married	0.155***
	(0.024)
At least one child	-0.111***
	(0.018)
Constant	0.134
	(0.075)
Age	✓
Tenure	✓
Wave	✓
Observations	25884
R^2	0.3243

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 B.2. Figure B.2 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.

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

Figure B.2: 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.

C 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:

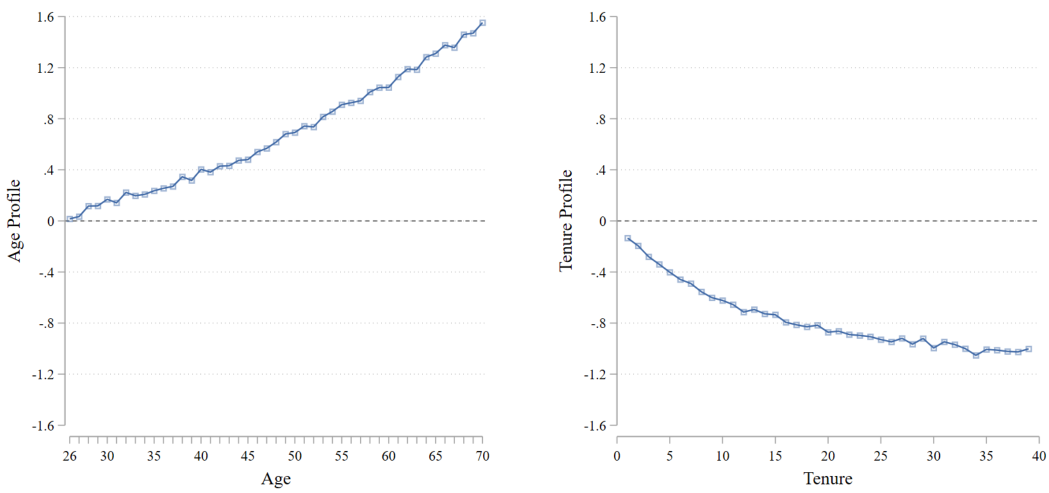
$$y_{it} = \sum_{a} \mathbb{1}\{a = \text{age}_{it}\}\alpha_a + \sum_{d} \mathbb{1}\{d = \text{tenure}_{it}\}\gamma_d + \sum_{w} \mathbb{1}\{w = \text{wave}_t\}\psi_t + \mathbf{X}_{it}\boldsymbol{\beta} + u_{it}$$

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 C.1 (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 C.1 (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 C.1: 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.

D Summary Statistics

Table D.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.

Table D.2: Summary statistics by household income

	Bottom quartile		Second quartile		Third quartile		Top quartile	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women	Men	Women	Men	Women	Men	Women	Men
Age	27.33	30.38	28.25	30.98	28.71	30.65	30.51	32.93
	(5.902)	(7.091)	(5.287)	(6.259)	(4.214)	(4.930)	(4.245)	(5.433)
College	10.31	5.250	25	19.88	46.52	36.80	65.28	63.34
	(30.42)	(22.32)	(43.32)	(39.93)	(49.90)	(48.25)	(47.63)	(48.21)
Weekly paid hours	30.32	33.57	34.26	38.04	35.05	39.06	35.92	38.07
	(14.52)	(15.27)	(9.098)	(10.46)	(8.308)	(7.988)	(8.574)	(9.104)
Weekly housework hours	10.99	5.318	9.888	5.256	8.413	5.166	8.100	4.998
	(7.267)	(4.536)	(5.585)	(4.133)	(4.812)	(3.893)	(4.815)	(3.528)
Gross monthly income	809.8	1008.6	1173.1	1497.5	1622.5	1992.9	2405.0	3141.8
	(490.7)	(531.4)	(549.5)	(699.4)	(841.8)	(1035.7)	(1347.8)	(1737.6)
Gross monthly income (hh)	1902.7	1838.1	2803.4	2779.3	3740.5	3761.2	5791.5	5810.9
	(1024.5)	(984.0)	(1260.1)	(975.2)	(1349.4)	(1301.7)	(2506.9)	(2490.6)
Gender norm index	0.172	0.231	0.450	0.140	0.523	0.404	0.467	0.204
	(0.896)	(0.865)	(0.723)	(0.933)	(0.822)	(0.830)	(0.929)	(0.817)
Observations	853	835	1060	1067	1140	1158	1277	1240

Standard errors in parentheses.

Table D.3: Summary statistics for differently attached couples

	Disattached		Attached	
	(1)	(2)	(3)	(4)
	Women	Men	Women	Men
Age	30.73	33.06	30.62	33.43
	(6.422)	(6.647)	(4.645)	(6.372)
College	57.20	36.02	67.36	56.78
	(49.53)	(48.05)	(46.93)	(49.58)
Weekly paid hours	32.98	36.17	33.73	38.55
	(10.75)	(11.42)	(9.766)	(7.987)
Weekly housework hours	8.278	5.627	8.186	5.715
	(7.009)	(4.045)	(5.921)	(3.679)
Gross monthly income	2000.8	2519.4	2222.9	2920.1
	(1276.0)	(1601.3)	(1317.7)	(1859.7)
Observations	513	512	580	584
Panel B: Couple characteristics				
Tenure	5.201		5.250	
	(4.830)		(3.416)	
Married (%)	56.92		72.07	
	(49.57)		(44.90)	
At least one child (%)	1.170		1.724	
	(10.76)		(13.03)	
Female share of paid hours (%)	48.58		46.84	
	(11.09)		(9.342)	
Female share of housework hours (%)	58.20		57.09	
	(18.93)		(18.06)	
Urban (%)	83.10		73.96	
	(37.51)		(43.92)	
Observations	521		592	

Standard errors in parentheses.

E Identification

E.1 A1. No anticipation effects

E.2 A2. Conditional parallel trends

Figure E.1: Sample mean and standard deviation of RQ around first child birth
What is in the y-axis? Predictive margins or simply averages?

The value of the raw sample moments. I don't know how to interpret the results for the standard deviation, though. It's around two-thirds of the standard deviation of the pooled measure. Is this considered low? Should we worry about extreme values cancelling out when computing the sample mean of RQ?

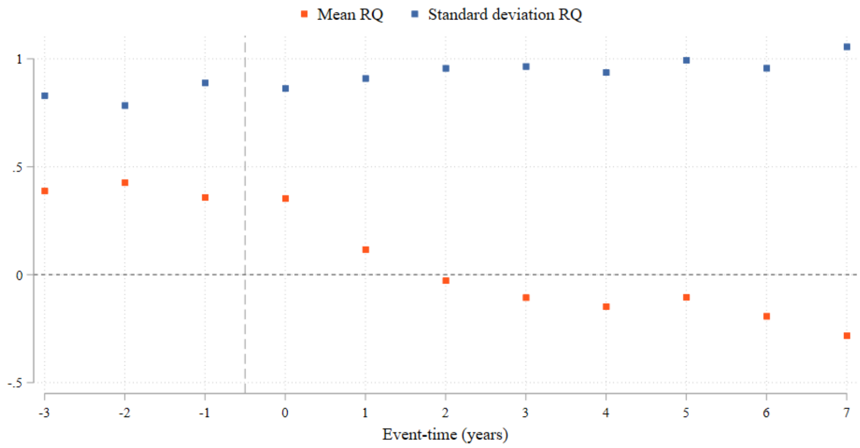
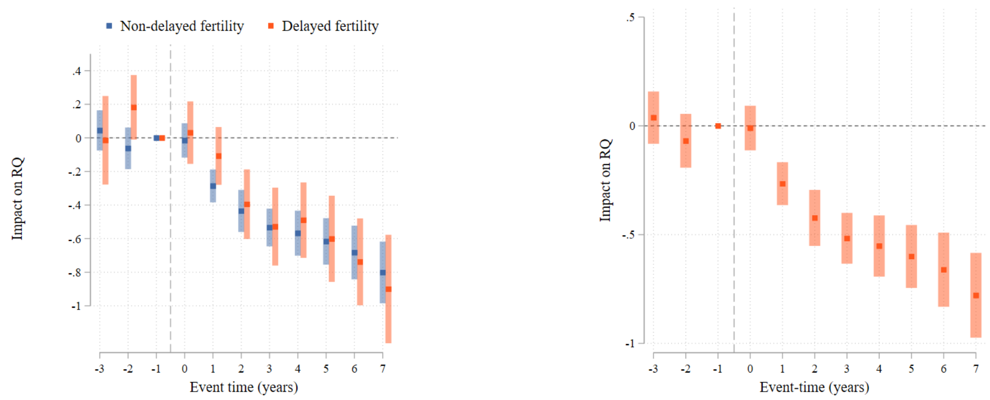


Figure E.2: Proxy delayed fertility vs non-delayed fertility

(a) Marginal effects by proxied delayed fertility

(b) Estimated effects on the subsample of parents reporting non-delayed fertility



F Robustness

F.1 Different estimation strategies

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

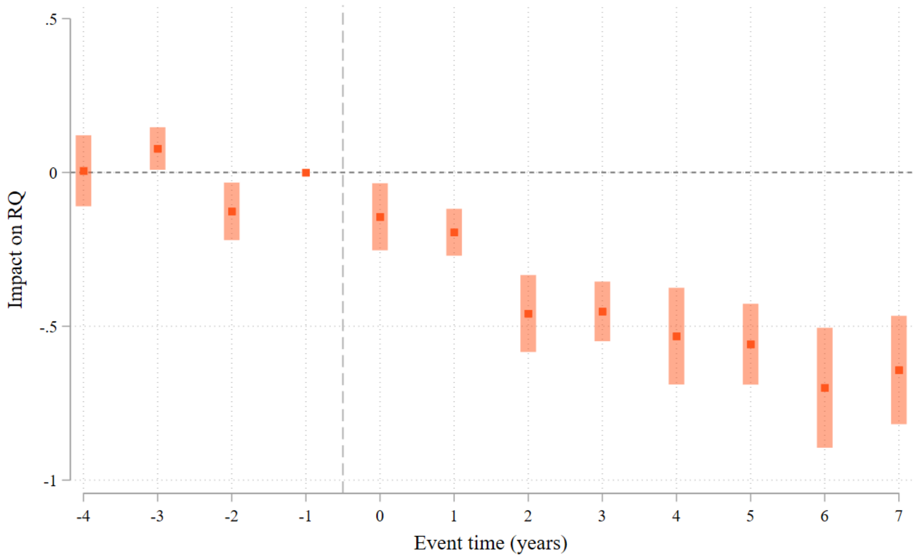
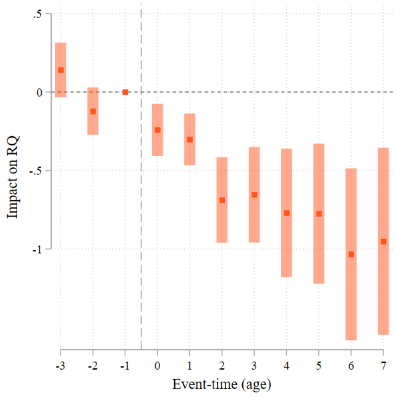
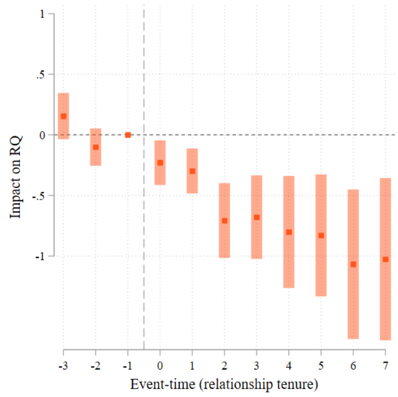


Figure F.2: TWFE specification with age and relationship tenure as time-varying variables

(a) Age



(b) Relationship tenure



F.2 Alternative measures of RQ

Figure F.3: Effects of first child birth on RQ, using Callaway and Sant'Anna (2021)

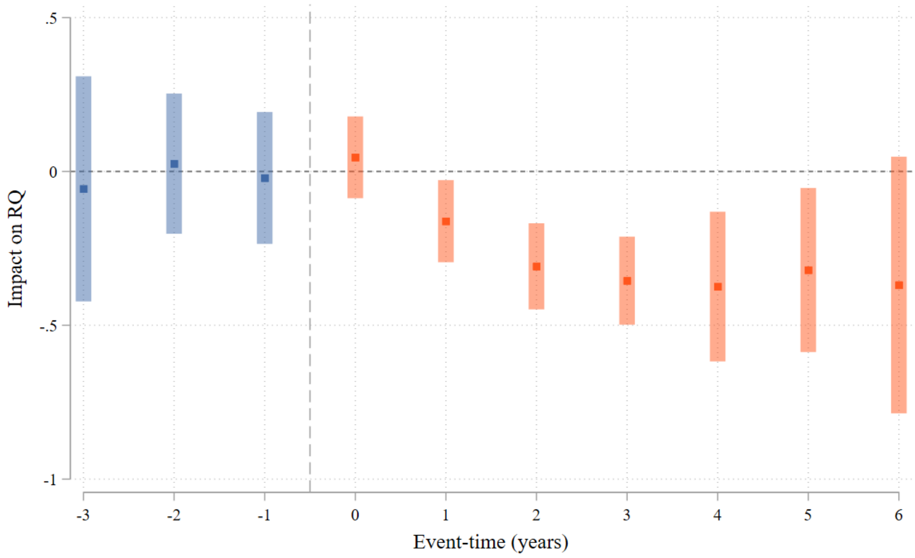
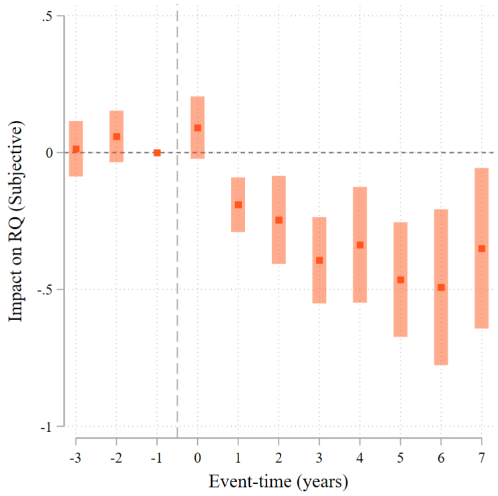
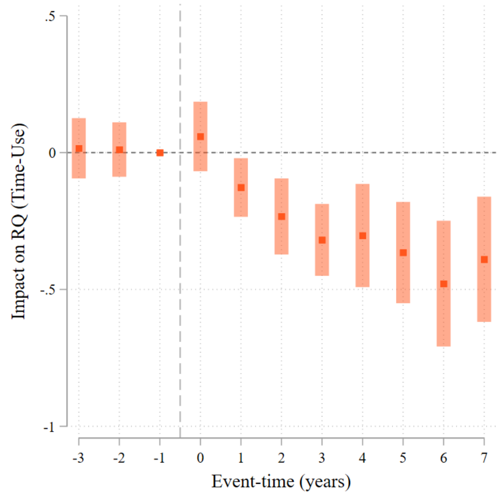


Figure F.4: RQ by item block

(a) Subjective assessment



(b) Couple time use



F.3 Different subsamples

Figure F.5: RQ using factor scores after birth

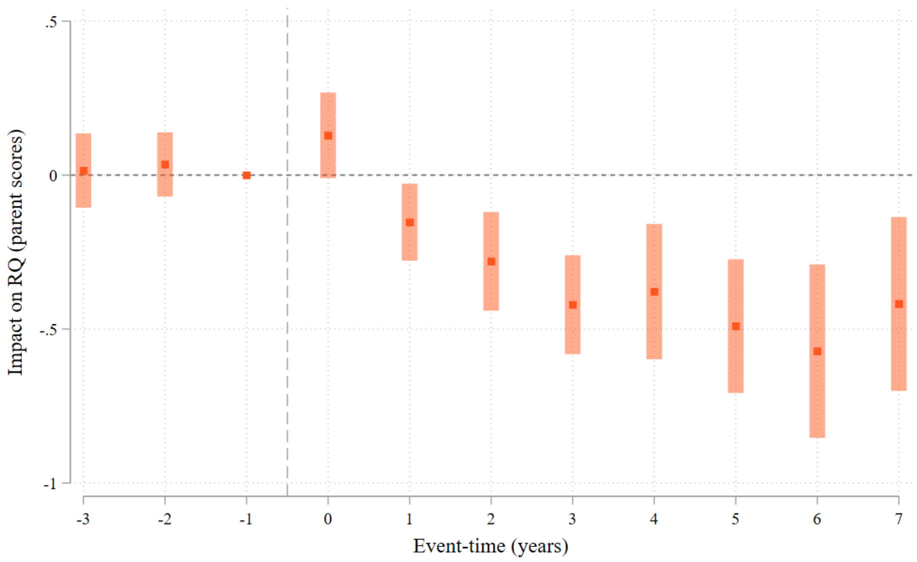
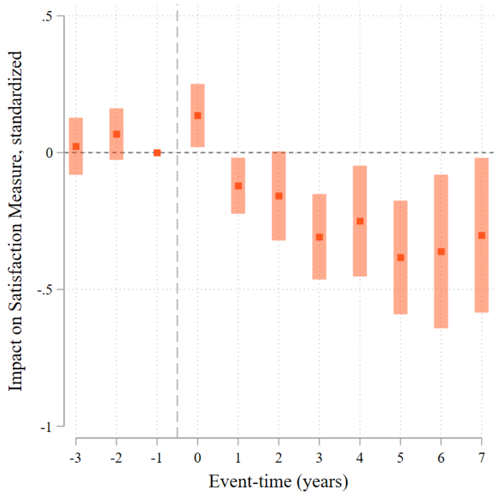


Figure F.6: Psychology measures

(a) RDAS satisfaction



(b) RDAS cohesion

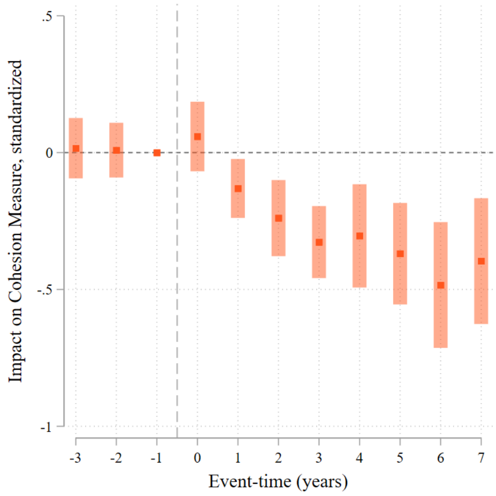
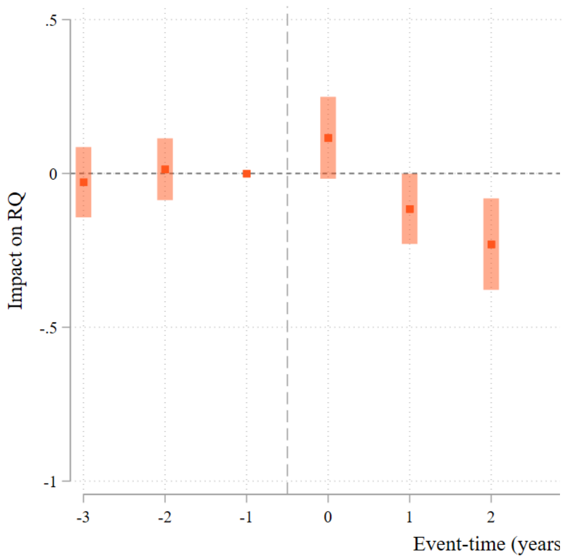


Figure F.7: Effects of first child birth on non-separating couples.

(a) Pooled sample



(b) Marginal effects by gender

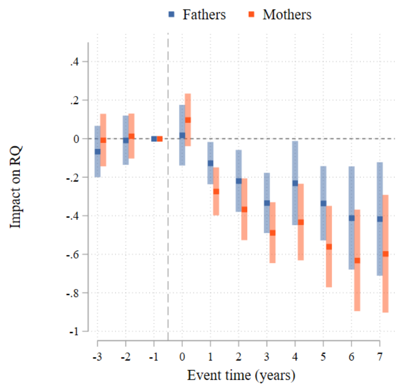
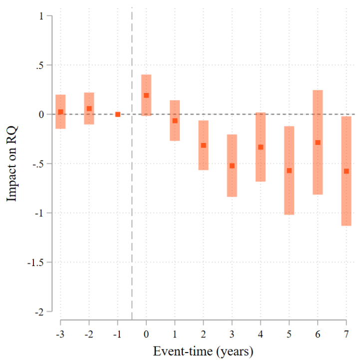
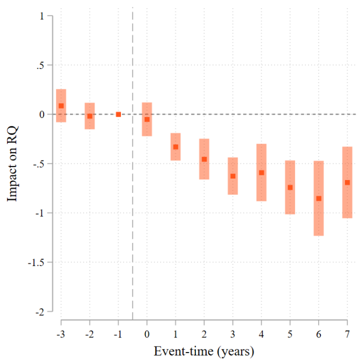


Figure F.8: Impact by final number of children.

(a) One child



(b) Two children



(c) Three + children

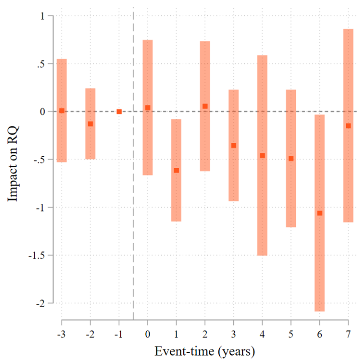


Figure F.9: Effects of first child birth on RQ, boys vs. girls.

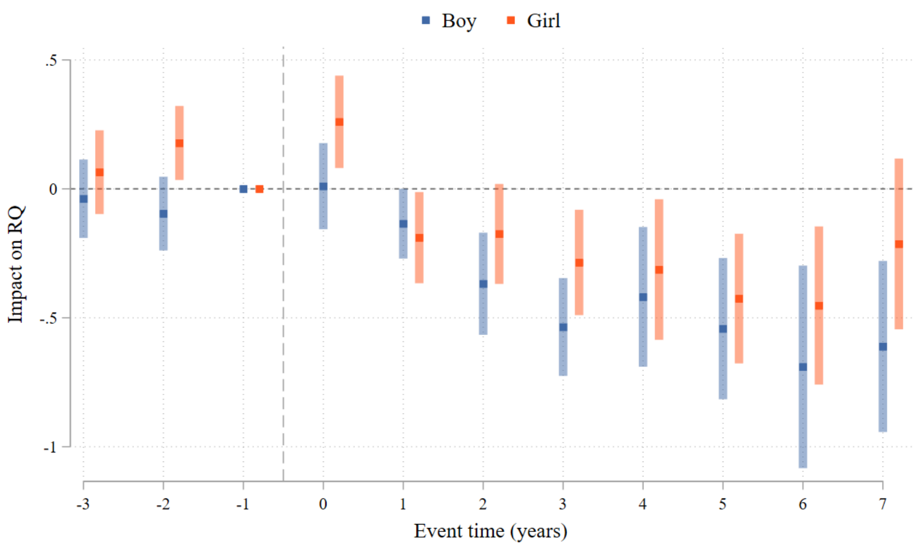
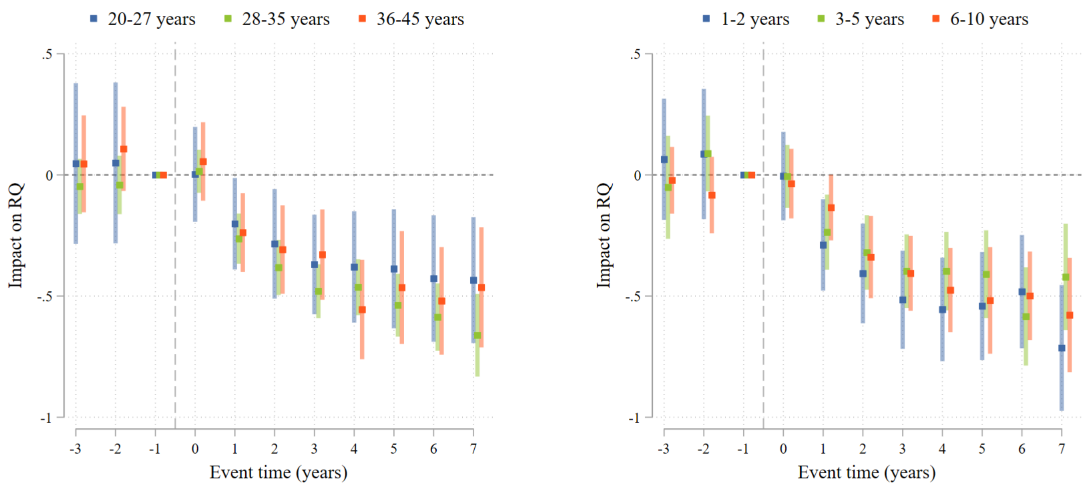


Figure F.10: By age and tenure bin

(a) Age bins

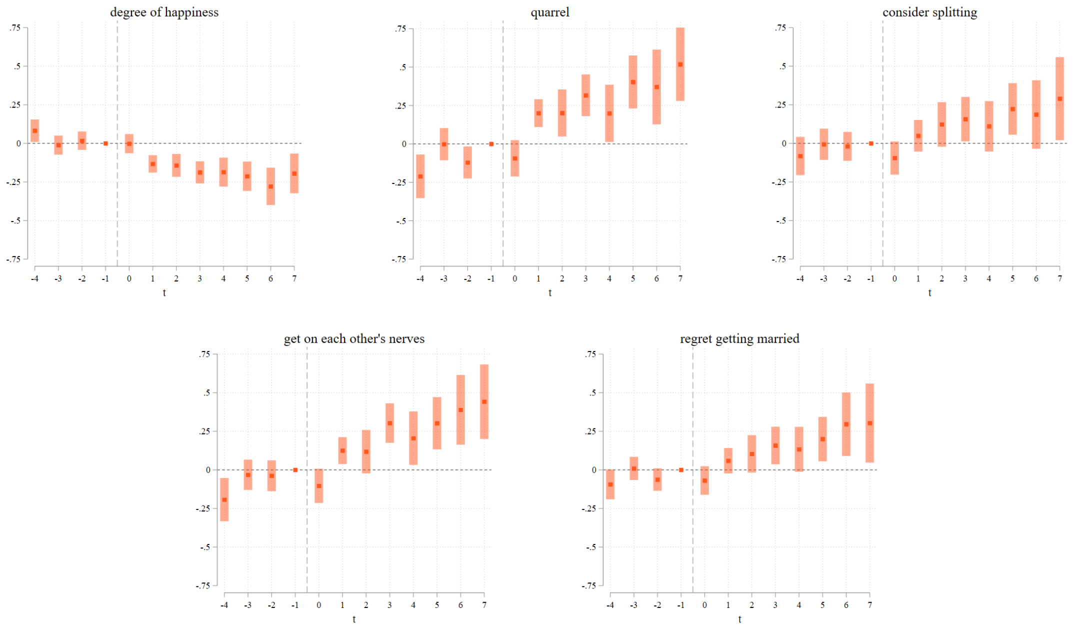
(b) Tenure bin



G Event Study per item

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

(a) Subjective assessment



(b) Couple time use

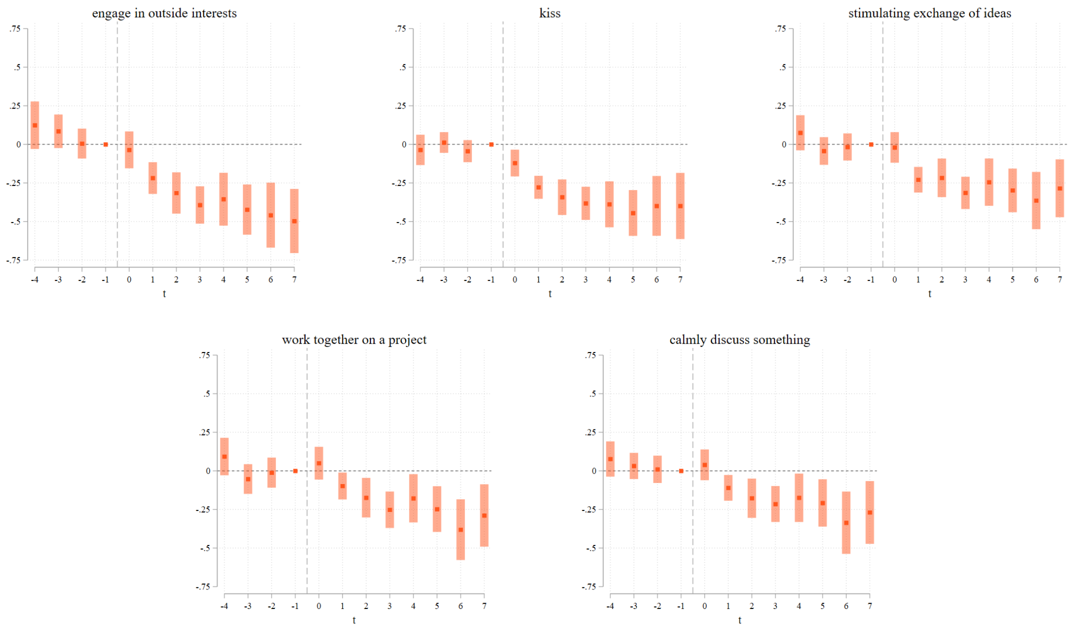
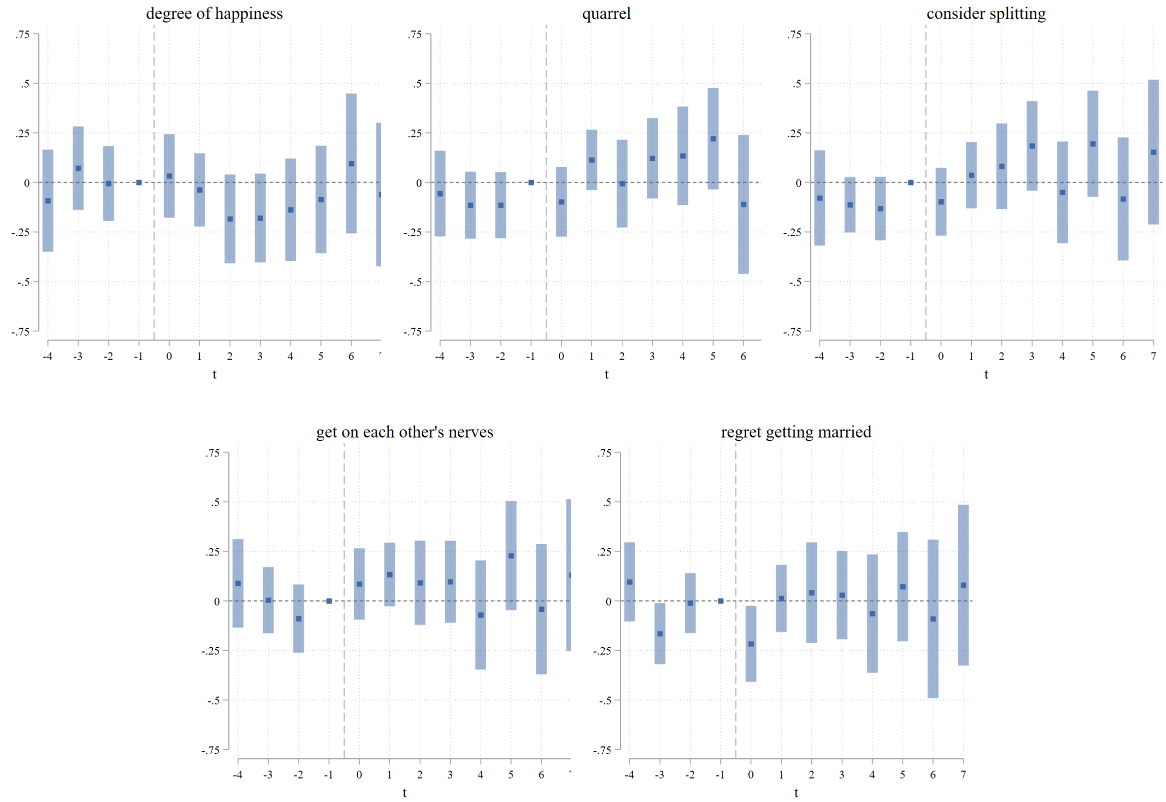
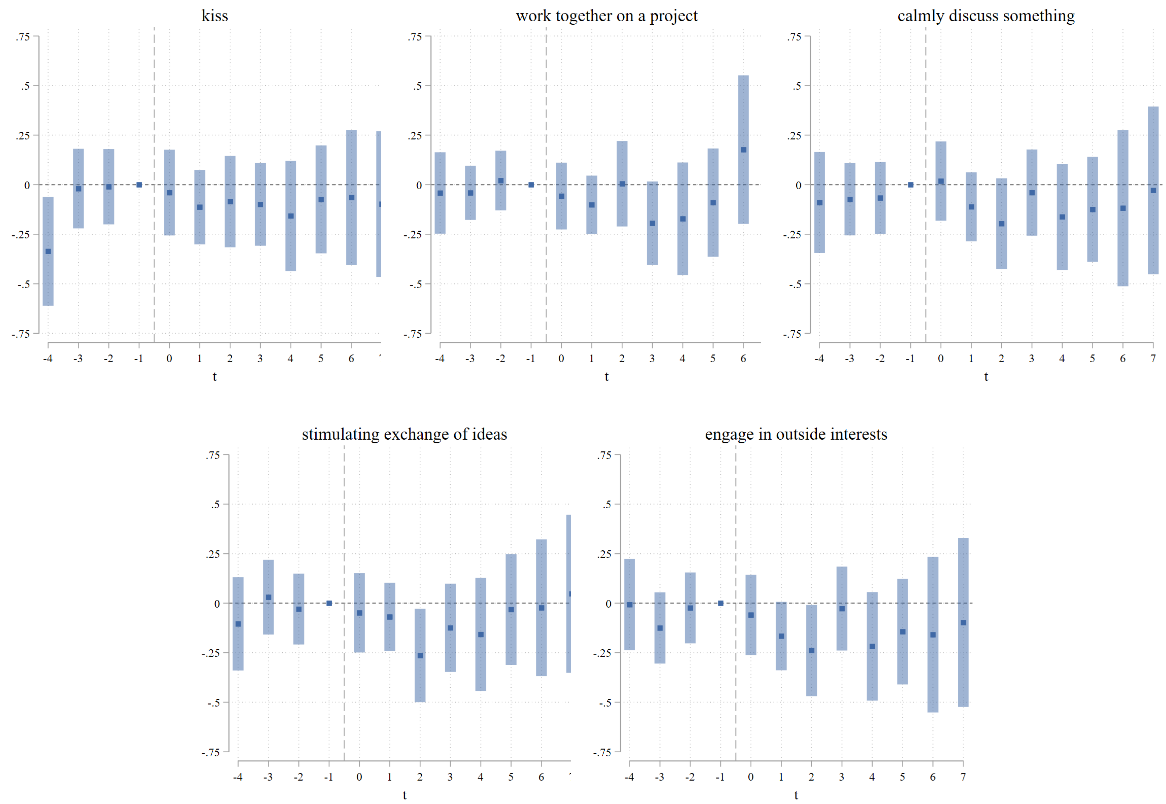


Figure G.2: Differential impact of first child on each item for mothers and fathers.

(a) Subjective assessment



(b) Couple time use



H Identification of the causal effects

H.1 Identifying assumptions of the dynamic ATTs

Our specification requires three main assumptions for the causal interpretation of the ATT parameter. Namely, we need (i) no anticipation effects; (ii) parallel trends; and (iii) homogeneous treatment effects across units and periods. We discuss the plausibility of each of these assumptions in the context of our study.

Using a sample of new parents, we study the impact of first child birth (treatment variable) on relationship quality (outcome variable). Our treatment is staggered because individuals have their first child birth in different observational periods. This divides our sample into different cohorts of new parents (treated units) depending on the calendar period they become parents.[15] The main issue regarding the estimation of the ATT parameter is that it relies on clean comparisons between new parents at period t and not-yet-parents (*late-treated* units), and on "forbidden" comparisons between new parents and already-parents (*early-treated* units). These latter comparisons demand the assumption of homogeneous treatment effects across periods and units when estimating the causal effect in the long run.

We illustrate this fact by means of an example based on our main specification.[16] Consider two units, A and B, of the same age, a, and relationship tenure, d, that are observed in period s. Units A are *early-treated* and receive the treatment in period $s = 2$, whereas units B are *late-treated* and receive the treatment in period $s = 3$. We denote the causal effect of the treatment for each unit and period as δ_{Zs}, where $Z \in \{A, B\}$. We present the expected outcomes for each unit type at each period s in Table H.1.

Table H.1: Example 1. Estimation of the ATT for treated units at event time periods.

$E[Y_{ist}]$	i = A	i = B
s = 1	$\alpha_a + \gamma_d$	$\alpha_a + \gamma_d$
s = 2	$\alpha_{a+1} + \gamma_{d+1} + \psi_2 + \delta_{A2}$	$\alpha_{a+1} + \gamma_{d+1} + \psi_2$
s = 3	$\alpha_{a+2} + \gamma_{d+2} + \psi_3 + \delta_{A3}$	$\alpha_{a+2} + \gamma_{d+2} + \psi_3 + \delta_{B3}$
Event date	s= 2	s = 3

The standard DiD estimator makes the following comparisons when estimating the

[15]In our sample everyone is treated at some point. This means that we will be comparing always-treated versus not-yet-treated units. Given our setting, we believe not-yet-treated units are a better comparison group than never-treated units. The reason is that we do not observe individuals that wanted to have a child at time t but did not do it at that time or did not do it at all. "Unwanted" delay/lack of fertility may be a threat to the parallel trend assumption to the extent to which it affects the RQ of couples. The exposure to this threat is smaller mechanically for not-yet-treated units since they end up having a child.

[16]Borusyak et al. (2022) derive a similar table to clarify the required assumptions of the DiD estimator in the standard TWFE setting. We follow their approach and adapt the example to our specification.

short-run effect for the *early-treated* units (A):

$$
\begin{aligned}
\delta_{A2} &= (Y_{A2} - Y_{A1}) - (Y_{B2} - Y_{B1}) = \\
&= \big((\alpha_{a+1} + \gamma_{d+1} + \psi_2 + \delta_{A2}) - (\alpha_a + \gamma_d)\big) - \big((\alpha_{a+1} + \gamma_{d+1} + \psi_2) - (\alpha_a + \gamma_d)\big) = \\
&= \big((\alpha_{a+1} - \alpha_a) + (\gamma_{d+1} - \gamma_d) + \psi_2 + \delta_{A2}\big) - \big((\alpha_{a+1} - \alpha_a) + (\gamma_{d+1} - \gamma_d) + \psi_2\big) = \\
&= \delta_{A2}
\end{aligned}
$$

The identification of the causal effect, δ_{A2}, relies on the assumptions of (i) no anticipation effects and (ii) conditional parallel trends. In our setting, the parallel trends assumption implies that, in the case of delayed fertility, the RQ of individuals would have evolved according to the functional form $y = \alpha_a + \gamma_d + \psi_s + u$. The presence of age and tenure effects ensures that the estimates of δ_{A2} do not capture additional time-varying effects that would otherwise result from the comparison of units of different ages and tenures:

$$
\delta_{A2} = \delta_{A2} + \big((\alpha_{a+1} - \alpha_a)) - (\alpha_{a'+1} - \alpha_{a'})\big) + \big((\gamma_{d+1} - \gamma_d)) - (\gamma_{d'+1} - \gamma_{d'})\big)
$$

$\forall a, a' : a \neq a'$ and $\forall d, d' : d \neq d'$.[17] Without the inclusion of age and tenure effects, biased estimates of the ATT would arise unless age and tenure effects have a linear relationship.[18]

As mentioned before, the assumption of homogeneous treatment effects across periods is key for identifying the causal effect in the long run. Using the previous example in Table H.1, we describe the comparisons carried by DiD estimator when estimating the impact of first child birth after two periods:

$$
\begin{aligned}
\delta_{A3} &= \big[(Y_{A3} - Y_{A1}) - (Y_{B3} - Y_{B1})\big] - \big[(Y_{A2} - Y_{A1}) - (Y_{B2} - Y_{B1})\big] = \\
&= \big[\big((\alpha_{a+2} + \gamma_{d+2} + \psi_3 + \delta_{A3}) - (\alpha_a + \gamma_d)\big) - \\
&\quad - \big((\alpha_{a+2} + \gamma_{d+2} + \psi_3 + \delta_{B3}) - ((\alpha_a + \gamma_d)\big)\big] - \\
&\quad - \big[\big((\alpha_{a+1} + \gamma_{d+1} + \psi_2 + \delta_{A2}) - (\alpha_a + \gamma_d)\big) - \big((\alpha_{a+1} + \gamma_{d+1} + \psi_2) - (\alpha_a + \gamma_d)\big)\big] = \\
&= \delta_{A3} - \delta_{B3} - \delta_{A2}
\end{aligned}
$$

The DiD estimator of δ_{A3} is unbiased only if the short-run treatment effect is homogeneous for units A and B, and across periods 2 and 3, $\delta_{B3} = \delta_{A2}$. This assumption is

[17] In the presence of covariates, the resulting δ_{A2} is a weighted unbiased estimator of the treatment effect for each combination of age and relationship tenure bins. These weights are determined by the proportion of units A that receive the treatment in each bin combination.

[18] Age and tenure linear effects imply that each additional year of life or relationship has the same impact on RQ. We estimate RQ profile along age and tenure in Appendix C, and shows that it is not the case.

necessary because the DiD estimator involves "forbidden" comparisons between treated units in period 3 and earlier treated units in period 2. As the time horizon increases and more *early-treated* units are used as the control group, this assumption becomes more restrictive. Additionally, our specification requires this assumption to hold for *every* combination of age and relationship tenure.

To further illustrate the implications of this assumption, we provide a second example. Consider an individual A who receives the treatment in period $s = 1$, at the age of twenty and after one year of relationship tenure. We denote the short-run effect of first child birth for unit A as δ_{A2}. Similarly, suppose there is another individual B, of the same age and relationship tenure as A by period 1, who receives the treatment in period 2. We refer to the short-run effect of first child birth for unit B as δ_{B3}. If we want to estimate the causal effect of first child birth two periods after for individual A (δ_{A3}), we need to assume that the short-run effects of first child birth are the same for units A and B ($\delta_{A2} = \delta_{B3}$). That is, we need to assume that the short-run impact of first child birth for an individual aged twenty and in the first year of her relationship is the same as for an individual aged twenty-one in the second year of her relationship.

This assumption becomes less plausible when estimating longer-run effects. Suppose we target the estimation of δ_{A12}, which represents the causal effect for individual A ten years after the first child birth. The DiD estimator extrapolates the type of forbidden comparisons and requires that $\delta_{A2} = \delta_{Z11}$, where Z is an individual of the same age and tenure as individual A by period 1 (when A is treated), that had her first child in period 11, at the age of 30 and 11 years of relationship tenure.

Panels (a) and (b) in **??** plots the marginal effects of first child birth for individual that become parents at a young age (< 25) and an old age (> 35). Panels (c) and (d) compare old-aged parents with average-age parents (25-35 years old). We distinguish by the total number of children to uncover heterogeneity that could arise from younger parents being more likely to have more children.

The impact of first child birth is the same for old parents, regardless of whether we condition on total fertility, which suggests that these parents usually have only one child. The RQ paths of young and old parents follow a similar pattern until the age of four, when the RQ of young parents starts declining again. The drop in RQ for young parents differs by total fertility. Young parents that only have one child suffer a sharper drop, indicating a certain correlation between the lack of more children and RQ. On the contrary, young parents who may have more than one child experience a smoother decay, consistent with the observed patterns for the pooled sample that have more than one child.[19]

The RQ paths for average-age and old parents are almost similar in the case of unconditional total fertility. Conditional to having only one child, average-aged parents

[19]The drop in RQ coincides with the average age of the first child at which couples usually have a second child.

experience a similar fall in RQ when the first child is seven years old. Once again, this is indicative of a negative correlation between RQ and the lack of more children among those couples that only had one child.

H.2 Main specification vs standard TWFE regression

In this section, we compare the differences estimators for the ATT parameters produced by the OLS estimation of the main specification and a standard dynamic TWFE regression. We present both algebraic and empirical explanations.

We start by specifying our estimation target:

$$ATT_t = \sum_j w_j ATT(j,t) \tag{3}$$

where j refers to treated units and is an indexation that varies across specifications, and w_j are the weights given to each treated unit. ATT(j,t) is the ATT for units j and event time t.

We consider two different equations to estimate ATT_t:

$$y_{ist} = \sum_{t=-5}^{-2} \delta_t^{leads} D_{is}^t + \sum_{t=0}^{10} \delta_t^{lags} D_{is}^t + \alpha_a + \gamma_d + \psi_s + \mathbf{X}_{is}\boldsymbol{\beta}^{(1)} + v_{ist} \tag{4}$$

$$y_{ist} = \tau_t + c_i + \beta_s + \mathbf{X}_{is}\boldsymbol{\beta}^{(2)} + u_{ist} \tag{5}$$

where δ_t are ATT_t parameters in the dynamic specification. $\forall g < 0, \delta_g$ captures the probability that unit i will be treated by the period s-t-g, where t is the treatment period for that unit.

Parameters α_a, γ_d are age and tenure fixed effects, and ψ_s are period effects. In the TWFE regression, the τ_j's parameters capture the dynamic effect. c_i, β_t are individual and period fixed effects.

Both specifications define treated units depending on the calendar period at which they had their first child. The main specification is, nevertheless, conditioning also on age and relationship tenure effects.

Using a similar example to the one in Table H.1 for the case of the TWFE regression, we can show that both specifications require the same set of assumptions to identify the $ATT(j,t)$ parameters. The only difference is that our main specification relaxes the parallel trends assumption and homogeneity assumption to a further set of covariates (age relationship tenure bins) than the TWFE regression.

Hence, the main difference between the two equations boils down to the weights that each of them assigns to treated units.

Different weighting schemes: We make use of Proposition 2 in Borusyak et al. (2022), and express the weights assigned by each specification as:

$$w_{is}^{(1)}(t) = \frac{\sum_i \sum_s \tilde{D}_{ist}^{(1)} E[y_{ist}^{(1)}]}{\sum_i \sum_s \tilde{D}_{ist}^{(1)2}} \tag{6}$$

$$w_{is}^{(2)}(t) = \frac{\sum_i \sum_s \tilde{D}_{ist}^{(2)} E[y_{ist}^{(2)}]}{\sum_i \sum_s \tilde{D}_{ist}^{(2)2}} \tag{7}$$

where t denotes the event time t, and (1) and (2) denote the main specification and the TWFE one. The $\tilde{D}_{ist}$ are the residuals of the following regressions:

For the shake of simplicity, we focus on the event time period 0. That is, each weight above simply denotes the predicted probability of belonging the treated unit t.

$$\tilde{D}_{is}^{(1)} = D_{is} - \hat{D_{is}}^{(1)} = D_{is} - \bar{D}_a - \bar{D}_d - \bar{D}_s - \bar{D}$$

$$\tilde{D}_{is}^{(2)} = D_{is} - \hat{D_{is}}^{(2)} = D_{is} - \bar{D}_s - \bar{D}_i - \bar{D}$$

D_{is} are treatment status indicators for each individual and period. $\forall a \in \{20, \bar{a}\}$ and $\forall d \in \{1, \bar{d}\}$, $\bar{D}_a$ and $\bar{D}_d$ are the probabilities of being treated at age a and relationship tenure d, s.t.:

$$\bar{D}_a = \frac{\sum_i \sum_s D_{is} \mathbb{1}\{age_{is} = a\}}{\sum_i \sum_s \mathbb{1}\{age_{is} = a\}}$$

$$\bar{D}_d = \frac{\sum_i \sum_s D_{is} \mathbb{1}\{tenure_{is} = d\}}{\sum_i \sum_s \mathbb{1}\{tenure_{is} = d\}}$$

$\bar{D}_i$ is the probability of being treated for individual i, and and $\bar{D}_s$ is the probability of being treated at period s, s.t.:

$$\bar{D}_i = \frac{\sum_s D_{is} \mathbb{1}\{individual_{is} = i\}}{\sum_s \mathbb{1}\{individual_{is} = i\}}$$

$$\bar{D}_s = \frac{\sum_i D_{is} \mathbb{1}\{period_{is} = s\}}{\sum_i \mathbb{1}\{period_{is} = s\}}$$

The main specification exploits existing within age, relationship tenure and period variation in the treatment status. The TWFE exploits within individual and period variation instead.

Let's illustrate what these differences mean by an example again. Consider individual i of age a and relationship tenure d at period s.

The main specification computes the probability that this individual is treated at s by adding up the sample probabilities of being treated at age a, being treated at relationship tenure d, being treated at period s, and being treated at all.[20] Since these probabilities are approximated nonparametrically, they are just the sample averages of the total number of treated individuals of each age, tenure, and at each period.

The TWFE regression exploits a different observed variation in the data. The predicted probability that individual i is treated at period s adds up the probability that individual i is treated at all, and the probability that individuals are treated at period s, and the average probability of being treated.

[20]The probability of being treated at all is one in our case since every one since there are not never-treated individuals in our sample.

I Heterogeneity by couple time attachment

Table I.1: Regression analysis by couple type

	Weekly paid hours		Weekly housework hours	
	(1)	(2)	(3)	(4)
	Women	Men	Women	Men)
Disattached	-8.527***	-0.390	2.821***	0.0591
	(1.030)	(0.902)	(0.787)	(0.399)
Baseline	32.98	36.17	8.27	5.62
Attached	-9.585***	-1.390	3.179***	0.264
	(1.023)	(0.808)	(0.674)	(0.362)
Baseline	33.73	38.55	8.187	5.71
Observations	2676	2423	1471	1435

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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

Figure J.1: Predictive margins of the impact of fertility on mothers' supply of paid and housework time for rich and poor households

(a) Paid hours

figures/income/qincome_heter_jbrhs.png

(b) Housework hours

figures/income/qincome_heter_howlng.png

J Heterogeneity by income quartile