Discussion Paper Series – CRC TR 224

Discussion Paper No. 719
Project A 03

Children, Household Specialization and Relationship Quality

Belén Rodríguez Moro[1]
Olatz Román[2]

December 2025

[1] Universitat Pompeu Fabra,Email: belen.rodriguez@eui.eu.
[2] University of Mannheim, Email: olatz.roman@uni-mannheim.de.

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Children, Household Specialization and Relationship Quality*

Belén Rodríguez Moro† Olatz Román‡

November 11, 2025

Abstract

We investigate how having children impacts the quality of couples' relationships, a proxy of the non-material gains from being in a relationship. Using a novel measure of relationship quality (RQ), we perform a dynamic difference-in-differences estimation around the birth of the first child. We find a sharp and lasting decrease in RQ immediately after birth. We attribute this effect to changes in household specialization. Traditional gender-based specialization prevails after birth, regardless of the baseline distribution of tasks within the couple. Leveraging heterogeneous changes in household specialization after birth, we find that couples undergoing larger rearrangements also suffer larger RQ drops.

Keywords: Fertility, Marital decisions, Time allocation, Household specialization

JEL Codes: J12, J13, J22, D13

*We are grateful to Sule Alan, Thomas F. Crossley, David K. Levine, Libertad González and Alexander Monge-Naranjo for their invaluable guidance and support. We would also like to thank Abi Adams, Monica Costa-Dias, Andrea Ichino, Jonas Jessen, Peter Levell, Yannick Reichlin, Almudena Sevilla, Alessandro Tarozzi, 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, the Understanding Society Scientific Conference 2023, the 2023 EEA-ESEM congress, the 2023 EALE conference, the 48th SAEe symposium, IZA Institute for Labor Economics, SOLE 2024, University of the Basque Country, the XVI AEET Labor Economics Conference, the 17th UniTO-Collegio Carlo Alberto Ph.D. Workshop in Economics, the ifo Institute, Masaryk University and the 18th Vigo Workshop on Dynamic Macroeconomics for helpful comments and suggestions. Support by the German Research Foundation (DFG) through CRC TR 224 (Project A03) is gratefully acknowledged.

†Universitat Pompeu Fabra, belen.rodriguez@eui.eu.

‡University of Mannheim, olatz.roman@uni-mannheim.de.

1 Introduction

Having children changes peoples' lives. With more responsibilities at home, including housework and childcare, couples need to reorganize how they allocate their time. Mothers experience a significant reduction in the time spent in the labour market (Goldin, 2021), home production is readjusted accordingly (Aguilar-Gomez et al., 2019; Siminski and Yetsenga, 2022), and leisure and resting become second order for both parents (Aguiar and Hurst, 2007; Costa-Font and Flèche, 2020). Despite these significant adjustments that couples undergo when having children, the deeper consequences that parenthood and these accompanying changes may have on couples have received limited attention.

This paper studies the impact of having children on the quality of couples' relationships. Relationship quality is a measure of the non-material benefits that individuals derive from their relationships, and is widely used in the theoretical literature of family economics as a component of utility that influences marital decisions. While models studying family formation and dissolution widely consider the notion of relationship or match quality, its underlying determinants are little understood (Browning et al., 2014; Chiappori, 2020). To our knowledge, only a few papers have attempted to empirically measure relationship quality to guide such models (Weiss and Willis, 1997; Chiappori et al., 2018).

The importance of relationship quality in shaping family behaviour has grown over time. As women have become more financially independent and their outside options have improved, non-material factors have gained importance within the value of being in a couple (Lundberg and Pollak, 2007), making relationship quality more decisive for couple formation and dissolution (e.g., Browning et al., 2014). These shifts also have consequences on fertility, a key concern for policymakers: greater female bargaining power means that the decision to have children increasingly reflects both partners' expectations about the division of childcare and housework (Doepke and Kindermann, 2019), and the realized experience of a first birth updates these expectations in ways that shape subsequent fertility decisions (d'Albis et al., 2017). Understanding how the transition to parenthood affects relationship quality is therefore essential to explaining patterns of couple formation, dissolution and fertility.

Beyond shaping marital and fertility decisions, relationship quality is itself a component of utility (Chiappori et al., 2018), implying that the welfare effects of parenthood cannot be assessed through time and income alone. It also governs intra-household cooperation and bargaining, influencing how parents allocate time and resources to children (Chiappori and Weiss, 2007). Taken together, these considerations make relationship quality central to understanding family behaviour and point to the transition to parenthood as a critical margin.

To investigate the effect of having children on parents' relationships, we construct a novel measure of relationship quality (RQ). We use a questionnaire from Understanding Society, a household longitudinal panel in the United Kingdom, which collects information about rela-

tionships with cohabiting partners. The questionnaire items provide insights into subjective assessments about the relationship, such as the degree of happiness with the couple or the frequency with which they consider splitting, and couple time use, like how often they work together on a project or kiss. This rich set of information allows us to gain a nuanced understanding of the non-material benefits from being in the current relationship. To capture the multiple dimensions of couples' relationship quality in a parsimonious way, we combine the responses to the questionnaire in a factor analysis and construct the RQ measure. We conduct several validation exercises that ensure the capacity of this measure to predict marital outcomes such as couple dissolution and marriage decisions. Our companion paper, Rodríguez-Moro and Román (2024), provides more details on the construction, validation and analysis of the measure.

We leverage variation in the timing of the birth of the first child to estimate the dynamic impact of having children on the quality of parents' relationships. Using a dynamic difference-in-differences approach, we remove individual unobserved heterogeneity and focus on the evolution of RQ with time relative to birth. We estimate this specification following the method proposed by Callaway and Sant'Anna (2021).

We find a sharp and steady decrease in RQ in the four years after the birth of the first child. This decline stabilizes at approximately half a standard deviation below pre-birth levels. Illustratively, the RQ of individuals who ranked in the 75th percentile of RQ before having their first child is reduced to median RQ within the first four years after birth, and such a decrease in RQ is associated with a 17.5% higher probability of dissolution. This impact lasts for the entire seven-year period examined. In contrast to most of the child penalty literature, this impact is largely symmetric for mothers and fathers, challenging the notion that fathers are unaffected by parenthood.

The magnitude and persistence of the estimated effect are sizeable when considering the large and positive association between RQ and couple dissolution, and in comparison to the effect that fertility has on general happiness. Notably, the effect on general happiness is positive, although about half the size of the impact on RQ, and it is short-lived, lasting only one period. We consider RQ a component of happiness exclusively attached to the couple and distinct from the happiness derived from having a child (Chiappori et al., 2018). Based on that, we interpret the impact on happiness as the sum between a decrease in happiness from one's partner and an increase in happiness from the child. Moreover, the impact of childbirth on RQ is larger than other major life events, such as becoming unemployed, which has virtually no impact on RQ.

The causal interpretation of these results is contingent on two assumptions: limited anticipation and conditional parallel trends. We provide evidence supporting the plausibility of these assumptions. Crucially, we observe flat trends in RQ in the periods leading up to birth, which is not masking any significant heterogeneity in pre-trends. Moreover, delays in fertility, such as miscarriages or the need for fertility treatments, do not lead to differences in the impact of

childbirth on RQ.

We establish the robustness of our findings addressing five potential concerns. First, we verify that our results are robust to using alternative counterfactuals, using the group of infertile individuals and never parents as control groups. Second, we show that our results are not solely driven by a reduction in time spent together as a couple by conducting separate analyses on the different item blocks, all of which are impacted by the arrival of the first child. Third, we rule out the possibility that subsequent children drive the lasting impact on RQ repeating the analysis separately for couples with different total number of children. Fourth, we address concerns about potential selection bias associated with the unavailability of RQ data after a couple dissolves, showing that attrition from the sample due to separation is minimal and that the results do not change when repeating the analysis solely on couples who do not dissolve. Lastly, we verify that the timing of birth, in terms of parental age and relationship tenure, does not influence our results.

In light of these findings, we delve into the mechanisms at play. We start by documenting that the first birth triggers significant and unprecedented changes in how couples allocate their time, and substantially increases the demand for routine housework. This increase is borne almost entirely by women: excluding childcare, they devote about five additional hours per week to housework on average, offset by a reduction in their labour market time. At the same time, men's housework increases by about one hour per week on average, while their labour market hours remain unchanged.

To study how household roles and specialization change, we compute women's share of total paid and unpaid work hours within the household and classify couples into groups based on their pre-birth division of responsibilities. We establish that, regardless of the pre-birth division of paid and unpaid work, all types of couples adopt gender-based household specialization on average after birth. While prior studies have documented an average move toward specialization after first birth, we are the first to show that this shift prevails across various degrees of pre-birth specialization. This pattern is not consistent with household specialization driven purely by comparative advantages within the couple and points instead to the existence of labour market frictions or identity considerations after birth (Akerlof and Kranton, 2000; Ichino et al., 2019).

Leveraging the documented heterogeneity in the reallocation of labour market and home production after the first birth, we study how child-induced changes in household specialization mediate the effects on RQ. In doing so, we conduct separate analyses for different couple types. We find that couples experiencing the largest changes in the share of housework done by women tend to experience more pronounced declines in RQ.

Our findings support the notion that household specialization driven by the arrival of children may have adverse consequences for couples' relationship quality. This has implications for the design of policies aimed at promoting fertility. Policies that aim to support more equi-

table divisions of responsibilities within households may mitigate these negative effects. Such policies have been linked to improved female labour market outcomes (Olivetti and Petrongolo, 2017) and increases in men's contribution to housework (Farré and González, 2019). At the same time, Avdic and Karimi (2018) find that these policies may bring couple separations forward, possibly by accelerating the revelation of the fully unravelled decrease in RQ. Therefore, implementing such policies could yield both immediate benefits, by improving parents' RQ, and long-term advantages, by perpetuating more equitable household arrangements that persist across generations.

Related literature. This paper connects to the extensive literature on the economics of family formation and dissolution. Standard models studying family formation and dissolution commonly integrate the concept of relationship or match quality into the decision-making process (Weiss and Willis, 1997; Browning et al., 2014; Chiappori and Mazzocco, 2017; Chiappori, 2020). However, the underlying determinants of this variable are little understood.[1] Empirical efforts to guide such models are limited, proposing general characteristics of partners or subjective well-being as proxies for match quality (Weiss and Willis, 1997; Bertrand et al., 2015; Chiappori et al., 2018). We develop a novel measure that explicitly integrates information about relationships, which has the potential to yield a more precise and reliable assessment in this context, shedding light on the process followed by this variable, in terms of its evolution over the life cycle, and on how it reacts to life events.

Previous studies in psychology have proposed and tested various measures of marital satisfaction and conflict (e.g., Norton, 1983; Busby et al., 1995; Joel et al., 2020). These variables have previously been studied across social sciences in relation to different outcomes. For instance, Carlson and VanOrman (2017) study the relationship quality trajectories of married and unmarried parents, only after birth. Our measure and analysis present a number of advantages over prior research. First, RQ integrates multiple aspects of relationships, for which only separate measures have been available, into a single, parsimonious measure of match quality (Busby et al., 1995). Second, the combination of a larger sample and longitudinal data allows us to employ causal identification methods that were previously unfeasible, overcoming limitations acknowledged in earlier studies (Amato and Booth, 2001; Hassebrauck and Fehr, 2002; Amato and Patterson, 2017).

This paper adds to our understanding of the consequences of having children. Seminal

[1]The models of couple formation and dissolution incorporate match quality in two ways: through a match quality parameter and through partner quality. A set of papers have modelled relationship quality as a parameter that is known in the current period but that receives stochastic shocks in the future (Chiappori and Weiss, 2006, 2007; Gemici and Laufer, 2011; Bruze et al., 2015; Voena, 2015; Goussé et al., 2017; Low et al., 2018). Others consider it to be deterministic but unobserved, where couple members only get a noisy signal each period about true quality (Brien et al., 2006; Blasutto et al., 2020; Antler et al., 2022). A second strand of models consider match quality to be contingent on partner quality in terms of different socio-economic characteristics (Greenwood et al., 2017; Eckstein et al., 2019; Low, 2024).

papers in gender economics have documented the differential effect of fertility on women relative to men on several labour market outcomes (Blau and Kahn, 2017; Bertrand, 2020; Goldin, 2021). A great volume of work has focused on accounting for the share of the gender wage gap explained by fertility and on studying the channels through which it operates (Goldin, 2014; Adda et al., 2017; Kleven et al., 2019b; Cortés and Pan, 2023; Cavapozzi et al., 2021; Kleven, 2022; Albanesi et al., 2023), in different countries (Kleven et al., 2019a, 2025) and across different cohorts (Goldin, 2021). Other work has discussed the deeper implications of having children on outcomes like identity (Akerlof and Kranton, 2000, 2010; Bertrand et al., 2015), well-being (Dolan et al., 2008; Clark et al., 2008; Blanchflower, 2009; Ferrer-i Carbonell, 2013) and mental health (Ahammer et al., 2023).[2]

The findings across these studies consistently reveal a stark asymmetry: fathers experience little to no impact, while mothers face significant consequences. In contrast, this paper identifies a consequence of childbirth shared by both parents, offering novel insights into the experiences of fathers. Importantly, this differs from studies of couple-level outcomes like relationship tenure or divorce, which inherently involve both partners (Lillard and Waite, 1993; Svarer and Verner, 2008). By focusing on an individual-level outcome like RQ, we can test whether childbirth uniquely affects each parent. Moreover, the implications of RQ extend beyond parents to future outcomes of their children, who are affected by exposure to a poor-quality parental relationships (Piketty, 2003; Björklund and Sundström, 2006) and by parental divorce (Gruber, 2004; Björklund et al., 2007).

Finally, this paper contributes to our understanding of household time allocation, which is a long standing issue (Lundberg and Rose, 2000; Lundberg, 2005). This topic has gained even greater relevance during and after the COVID-19 pandemic, which introduced unprecedented disruptions to housework and childcare (Sevilla and Smith, 2020; Hupkau and Petrongolo, 2020; Alon et al., 2020; Farré et al., 2021; Del Boca et al., 2020). The dynamic impact of the birth of the first child on paid and unpaid work has previously been studied by Aguilar-Gomez et al. (2019) in Mexico and Siminski and Yetsenga (2022) in Australia. We replicate their findings for the United Kingdom and extend them to uncover heterogeneity in this impact based on household time arrangements before birth, providing the novel finding that gender-based specialization occurs after birth, irrespective of pre-birth arrangements. Furthermore, we provide evidence of the non-material costs of household specialization, highlighting its impact on relationship quality.

Roadmap. The rest of the paper is organized as follows. Section 2 describes the data and presents our measure of relationship quality. Section 3 describes the empirical strategy and discusses identification. Section 4 presents the main results and considers their robustness.

[2]The papers cited on identity and well-being do not directly address the influence of children on those outcomes, but they simply discuss their potential implications.

Section 5 explores changes to household specialization as the potential mechanism at play. Section 6 concludes.

2 Data and Measures

The analysis is based on data from Understanding Society, a longitudinal household survey in the United Kingdom representative of the country's population (University of Essex and Research, 2024). Information regarding the primary outcome spans from 2009 to 2022 and is based on data collected about couples in waves 1, 5, 7, 9, 11 and 13. We source additional data from the British Household Panel Survey (BHPS), which predates Understanding Society and covers the period from 1991 to 2008. Crucially, the BHPS allows us to identify both past and current partners, along with detailed characteristics of the relationships, including tenure.

2.1 Measure of Relationship Quality

Understanding Society conducts a 10-question survey every other data collection wave. These questions revolve around individuals' relationships with their cohabiting partners, including questions like "How often do you and your partner quarrel?". Respondents rate these questions on a scale from "All of the time" to "Never" on a six-point Likert scale. The survey also includes questions regarding relationship happiness and shared interests. Table 1 contains the full set of questions.[3] These questions are asked individually to all respondents living with a partner, regardless of marital status. This data is available in alternating waves spanning from 2009 to 2022.[4]

We categorize the items in Table 1 in two blocks based on the information they convey: (a) subjective assessments of the relationship and (b) couple time use items. The items in the first block contain information related to the degree of happiness and conflict in the relationship, which are individual perceptions that may not be shared by both partners. In the second block, the items refer to joint activities and how couples spend their time together. Including both types of items provides a nuanced overview of the relationship.

To construct the measure, we first transform the responses to all the items in Table 1 such that lower values correspond to poorer couple attitudes. Using responses from the complete dataset, we conduct a factor analysis and retain the first factor as the comprehensive measure of relationship quality (RQ). All items have positive loadings and the factor accounts for 40.61%

[3]This questionnaire, excluding the questions on happiness and kissing, was originally used in psychology research to construct the cohesion and satisfaction sub-scales of the Revised Dyadic Adjustment Scale (Busby et al., 1995). Appendix A.1 provides the exact phrasing of the questions and the response options of the questionnaire.

[4]An implementation error in data collection wave 3 (starting on year 2011) makes the item "Do you and your partner engage in outside interests together?" incompatible with the rest of the waves. Hence, we cannot construct the outcome in this wave.

Table 1: Questions in the Understanding Society Partner module.

(a) Subjective assessment	(b) Couple time use
How often do you... ?	*How often do you... ?*
discuss or consider splitting	work together on a project
regret that you married or lived together	have stimulating exchanges of ideas
quarrel	calmly discuss something
get on each other's nerves	kiss your partner
What is the... ?	*Do you and your partner... ?*
degree of happiness of your relationship	engage in outside interests together

Notes: This table lists all the questions used to construct the relationship quality measure (RQ), drawn from the Understanding Society Partner module. This self-completion questionnaire is administered in waves 1, 5, 7, 9, 11 and 13, covering the period from 2009 to 2022. We group these questions in two blocks based on the information they convey. Panel (a) includes items on subjective assessments of the relationship. Panel (b) contains items about how couples spend their time together. The responses to the first four questions in each block are given in a six-point Likert scale of frequency; the degree of happiness question is reported in a seven-point scale; the number of outside interests is given in a five-point scale. The exact phrasing of the questions and response options is reported in Appendix A.1.

of the variation in the data.[5] The resulting variable is standardized with a mean of zero and a standard deviation of one. In this context, higher values of RQ indicate higher quality relationships. A more in-depth discussion of the construction, validation and analysis of the RQ measure can be found in the companion paper, Rodríguez-Moro and Román (2024).

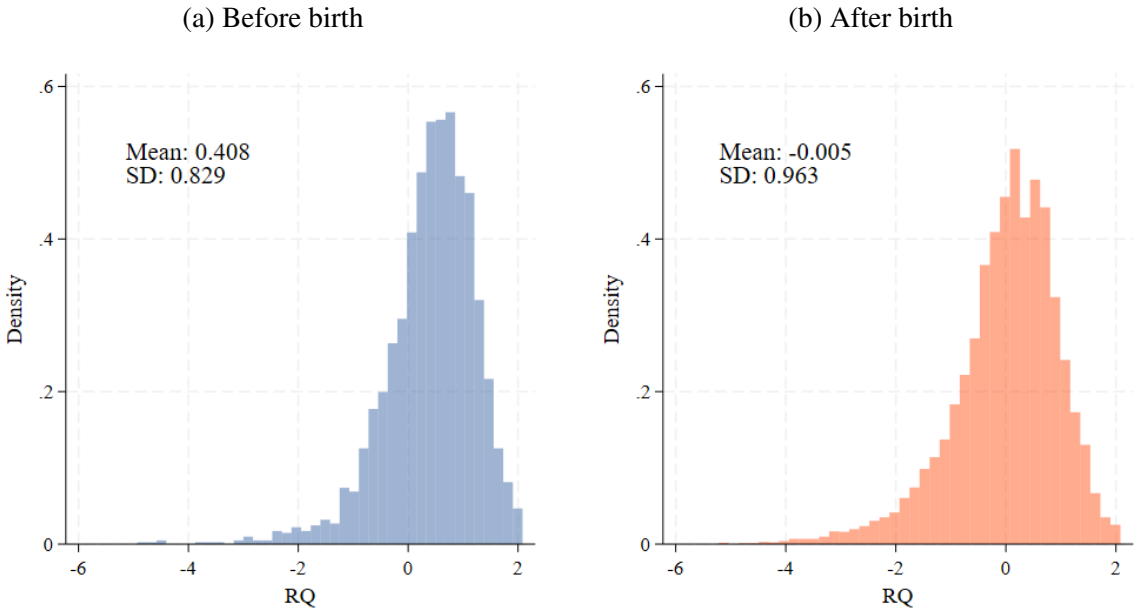
Figure 1 shows the distribution of RQ for individuals who become parents, separately for periods (a) before and (b) after the birth of their first child. In both cases, the distribution is left-skewed, indicating a higher frequency of high-quality relationships. Before birth, average RQ is notably higher than in the full sample and less disperse. After birth, instead, mean RQ drops below the full-data average. This is preliminary evidence of a decrease in RQ after having a child.

Validity of the measure. Given the novelty of the measure, in Appendix A.3 we conduct a series of tests to verify that RQ provides sensible information about the quality of relationships. We follow the life satisfaction literature and aim to validate two fundamental theoretical assumptions: informativeness and interpersonal comparability (Ferrer-i Carbonell and Frijters, 2004). The validation exercises are computed using all the available information on RQ (full data), instead of restricting to parents like in the remainder of the paper.

First, we evaluate the informativeness of RQ by assessing its ability to predict couples' decisions. Specifically, we test whether the distribution of RQ in periods preceding marital transitions (marriage and separation) differs from its distribution in the full data. Similarly, we compare the distribution of RQ in the analysis sample of parents with its distribution in the

[5]The retained factor has an eigenvalue of 4.06, while the next one has an eigenvalue of 1.45. 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 main analysis sample of parents, separately for the periods (a) before and (b) after the birth of their first child. The sample consists of individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18-45 (women) or 18-50 (men) at the time of birth, and were not cohabiting with children from previous relationships. Panel (a) includes observations up to four years before birth. Panel (b) spans from the year of birth to seven years after. These time frames align with the main analysis period. In the full data, RQ has a mean of 0 and a standard deviation of 1.

period before the birth of the first child. We further explore the link between RQ and couple dissolution in the regression analysis in Table A.3.

We find strong evidence that, before a couple's dissolution, the distribution of RQ lies below that of the full sample. In the regression analysis, a one-standard-deviation higher RQ is associated with a 0.9 percentage point lower probability of separating in the subsequent period. The opposite pattern holds for periods preceding marriage and fertility events. These results are consistent with the theoretical assumption that RQ is informative of relationship outcomes.

Second, we substantiate the interpersonal comparability of the measure, that is, we evaluate whether there is a degree of commonality in the concept captured by RQ. To do this, we study the correlation between partners' responses by regressing women's RQ on their male partners' RQ. Our analysis reveals a high level of correlation between responses, which remains robust even after controlling for individual and couple characteristics. This evidence supports the notion of objectivity in the measured concept.

Additionally, we verify that RQ evolves smoothly with age and relationship tenure, and examine its correlation with observable individual and couple characteristics. To assess the stability of RQ, we estimate a two-way fixed effects regression that includes age and relationship tenure non-parametrically. This allows us to obtain the age and tenure profiles of RQ from the full data. Both profiles are smooth and show no significant jumps. While the age profile

is increasing and nearly linear, the tenure profile of RQ displays a rapid decline in the early years of the relationship, which slows down as tenure increases. Lastly, we find that women and individuals with children consistently report lower RQ levels, whereas college educated and married individuals report higher levels.

2.2 Household Specialization

Household specialization refers to the division of labour between partners in terms of home and market production. Understanding Society provides data on the number of hours each individual works in the labour market and spends on routine housework per week.[6] We use this information to compute the shares of home and market production out of the couple total carried out by women as our proxies for household specialization. We refer to these as the *female share* of unpaid housework and paid labour market hours. A 50% share in both types of work indicates no specialization, larger shares imply female specialization and smaller ones male specialization.

Figure E.1 plots the distribution of the female share of paid and unpaid hours (a) before and (b) after the birth of the first child. Before birth, household specialization varies considerably. The distribution of the female share of paid hours is centred around 50%, while the distribution of unpaid work is more evenly spread, primarily above the 50% mark. After birth, both distributions experience noticeable changes: the mass of paid hours moves below the 50% threshold, while the share of unpaid hours becomes more concentrated above this point.

We classify couples based on the female share of paid and unpaid hours *before* the birth of their first child, distinguishing four types: (i) traditional couples, where women perform more than 50% of housework and men take a larger share of paid work; (ii) unbalanced couples, where women contribute more than half of both types of work; (iii) egalitarian couples, where men and women share both types of work roughly equally, with female shares between 40% and 60%; and (iv) counter-traditional couples, where men perform more than 50% of housework.[7]

[6] Specifically, the questions asked 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?". The responses to these questions are available for approximately half of the sample of parents. Following Borra et al. (2021), we do not consider childcare to be part of routine housework. Time use information on childcare is unavailable.

[7] Figure E.2 illustrates this classification based on the female shares of paid and unpaid hours. Each couple belongs to a single category, with no overlap between classes. Couples with female shares between 40% and 60% in both paid and unpaid work are always classified as egalitarian.

2.3 Sample Restrictions and Descriptives

The population of interest consists of individuals in cohabiting relationships, whether married or not, who become parents.[8] Our analysis focuses on the subset of individuals who become parents for the first time within the observation period. To ensure the robustness of our sample, we impose several criteria. First, we restrict our analysis to individuals who were aged 18 to 45 (women) or 50 (men) when they first became parents. Second, we exclude couples cohabiting with children from previous relationships. Third, we use only heterosexual couples. The resulting sample is an unbalanced panel of 1,760 individuals observed up to 6 times, capturing 97% of the fertility events in the data.

Table 2 presents summary statistics for the analysis sample in the period preceding childbirth. Panel (a) reports individual characteristics separately for fathers and mothers. On average, fathers are 32 years old, while mothers are four years younger. A smaller share of mothers than fathers are active in the labour market or employed. Mothers work four hours less per week in the labour market, earn lower monthly labour income and spend three and a half more hours on housework in comparison to fathers. Both report above-average RQ compared to the full sample. Panel (b) summarizes couple characteristics. On average, respondents cohabit for four years before having their first child. Consistent with the time-use patterns, mothers contribute slightly less than 50% of paid labour market hours and take on more than 60% of unpaid housework.

Table E.1 summarizes the same set of characteristics separately for each couple type based on pre-birth household specialization.[9] Importantly, there are no significant differences across couple types in the timing of childbirth in terms of age and relationship tenure, or in marital status. The most pronounced differences in labour market outcomes emerge between traditional couples and the other groups. Traditional couples have, on average, lower educational attainment and lower household income. They also exhibit the largest within-household differences in paid and unpaid work hours (differing by 10 and 7 hours, respectively), and in monthly labour earnings. RQ is lowest among unbalanced couples, aligning with the full data average, and highest among egalitarian couples. Egalitarian couples also have the highest household income.

[8] Some studies also examine infertile individuals who attempt to have children but are unable to conceive, offering insights into the counterfactual scenario where no child is born (Bögl et al., 2024). While our main analysis focuses on individuals who become parents, since we argue this is the policy-relevant population, we conduct robustness checks using infertile individuals as the control group in Appendix B.3.

[9] The reduced sample sizes result from the limited availability of responses to time use questions and the unbalanced nature of the sample, where not all individuals are observed in the period immediately before birth.

Table 2: Summary statistics for the period before the birth of the first child.

(a) Individual characteristics			(b) Couple characteristics	
	(1) Fathers	(2) Mothers		(3) Couples
Age	32.00	28.38	Tenure	4.187
	(6.335)	(6.053)		(3.310)
College educated (%)	33.76	36.27	Married (%)	42.57
	(47.30)	(48.09)		(49.21)
Active in labour mkt (%)	87.13	84.07	Monthly household income (£)	4040.0
	(33.48)	(36.59)		(2992.0)
Employed (%)	82.60	77.91	Female share of paid work	0.472
	(37.88)	(41.47)		(0.210)
Monthly labour income (£)	2347.7	1654.2	Female share of housework	0.630
	(1685.9)	(1306.7)		(0.204)
Weekly work hours	31.49	27.27		
	(16.99)	(16.15)		
Weekly housework hours	5.147	8.582		
	(4.039)	(6.254)		
RQ	0.353	0.383		
	(0.859)	(0.895)		
Observations	2715	3261	Observations	4125

Notes: This table presents summary statistics for the main analysis sample of parents in the period before the birth of their first child. The sample consists of individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18-45 (women) or 18-50 (men) at the time of birth, and were not cohabiting with children from previous relationships. Panel (a) reports the mean values of individual characteristics separately for women and men, and panel (b) summarizes couple characteristics. Couple characteristics can be retrieved even when only one partner is observed. Female shares of paid and unpaid work represent the share of total weekly household paid labour market hours and unpaid housework hours performed by women. Income measures are presented in gross terms. Standard deviations are reported in parentheses.

3 Empirical Strategy

Our primary objective is to identify the causal effect of first child birth on RQ. To do so, we exploit variation in the timing of first births and adopt a dynamic difference-in-differences (DiD) approach. Consider the usual two-way fixed effects (TWFE) regression model:

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

where $y_{i,t}$ denotes the RQ of individual i in period t. We denote as G_i the year in which individual i has a first child, so $t - G_i$ represents event-time, the number of periods since birth. Equation 1 includes the full set of event-time dummies, as well as individual fixed effects

and period dummies. The coefficients of interest are δ_j, which capture changes in RQ around first child birth. Recent econometric literature has highlighted biases derived from estimating TWFE models via OLS in settings with staggered treatment adoption (De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak et al., 2022; Gardner, 2022; Roth et al., 2023). To address these concerns, we use the estimator proposed by Callaway and Sant'Anna (2021).[10]

The causal interpretation of the event-time coefficients δ_j as average treatment effects on the treated (ATT) relies on two identification assumptions. First, changes in RQ do not systematically predict when individuals have their first child (*no anticipation*).[11] This assumption would be violated, for example, if couples decided to have children in response to temporary periods of high RQ.[12] Second, in absence of treatment (first child birth), RQ would have evolved similarly across individuals, regardless of when they had their first child (*parallel trends*). While these assumptions cannot be directly tested, we assess their plausibility in Appendix B.2.

First, we verify that the evolution of RQ is flat during the periods preceding the birth of the first child, which supports the plausibility of both assumptions. However, flat pre-trends may arise from averaging out couples with varying circumstances: some may decide to have a child to address relationship issues, while others may do so because they are momentarily very happy with their relationship. This would be a violation of no anticipation. We provide evidence that this is not the case showing that variation in RQ is not larger during the periods preceding birth, and that there are no significant deviations from individual-specific means that preface birth. Second, we perform "long" comparisons of pre-birth RQ values relative to the baseline period, following Roth (2024), and find that the coefficients are not statistically different from zero. Third, we show that parents who experienced delays in the birth of their first child, namely those who have used fertility treatments or faced involuntary pregnancy interruptions, do not differ from those who did not before childbirth. Fourth, our results hold when restricting comparisons to individuals with similar RQ levels before birth, comparisons which would make the parallel trends assumption more plausible. Finally, we do not find evidence of heterogeneity in treatment effects for parents whose first children were born in different years, which supports the assumption of homogeneity of treatment effects across cohorts needed for the time aggregation in our main results.

Importantly, this methodology does not require childbirth to be random. Having a child is a deliberate decision and individuals who become parents differ from those who remain childless in several aspects, including their level of RQ, as shown in Table B.2. To abstract from these

[10] A detailed summary of this methodology is provided in Appendix B.1. We implement it using the `csdid` package on Stata.

[11] This assumption can be relaxed to limited anticipation if the anticipation horizon can be specified. We relax this assumption in complementary analyses to allow for outcomes to react during pregnancy.

[12] Importantly, this assumption relates to *changes* in RQ before childbirth, pertaining to the longitudinal dimension of the data. This differs from the finding in Figure A.3 (b) that the average *level* of RQ is higher before having a child, in cross-sectional comparisons.

differences, our main analysis is restricted to parents and identification exploits only variation in the timing of the first birth. Accordingly, our estimates should be interpreted as the effect of childbirth on those who do become parents. In complementary analyses in Appendix B.3 we use individuals who never try to have children and those who try unsuccessfully as alternative comparison groups. Since our design relies on timing, we also examine in Table B.1 whether the timing of childbirth is predictable. Specifically, we regress an indicator for birth in the subsequent period on pre-birth RQ, economic characteristics and couple characteristics, controlling for individual and period fixed effects. We confirm that changes in RQ before birth do not predict timing, providing additional support for the no-anticipation assumption. The only significant predictor is marriage.

4 Impact of Children on Relationship Quality

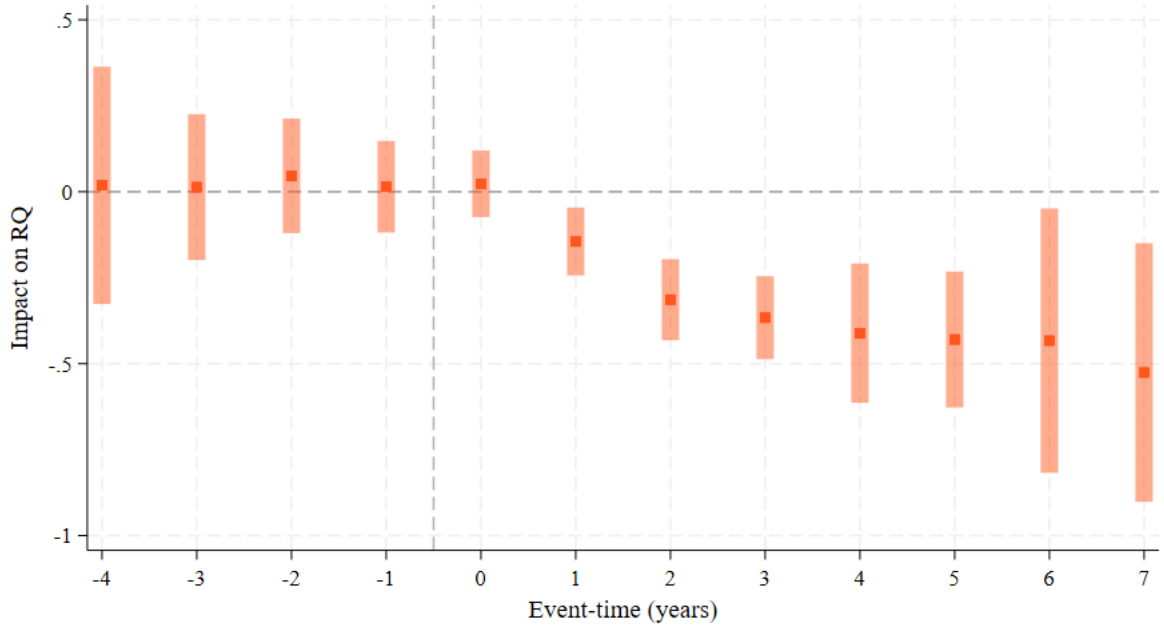
Figure 2 presents the results from a dynamic difference-in-differences estimation of the effect of first child birth on RQ, using the method proposed by Callaway and Sant'Anna (2021). The reported coefficients represent the impact of childbirth on RQ at each period relative to birth. Since RQ is standardized, the estimates are interpreted in standard deviations. The available data allows us to examine up to four periods before and seven periods after the birth of the first child. We also plot 95% confidence intervals.

The coefficients corresponding to periods before birth are not significantly different from zero. These flat pre-trends are consistent with the identification assumptions of no anticipation and parallel trends discussed in Section 3. The coefficient corresponding to the period of birth is also statistically non-distinguishable from zero. This means that RQ does not immediately react to the birth of the first child.

There is a significant decrease in RQ during the first four years after birth. At this time, the decrease in RQ stabilizes at around half a standard deviation. As an illustration, individuals ranked in the 75^{th} percentile of the RQ distribution before the birth of their first child are pushed down to median RQ over this time period. This negative impact stabilizes after four years and until seven years after child birth. This suggests that, although having a child shifts RQ down, it does not change how this variable evolves over time.

Impact on mothers and fathers. Most of the literature on child penalties finds strongly asymmetric effects of first child birth on mothers and fathers. Studies on labour market outcomes find that women reduce their participation in the labour force, work fewer hours, and experience declines in wages and earnings following childbirth, while men show no such impact. This contributes to increasing within-household inequality and widens the gender wage gap (Adda et al., 2017; Blau and Kahn, 2017; Kleven et al., 2019b; Cortés and Pan, 2023;

Figure 2: Dynamic effect of first child birth on RQ.



Notes: This figure presents the results from a dynamic difference-in-differences estimation of the effect of first child birth on RQ, using the method proposed by Callaway and Sant'Anna (2021). The sample includes individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18–45 (women) or 18–50 (men) at the time of birth, and were not cohabiting with children from previous relationships. We also plot 95% confidence intervals.

Goldin, 2021; Albanesi et al., 2023). Beyond labour market outcomes, gender differences in the impact of childbirth also emerge on time spent on housework (Aguilar-Gomez et al., 2019; Siminski and Yetsenga, 2022), leisure (Aguiar and Hurst, 2007), sleep (Costa-Font and Flèche, 2020) and even mental health (Ahammer et al., 2023).

Motivated by this, we study potential gender differences in RQ around first child birth. Despite RQ being a shared concept, as discussed in Subsection 2.1, individual perceptions may vary. However, as shown in Table 2, mothers and fathers report nearly identical levels of RQ prior to childbirth. In Figure C.1 (a), we repeat the main analysis for men and women separately. We find a comparable decline in RQ for both parents during the first three years post-birth. After three years, RQ stabilizes for men, while it declines further for women in year four. Figure C.1 (b) examines whether these differences are statistically significant, comparing RQ responses between both partners. The results show no significant gender difference.

The symmetric effect of childbirth on both mothers' and fathers' RQ makes our results distinct from prior child penalty literature. While most penalties primarily affect women, RQ reflects a broader effect that involves both partners, indicating a utility loss on both sides. Notably, this decline occurs for fathers, even though they typically do not see significant changes in many outcomes traditionally studied, providing valuable insight into their experiences and contrasting with the notion that fathers are unaffected by parenthood.

Implications for couple dissolution. To better understand the relevance of these results, we compute a back-of-the-envelope calculation of the implications for couple dissolution. In Section 2 we provide evidence of the capacity of RQ to predict marital transitions, finding lower values of RQ preceding couple separation. The regression analysis in Table A.3 indicates that, in the main analysis sample of parents, a standard deviation increase in RQ is associated with a 0.7 percentage point lower probability of separation.

The magnitude of this coefficient highlights the relevance of RQ in dictating couple dissolution. Based on that, the observed half standard deviation decline in RQ following childbirth would be associated with a 17.5% higher probability of separating.[13] Furthermore, this result is in line with the findings of previous research on the impact of children on couple dissolution (Lillard and Waite, 1993; Svarer and Verner, 2008).

As argued in Browning et al. (2014), RQ is just one aspect of how the presence of children influences couple dissolution decisions. Having a child tends to raise separation costs for parents while also increasing the value of maintaining the relationship through the creation of economies of scale within the couple, which would reduce the probability of dissolution. While exploring these factors jointly is crucial given the detrimental effects that separation can have on children, studying them separately may be equally relevant, especially considering the adverse effects of being exposed to a low parental RQ.

General happiness and RQ. We next study the effect of the birth of the first child on general happiness to obtain a more comprehensive picture of the impact of children on individual utility. RQ can be interpreted as part of the utility exclusively attached to the couple, distinct from the utility derived from having a child. Therefore, the impact of childbirth on happiness and RQ would not necessarily align in the same direction.

In Figure C.3 (a) we repeat the main analysis using standardized general happiness as the outcome.[14] We find that the birth of the first child induces an immediate increase in general happiness during the pregnancy and birth periods. However, the absolute magnitude of this impact is smaller than the one documented for RQ. Moreover, the impact on general happiness only lasts one period, not mirroring the persistence observed in the impact on RQ.[15]

The depicted patterns align with the hypothesis that happiness encompasses a component

[13] Figure C.2 plots the share of couples dissolving out of all the couples observed each year in the full data and in the parent sample, showing that 2% of the existing couples in our data separate each year. According to the Office for National Statistics (2022), on average 1.44% of the married couples in fertility ages (20-45 year-olds) living in England and Wales divorced yearly between 2009 and 2021. We do not have the equivalent data for cohabiting couples.

[14] The question used is "Have you recently been feeling reasonably happy, all things considered?". We take two periods before birth as the baseline to allow for anticipation during the pregnancy period.

[15] There is a decreasing trend in general happiness after birth. Blanchflower and Oswald (2004) find that general happiness is U-shaped in age, reaching the minimum at ages 37-41 in the United Kingdom. In our sample, the average age at birth is 30, as seen in Table 2. Thus, the documented decline corresponds to a life-cycle trend that our main methodology does not allow accounting for.

attached to the couple (RQ) and another attached to the child.[16] The main finding of this paper indicates that the couple component decreases after birth. Hence, the fact that the sum of both components - couple and child - does not decrease suggests that the child component is positive. Conclusively, while the event of childbirth may indirectly reduce happiness, entailing negative consequences for RQ, this evidence suggests that the direct contribution of children to happiness is not negative.

Other measures and shocks. We compare the magnitude and persistence of our main finding with the effect of this event on other subjective outcomes at the individual level. Figure C.3 (b) displays the results for subjective well-being.[17] In contrast to our main finding, we do not document an effect of first childbirth on subjective well-being. Ahammer et al. (2023) use a similar approach to study the impact on mental health and find a significant increase in antidepressant prescriptions for the sub-population of individuals that visit the psychiatrist. They find a greater effect on mothers, which they attribute to the higher childcare burden borne by them.

Finally, we examine the significance of the birth of the first child as a distinct shock to RQ by comparing it to another major life event: becoming unemployed. Previous studies have found a positive correlation between male unemployment and divorce probabilities (Jensen and Smith, 1990; Charles and Stephens, 2004; Eliason, 2012; Doiron and Mendolia, 2012). Figure C.4 displays the results obtained from computing our main estimation around unemployment events on RQ.[18] Unlike childbirth, unemployment is not significantly associated with changes in RQ, which contrasts with the findings on divorce.

These insights can inform models of family formation and dissolution in multiple ways. As mentioned earlier, RQ is a measure of the non-material benefits of being in a relationship. The difference between the association of unemployment with RQ and with divorce may indicate that, while the job loss of a partner reduces the material gains from being in a relationship with them, it may not have such impact on non-material gains. Furthermore, our main finding evidences the smooth and path-dependent nature of RQ, which could guide the modelling of the distribution of this factor within family formation and dissolution models.

[16] In Table C.1, we quantify the association between happiness and RQ for individuals who have not become parents by age 50, the end of the fertility cycle. We find that a one standard deviation higher RQ is associated with a 0.2 standard deviation higher happiness measure, on average.

[17] The subjective well-being measure combines the responses to the 12 General Health Questionnaire (GHQ-12), related to physical and mental health. Some of the questions relate to sleep and being under strain, which are mechanically affected by having a child. Previous literature has discussed the association between children and well-being, although not in a causal way (Dolan et al., 2008; Clark et al., 2008; Blanchflower, 2009; Ferrer-i Carbonell, 2013).

[18] We focus on the first instance of unemployment for an individual within our observation period and exclusively consider individuals who become unemployed at some point ("not yet unemployed") as our control group. We do not make any causal claims about these results.

4.1 Robustness

We address five potential concerns that could affect the interpretation of our results. First, we study whether the results are robust to using alternative counterfactuals. Second, we explore whether the impact on RQ is driven by specific items within the measure. This could result from an uneven response of items across different blocks, or from a change in how the items are valued after the birth of a child. Third, we consider whether the persistent impact on RQ is solely due to subsequent children or if it represents a lasting effect of the first child's arrival. Fourth, we take into account the limitation that RQ cannot be observed after a couple dissolves, which may introduce some selection bias that could distort our results. Fifth, we examine whether the timing of child birth, in terms of the age of the parents and the duration of the relationship, has any influence on our findings.

Alternative counterfactuals. Our main analysis compares individuals who become parents to those who will become parents later in the observation window. We argue that this is the most policy-relevant study sample, since childbirth is a meaningful event for individuals who have children. To test the robustness of our findings in Appendix B.3, we consider two alternative counterfactuals: individuals who want children but do not manage to have them (infertile individuals) and individuals who neither want nor have children (never-parents).

First, we consider individuals who want children but do not manage to have them, which could provide insight into the counterfactual scenario where childbirth does not occur (Bögl et al., 2024). We identify them as individuals who become pregnant but never experience a live birth. Using this group as the control and exploiting the timing of their first failed pregnancy, we find that the estimated effect of childbirth on RQ is halved in comparison to the main result as shown in Figure B.6. This reduction in the estimated coefficients occurs because RQ also declines for infertile individuals following a failed pregnancy, implying that the estimated effect should be interpreted as a lower bound.

Second, we consider individuals who neither want nor have children, identified as those who have never been pregnant and remain childless by the end of their fertility cycle. Following Kleven et al. (2019b), we employ a design that assigns placebo births to never parents. While this group was initially excluded due to concerns about comparability to future parents and about the plausibility of the identification assumptions, the results in Figure B.6 are similar to our main findings, with improved precision.

Time invariance of RQ. It is a concern whether the decrease in RQ might be solely attributed to specific items within the measure, particularly those related to couple time use. This concern arises from the fact that, after the birth of the first child, the available time for couples to spend together decreases significantly due to the new responsibilities that occupy a significant portion

of their time (Bianchi, 2011). Given that individuals highly value spending leisure time with their partners (Georges-Kot et al., 2024), this decrease could directly reduce RQ, potentially being the only driver of the documented decrease.

To explore this, we repeat the factor analysis excluding the time use items and using solely the items containing subjective assessments of the relationship. Figure D.1 (a) plots the results of conducting our main analysis on the resulting measure. The documented decrease is much more pronounced than what we observed on RQ. For completeness, we repeat this procedure using only the time use items in Figure D.1 (b). In this case, the impact is less pronounced and not statistically different from zero after three years. This suggests that our primary finding is not only driven by reduced time together, but is mainly influenced by changes in how individuals assess their relationship.[19]

However, the possibility remains that individuals change how they value different items within the RQ measure after becoming parents. To address this, we construct a new RQ measure using data solely from individuals who are already parents. We use those observations to obtain the factor loadings in the factor analysis and construct the measure for the entire sample. Figure D.2 displays the results of our main analysis using this measure, showing that our primary result remains consistent.

Subsequent fertility. Our analysis focuses exclusively on the birth of the first child. It is unclear whether the persisting impact on RQ represents a lasting effect of the first child's arrival or if it is merely a consequence of subsequent children. In Figure D.5 we split the analysis sample in two groups: (a) individuals with only one child or (b) those with subsequent children. About 45% of the individuals in our sample are observed having an only child. The impact occurs slightly faster and is somewhat more pronounced for individuals with more than one child, but they experience a recovery four years after the first child's birth. Thus, we cannot attribute the documented persistence on RQ to parents of more than one child.

However, this analysis masks censorship in the sample, as many individuals have not reached the end of their fertility cycle at the last observation. We restrict the analysis to individuals observed at age 40 or older, which better approximates their lifetime fertility. This restriction remarkably reduces the sample and does not allow for a dynamic analysis. Table D.1 displays the static DiD results obtained through the usual estimation method. Note that the obtained ATT is an average of the impact all periods after, where periods with more observations get higher weights (Callaway and Sant'Anna, 2021). The results suggest that subsequent children do not have an additional negative impact on parental RQ. In fact, they may somewhat offset the initial decline in RQ following the birth of the first child.

[19] Figure D.3 and Figure D.4 display the impact of first child birth on each subjective assessment and couple time use item separately. The item responses are standardized for comparability. The degree of happiness with the couple experiences the largest impact.

It is important to note, however, that this could be a selected sample, as couples who choose to have a second child may have had a different experience compared to those who did not. We focus exclusively on second-time parents to examine whether the birth of a second child has any impact on RQ. Figure D.6 shows the results of the main analysis around the birth of the second child. Although we observe a pre-trend, likely reflecting the decline in RQ following first child birth, there is no significant effect associated to the birth of the second child.

Selected sample. Due to the nature of RQ, it cannot be observed once a couple dissolves, potentially introducing selection bias that could distort our results. We investigate the attrition from the sample due to separation, by estimating the impact of birth on the likelihood of separation. Note that, by construction of the sample, we do not observe couples dissolving before the first child is conceived. Thus, the results should not be interpreted as causal. Figure D.8 (a) indicates that dissolution probabilities are around 2% over the periods analyzed after birth, suggesting that the attrition from the sample is not a significant concern. We repeat the analysis excluding from the sample couples who separate over the observation period, as shown in Figure D.8 (b). These results largely mirror our primary findings, alleviating concerns about selection bias.

Age and tenure at birth. In addition to the birth of the first child, the timing of this event may significantly affect RQ (Kleven et al., 2019b). We explore this by separately examining individuals who had children at different ages and relationship tenures. Figure D.9 presents the average RQ at each event-time period by (a) age bin and (b) tenure bin, normalizing RQ to zero in the period before birth. While baseline RQ levels differ for individuals who had children at different ages, the post-birth evolution of RQ remains consistent across groups.

Baseline levels also differ across tenure groups. In this case, individuals who had their first child at the earliest stages of their relationship experience a slightly greater decrease in RQ . However, this more substantial decrease could be partially attributed to the general trend of RQ decreasing during the early stages of a relationship, as seen in Figure A.5.

To ensure the robustness of our main findings, we include these variables as controls, as presented in Figure D.10. Note that Callaway and Sant'Anna (2021) does not allow to study the dynamic impact of age and tenure on RQ, since it only allows to control for pre-birth levels of covariates. The differences for all sets of controls remain minimal until six years after the birth of the first child. At that point the coefficients of the specification including both age and tenure become larger. Thus, our main results could be interpreted as a lower bound.

To account for potential dynamic effects across age and tenure, we follow the event-study approach popularized by Kleven et al. (2019b). Specifically, we estimate the effects using OLS while replacing individual fixed effects with the full set of age and relationship tenure dummies. This analysis relies on the assumption that treatment timing is independent of unob-

served factors once age, tenure and calendar time are conditioned on. The results displayed in Figure D.11 confirm that our main finding holds when accounting for the dynamic evolution of RQ over the life cycle and relationship duration.

5 Mechanism: Changes in Household Specialization

The birth of a child introduces a significant and unprecedented shift in how couples allocate their time. New responsibilities related to childcare emerge and the demand for routine housework increases, requiring couples to adjust how they divide paid work in the labour market and unpaid work at home. This reallocation of responsibilities shapes the roles that each partner assumes within the household, such as primary earner or main caregiver. Research shows that men's labour market outcomes remain largely unaffected by the transition to parenthood, while women experience substantial declines in labour force participation and working hours (e.g., Kleven et al., 2019b; Goldin, 2021) and take on longer home production times (Aguilar-Gomez et al., 2019; Siminski and Yetsenga, 2022).

Kuziemko et al. (2018) show that the magnitude of these changes is often unforeseen. Women, particularly high educated ones, underestimate the impact of having a child on their labour market outcomes and do not envision themselves as full-time caretakers after motherhood. We argue that the unanticipated, structural changes in the division of paid and unpaid work times within the couple may be mediating the deterioration in RQ that follows the birth of the first child.[20]

5.1 Impact of Children on Household Specialization

We start by documenting how mothers and fathers adjust their time use around the birth of their first child. Specifically, we re-estimate our main specification using weekly hours in paid employment and unpaid housework, as defined in Section 2, separately for mothers and fathers.[21] Figure 3 (a) plots the findings on paid work and Figure 3 (b) does it for unpaid work.

As seen in Table 2, before birth, fathers spend on average 31 weekly hours in paid work, while mothers work roughly four hours less. Figure 3 (a) indicates that, after childbirth, mothers reduce their labour market hours sharply and persistently, stabilizing at about 14 fewer hours within two years. In contrast, fathers display only a slight decline later in the child's life, which

[20] In Appendix F we consider other potential mechanisms, including changes in gender norms, household income and the within-couple earnings gap. The results suggest that household specialization is the most plausible channel.

[21] To account for life-cycle trends in employment and home production, we control for age in Equation 1. The identifying assumptions remain those in Section 3, with the nuance that the parallel trends assumption needed is conditional on age. Since the Callaway and Sant'Anna (2021) method can only account for baseline levels of control variables, overlooking non-linear trends, we may still observe patterns attributable to the life-cycle.

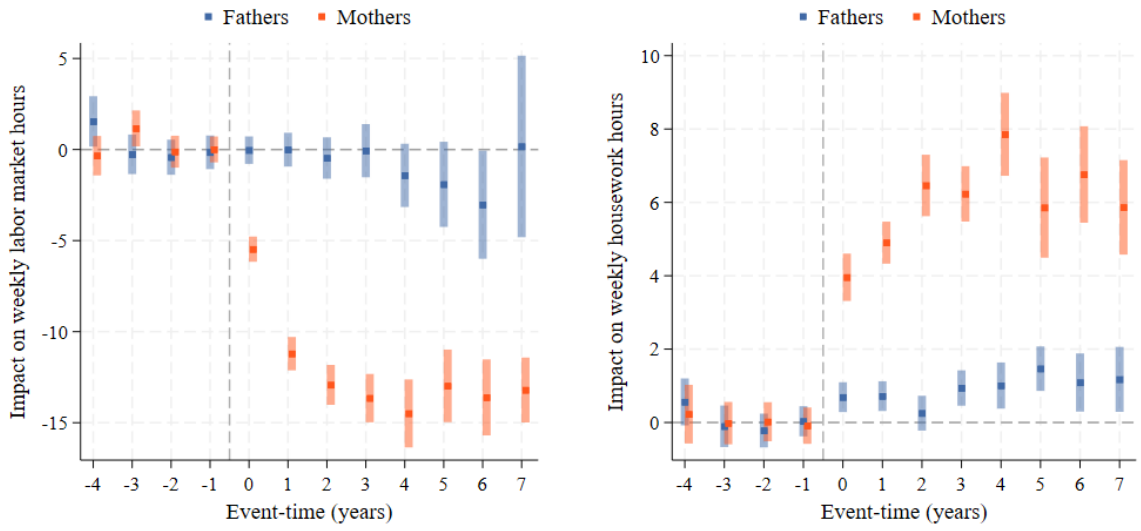
is consistent with life-cycle trends. In terms of housework, in Figure 3 (b), mothers increase their weekly unpaid work time by about five hours in the first year after birth, relative to a pre-birth baseline of nine hours. This increase is sustained for about four years, stabilizing at around seven hours above the baseline. Fathers raise their housework time from five to six hours per week. Although modest, this increase is statistically significant.

Taken together, these results show that, following first birth, women give up the equivalent of more than one and a half days of full-time work and assume nearly a full extra day of unpaid work each week. Conversely, men increase their housework time by one hour and maintain their labour market hours. These findings are consistent with prior evidence on the impact of having children on labour market outcomes (Kleven, 2022) and home production (Aguilar-Gomez et al., 2019; Siminski and Yetsenga, 2022). They reveal asymmetries in how couples absorb the additional workload from children, with women assuming the bulk of new household responsibilities.[22] Yet both partners report similar declines in RQ, suggesting that this reorganization generates tensions at the couple level. This reallocation of time not only reinforces gender gaps in paid and unpaid work, but also reshapes household roles and specialization.

Figure 3: Impact of first child birth on paid and unpaid hours.

(a) Weekly labour market hours

(b) Weekly housework hours



Notes: This figure presents the results from a dynamic difference-in-differences estimation of the effect of first child birth on time use variables, using the method proposed by Callaway and Sant'Anna (2021). The sample includes individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18–45 (women) or 18–50 (men) at the time of birth, and were not cohabiting with children from previous relationships. Panel (a) displays the results using average weekly paid labour market hours as an outcome, and panel (b) presents the results for average weekly housework hours (cooking, cleaning, doing laundry, etc.). On the period before first birth, fathers work 31 hours in the labour market and 5 hours at home, while mothers spend 27 hours in paid work and 9 in housework. We control for age at first birth to account for potential cohort-age differences. We also plot 95% confidence intervals.

[22] Andrew et al. (2021) document that mothers also spend more time on childcare than fathers do.

To study how roles and household specialization evolve, we track changes in women's relative contribution to market and home production. To that end, we calculate the female share of total paid and unpaid work hours within the couple and classify couples into four categories based on their pre-birth values, as outlined in Section 2: traditional, unbalanced, egalitarian and counter-traditional. These pre-birth values of the female shares provide information about the roles in place before parenthood and the existing degree of household specialization, which may reflect underlying comparative advantages within the couple.

We investigate the association between first childbirth and these shares by couple type. Figure 4 (a) plots the average female share of paid work at each period relative to birth. The contribution of women to paid work decreases after birth across all types, accounting for less than half of the household total. The largest drop occurs in unbalanced couples, where the female share falls by nearly 15 percentage points. In egalitarian and counter-traditional couples women's share stabilizes around 35%, representing a decrease of approximately 10 percentage points. Traditional couples are the least affected in this dimension, although the female contribution remains low compared to the other groups.

Figure 4 (b) displays the same shares for unpaid work. Regardless of the initial allocation, women in all couple types assume a larger share of unpaid work than men do after birth. Counter-traditional and egalitarian couples experience the largest changes, where female shares rise by about 20 and 15 percentage points, respectively, accounting for over 60% of the total. In contrast, traditional and unbalanced couples, where women already did over three quarters of unpaid work before birth, do not experience significant changes.[23]

This evidence suggests that couples across all types adjust their degree of household specialization after childbirth, transitioning towards more traditional, gender-based roles. However, the changes they undergo to establish these new roles vary considerably. Traditional couples experience the smallest adjustments, maintaining men's predominant role in the labour market and women's in home production. Among unbalanced couples, the division of housework remains unchanged, but the primary responsibility for paid work shifts from mothers to fathers. Notably, women in these couples are the only ones whose total time share declines after birth. Egalitarian couples experience substantial changes in the division of both types of work, abandoning the equal distribution of tasks. After birth, women adopt a dominant role at home whereas men specialize in the labour market. Counter-traditional couples experience the most pronounced shifts in the division of housework. Men abandon their predominant role in this domain, which is assumed by women, while preserving or strengthening their role in the labour market.

In traditional models of the household, paid and unpaid work are assumed to be allocated among household members based on the degree of substitutability between their labour market

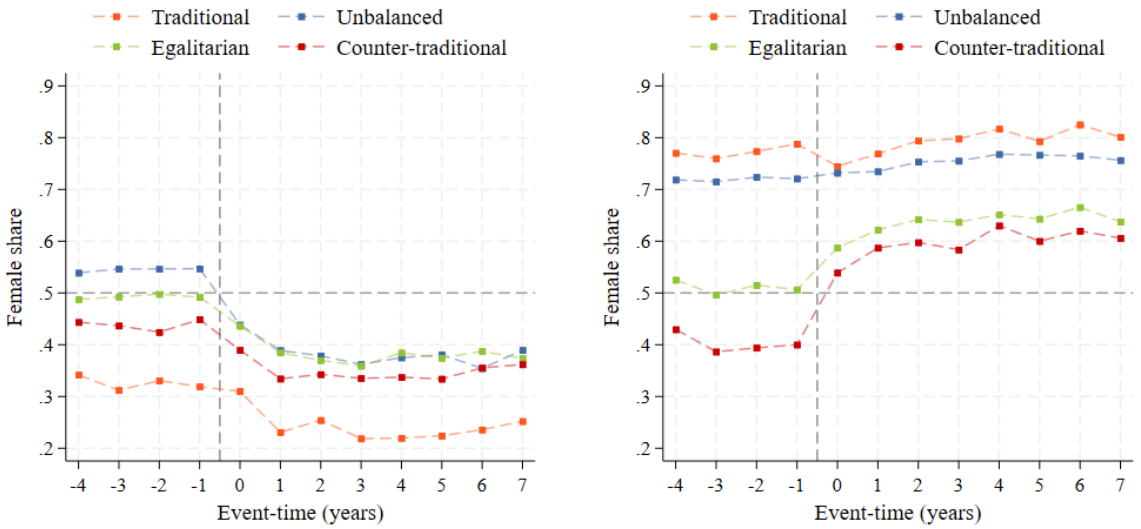
[23] In Figure E.3, we examine changes in labour market work and housework hours by couple type and gender, confirming that the shifts in female shares are driven by changes in women's time use.

and domestic inputs (Becker, 1991). Our findings challenge the plausibility of this assumption. If the pre-birth distribution of tasks reflects potential comparative advantages within the couple, then the post-birth exchange of responsibilities cannot be fully explained by differences in skill substitutability. Instead, it points to the presence of other forces, such as frictions in the labour market, for instance in the shape of statistical discrimination against mothers (Petit, 2007; Becker, 2010), or identity considerations (Akerlof and Kranton, 2000; Ichino et al., 2019; Farré et al., 2021). This evidence suggests that such factors become prevalent with parenthood, confirming the difficulties that individuals may face to anticipate changes in roles and in household specialization (Kuziemko et al., 2018). We argue that this mismatch in expectations may, in turn, affect RQ.

Figure 4: Household specialization around first child birth by couple type.

(a) Paid work share

(b) Unpaid work share



Notes: This figure displays the average the female shares of (a) paid and (b) unpaid work — defined as the share of total weekly household paid labour market hours and unpaid housework hours performed by women — at each time around the birth of the first child. These averages are plotted separately by couple type, defined based on the distribution of tasks before first childbirth. The sample includes individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18–45 (women) or 18–50 (men) at the time of birth, and were not cohabiting with children from previous relationships.

5.2 Household Specialization and Relationship Quality

Leveraging the heterogeneity in task reallocation documented in Figure 4, we study how changes in household specialization after first birth may mediate the effects on RQ. To do so, we repeat the main analysis on the impact of child birth on RQ separately for each couple type. Because the reduced sample size within each group does not allow us to carry out a dynamic analysis as in Equation 1, we adopt a static approach and estimate the following regression for each couple

type:

$$y_{i,t} = \alpha_i + \mu_t + \delta D_{i,t} + u_{i,t} \tag{2}$$

where $D_{i,t}$ is a binary variable equal to one if individual i already has a child at time t, and α_i and μ_t denote the full set of individual and time dummies.

We estimate this regression using the estimator proposed by Callaway and Sant'Anna (2021). In a static setting, the coefficient δ captures an average treatment effect on the treated across post-birth periods.[24] The results in Figure 2 indicate heterogeneity in treatment effects with time relative to childbirth, suggesting that the interpretation of δ as a single, time-invariant effect is not appropriate in our context. Due to this and since we cannot disentangle the impact of first birth from unobserved characteristics of each couple type, we do not interpret these estimates as causal.

Table E.1 reports the average pre-birth RQ levels for each couple type. All types report RQ values above the average in the full dataset. Egalitarian couples exhibit the highest levels, followed by counter-traditional ones. Unbalanced couples display the lowest levels, although they are similar to those of traditional couples. The largest gender gap in RQ is observed in traditional couples, with men reporting lower values.[25]

Table 3 reports the results from estimating Equation 2 by couple type. All the estimated coefficients are negative, meaning that RQ decreases after first child birth across couple types. However, this decline is only statistically significant for individuals in egalitarian and counter-traditional couples.[26] As seen in Figure 4, these are also the couples experiencing the largest changes in household specialization, where women start adopting primary roles in housework and men specialize in paid work after birth. Unbalanced couples exhibit the smallest decline in RQ. Although these couples experience the largest drop in the female share of paid hours, their division of housework remains largely constant, and the overall allocation of time is more balanced between partners.

There are two factors that could jointly explain these results. First, RQ might be particularly sensitive to unequal divisions of total household production time. Unbalanced couples, where women shoulder the bulk of labour and home responsibilities before birth, are the only ones distributing overall time more evenly between partners after they become parents. Despite substantial time rearrangements, the increased equality in the contribution to household production may be attenuating the associated tensions and the decrease in RQ. Second, the redistribution

[24] As explained in Appendix B, the static aggregation constitutes a weighted average of the treatment effects for different treatment cohorts at different times relative to treatment. These weights are assigned based on the number of observations in each time bin, with more populous bins receiving higher weights. Hence, without this assumption, the estimates would be skewed towards the event-times with a larger number of observations.

[25] Equation 2 assumes that the impact of first child birth on RQ is linear for all types of couples, and hence it does not depend on the pre-birth values of RQ.

[26] Note that the event-time periods with a larger number of observations are those closest to the time of birth. As seen in Figure 2, the dynamic impact of children on RQ is increasing with time since birth. Thus, these estimates should be considered a lower bound to the full impact.

Table 3: Impact of first child birth on RQ by couple type.

	(1) Traditional	(2) Unbalanced	(3) Egalitarian	(4) Counter-traditional
ATT	-0.107	-0.0992	-0.175*	-0.243**
	(0.180)	(0.086)	(0.069)	(0.075)
Observations	273	876	611	856

Notes: This table presents the results from a static difference-in-differences estimation of the effect of first child birth on RQ, using the method proposed by Callaway and Sant'Anna (2021). The sample includes individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18–45 (women) or 18–50 (men) at the time of birth, and were not cohabiting with children from previous relationships. The estimation is computed separately by couple type, as defined in Figure E.2. Standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

of housework may be the key factor triggering the deterioration of relationships, as it is less foreseeable than changes in labour market work. While labour interruptions and protections for mothers are a central concern of the most common fertility policies, changes in home production are arguably less salient. Thus, the unforeseen and unequal distribution of housework may have a dominant role in decreasing RQ.

6 Concluding Remarks

This paper investigates the causal relation between having children and RQ, a proxy of the non-material gains from being in a relationship. The arrival of children triggers a series of decisions related to time allocation in both the labour market and at home. These decisions have implications not only for individual well-being, but also for the quality of the couple's relationship. While having children can lead to increased overall happiness, it does not necessarily guarantee happiness in every aspect of one's life.

We find that having a child significantly reduces RQ, and that this effect endures over time. In other words, the level of RQ is shifted downward after the birth of a child, maintaining the pre-existing life-cycle trend. This impact on RQ is substantial and persistent compared to other life events, such as unemployment, and other well-being indicators, like general happiness.

The paper posits that the arrival of children increases the demands on couples, including childcare responsibilities and additional housework. Consistent with existing literature, we find that women largely bear the increased housework at the expense of their participation in the labour market, a consequence not typically foreseen by mothers before childbirth. We exploit heterogeneity in changes to household specialization to explore the extent to which these changes mediate the impact of child birth on RQ. The results indicate that couples undergoing the most significant changes in household specialization also experience the greatest declines in RQ.

These findings open avenues for further research on the consequences of having children. From a policy perspective, several measures addressing the impact of children on various outcomes have been explored. Policies like paternity leave, shared parental leaves, and childcare provisions have been shown to enhance female labour market outcomes, potentially by redistributing household tasks and responsibilities more equitably among couple members. Our findings suggest that such policies may mitigate the negative effects of children on RQ, an outcome closely tied to couple dissolution and the decisions surrounding investments in children, which ultimately influence their future well-being.

Moreover, this paper introduces a novel factor for consideration in the study of decisions to have children. As noted earlier, the impact on relationship quality is, to some extent, unforeseen when deciding to have the first child. Nevertheless, the observed decline in RQ and the associated utility losses, despite potential utility gains from the child itself, may become a significant factor in deciding to have subsequent children. This factor could contribute to explaining the decline in fertility observed on the intensive margin.

References

Adda, J., Dustmann, C., and Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, 125(2):293–337. 5, 13

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. 1, 14

Aguilar-Gomez, S., Arceo-Gomez, E., and De la Cruz Toledo, E. (2019). Inside the black box of child penalties: Unpaid work and household structure. *Available at SSRN 3497089*. 1, 5, 14, 20, 21

Ahammer, A., Glogowsky, U., Halla, M., and Hener, T. (2023). The parenthood penalty in mental health: Evidence from Austria and Denmark. 5, 14, 16

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753. 3, 5, 23

Akerlof, G. A. and Kranton, R. E. (2010). *Identity economics: How our identities shape our work, wages, and well-being*. Princeton University Press. 5

Albanesi, S., Olivetti, C., and Petrongolo, B. (2023). Families, labor markets, and policy. In *Handbook of the Economics of the Family*, volume 1, pages 255–326. Elsevier. 5, 14

Alon, T., Doepke, M., Olmstead-Rumsey, J., and Tertilt, M. (2020). The impact of COVID-19 on gender equality. *Covid Economics: Vetted and Real-Time Papers*, 4:62–85. 5

Amato, P. R. and Booth, A. (2001). The legacy of parents' marital discord: consequences for children's marital quality. *Journal of Personality and Social Psychology*, 81(4):627. 4

Amato, P. R. and Patterson, S. E. (2017). The intergenerational transmission of union instability in early adulthood. *Journal of Marriage and Family*, 79(3):723–738. 4

Andrew, A., Bandiera, O., Dias, M. C., and Landais, C. (2021). The careers and time use of mothers and fathers. *Institute for Fiscal Studies. Available at: https://ifs. org. uk/publications/careers-and-time-use-mothers-and-fathers*. 21

Antler, Y., Bird, D., and Fershtman, D. (2022). *Learning in the marriage market: The economics of dating*. Centre for Economic Policy Research. 4

Avdic, D. and Karimi, A. (2018). Modern family? paternity leave and marital stability. *American Economic Journal: Applied Economics*, 10(4):283–307. 4

Becker, G. S. (1991). *A treatise on the family: Enlarged edition*. Harvard University press. 23

Becker, G. S. (2010). *The economics of discrimination*. University of Chicago press. 23

Bertrand, M. (2020). Gender in the twenty-first century. In *AEA Papers and Proceedings*, volume 110, pages 1–24. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203. 5

Bertrand, M., Cortes, P., Olivetti, C., and Pan, J. (2021). Social norms, labour market opportunities, and the marriage gap between skilled and unskilled women. *The Review of Economic Studies*, 88(4):1936–1978. 75

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614. 4, 5

Bianchi, S. M. (2011). Family change and time allocation in american families. *The ANNALS of the American Academy of Political and Social Science*, 638(1):21–44. 18

Björklund, A., Ginther, D. K., and Sundström, M. (2007). Family structure and child outcomes in the USA and Sweden. *Journal of Population Economics*, 20:183–201. 5

Björklund, A. and Sundström, M. (2006). Parental separation and children's educational attainment: A siblings analysis on Swedish register data. *Economica*, 73(292):605–624. 5

Blanchflower, D. G. (2009). International evidence on well-being. In *Measuring the subjective well-being of nations: National accounts of time use and well-being*, pages 155–226. University of Chicago Press. 5, 16

Blanchflower, D. G. and Oswald, A. J. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, 88(7-8):1359–1386. 15

Blasutto, F., Kozlov, E., et al. (2020). (Changing) Marriage and cohabitation patterns in the US: Do divorce laws matter? Technical report, Job Market Papers. 4

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865. 5, 13

Bögl, S., Moshfegh, J., Persson, P., and Polyakova, M. (2024). The economics of infertility: Evidence from reproductive medicine. Technical report, National Bureau of Economic Research. 10, 17, 52

Borra, C., Browning, M., and Sevilla, A. (2021). Marriage and housework. *Oxford Economic Papers*, 73(2):479–508. 9

Borusyak, K., Jaravel, X., and Spiess, J. (2022). Revisiting event study designs: Robust and efficient estimation. *Available at SSRN 2826228*. 12

Brien, M. J., Lillard, L. A., and Stern, S. (2006). Cohabitation, marriage, and divorce in a model of match quality. *International Economic Review*, 47(2):451–494. 4

Browning, M., Chiappori, P.-A., and Weiss, Y. (2014). *Economics of the Family*. Cambridge University Press. 1, 4, 15

Bruze, G., Svarer, M., and Weiss, Y. (2015). The dynamics of marriage and divorce. *Journal of Labor Economics*, 33(1):123–170. 4

Busby, D., Christensen, C., Crane, D., and Larson, J. (1995). A Revision of the Dyadic Adjustment Scale for Use with Distressed and Nondistressed Couples: Construct Hierarchy and Multidimensional Scales. *Journal of Marital and Family Therapy*, 21:289–308. 4, 6

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230. 2, 12, 13, 14, 18, 19, 20, 21, 24, 25, 47, 48, 51, 52, 53, 54, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 77, 78

Carlson, M. J. and VanOrman, A. G. (2017). Trajectories of relationship supportiveness after childbirth: Does marriage matter? *Social Science Research*, 66:102–117. 4

Cavapozzi, D., Francesconi, M., and Nicoletti, C. (2021). The impact of gender role norms on mothers' labor supply. *Journal of Economic Behavior & Organization*, 186:113–134. 5

Charles, K. K. and Stephens, Jr, M. (2004). Job displacement, disability, and divorce. *Journal of Labor Economics*, 22(2):489–522. 16

Chiappori, P.-A. (2020). The theory and empirics of the marriage market. *Annual Review of Economics*, 12:547–578. 1, 4

Chiappori, P.-A. and Mazzocco, M. (2017). Static and intertemporal household decisions. *Journal of Economic Literature*, 55(3):985–1045. 4

Chiappori, P.-A., Radchenko, N., and Salanié, B. (2018). Divorce and the duality of marital payoff. *Review of Economics of the Household*, 16:833–858. 1, 2, 4

Chiappori, P.-A. and Weiss, Y. (2006). Divorce, remarriage, and welfare: A general equilibrium approach. *Journal of the European Economic Association*, 4(2-3):415–426. 4

Chiappori, P.-A. and Weiss, Y. (2007). Divorce, remarriage, and child support. *Journal of Labor Economics*, 25(1):37–74. 1, 4

Clark, A. E., Frijters, P., and Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, 46(1):95–144. 5, 16

Cortés, P. and Pan, J. (2023). Children and the remaining gender gaps in the labor market. *Journal of Economic Literature*, 61(4):1359–1409. 5, 13

Costa-Font, J. and Flèche, S. (2020). Child sleep and mother labour market outcomes. *Journal of Health Economics*, 69:102258. 1, 14

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996. 12

Del Boca, D., Oggero, N., Profeta, P., and Rossi, M. (2020). Women's and men's work, housework and childcare, before and during COVID-19. *Review of Economics of the Household*, 18:1001–1017. 5

Doepke, M. and Kindermann, F. (2019). Bargaining over babies: Theory, evidence, and policy implications. *American Economic Review*, 109(9):3264–3306. 1

Doiron, D. and Mendolia, S. (2012). The impact of job loss on family dissolution. *Journal of Population Economics*, 25:367–398. 16

Dolan, P., Peasgood, T., and White, M. (2008). Do we really know what makes us happy? a review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology*, 29(1):94–122. 5, 16

d'Albis, H., Gobbi, P. E., and Greulich, A. (2017). Having a second child and access to childcare: Evidence from european countries. *Journal of Demographic Economics*, 83(2):177–210. 1

Eckstein, Z., Keane, M., and Lifshitz, O. (2019). Career and family decisions: Cohorts born 1935–1975. *Econometrica*, 87(1):217–253. 4, 39, 44

Eliason, M. (2012). Lost jobs, broken marriages. *Journal of Population Economics*, 25(4):1365–1397. 16

Farré, L., Fawaz, Y., González, L., and Graves, J. (2021). How the COVID-19 lockdown affected gender inequality in paid and unpaid work in Spain. *Review of Income and Wealth*, 68(2):323–347. 5, 23

Farré, L., Felfe, C., González, L., Schneider, P., et al. (2022). Changing gender norms across generations: Evidence from a paternity leave reform. *Economics Working Paper Series*, (1812). 75

Farré, L. and González, L. (2019). Does paternity leave reduce fertility? *Journal of Public Economics*, 172:52–66. 4

Ferrer-i Carbonell, A. (2013). Happiness economics. *SERIEs*, 4(1):35–60. 5, 16

Ferrer-i Carbonell, A. and Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114(497):641–659. 7, 39

Gardner, J. (2022). Two-stage differences in differences. *arXiv preprint arXiv:2207.05943*. 12

Gemici, A. and Laufer, S. (2011). Marriage and cohabitation. *New York University, mimeo*. 4

Georges-Kot, S., Goux, D., and Maurin, E. (2024). The value of leisure synchronization. *American Economic Journal: Applied Economics*, 16(1):351–376. 18

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4):1091–1119. 5

Goldin, C. (2021). *Career and Family: Women's Century-Long Journey toward Equity*. Princeton University Press. 1, 5, 14, 20

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277. 12, 47

Goussé, M., Jacquemet, N., and Robin, J.-M. (2017). Marriage, labor supply, and home production. *Econometrica*, 85(6):1873–1919. 4

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

Gruber, J. (2004). Is making divorce easier bad for children? the long-run implications of unilateral divorce. *Journal of Labor Economics*, 22(4):799–833. 5

Hassebrauck, M. and Fehr, B. (2002). Dimensions of relationship quality. *Personal Relationships*, 9(3):253–270. 4

Hupkau, C. and Petrongolo, B. (2020). Work, care and gender during the COVID-19 crisis. *Fiscal Studies*, 41(3):623–651. 5

Ichino, A., Olsson, M., Petrongolo, B., and Thoursie, P. S. (2019). Economic incentives, home production and gender identity norms. Technical report. 3, 23

Jensen, P. and Smith, N. (1990). Unemployment and marital dissolution. *Journal of Population Economics*, 3(3):215–229. 16

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. 4

Kleven, H. (2022). The geography of child penalties and gender norms: Evidence from the United States. Technical report, National Bureau of Economic Research. 5, 21

Kleven, H., Landais, C., and Leite-Mariante, G. (2025). The child penalty atlas. *Review of Economic Studies*, 92(5):3174–3207. 5

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2019a). Child penalties across countries: Evidence and explanations. In *AEA Papers and Proceedings*, volume 109, pages 122–126. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203. 5

Kleven, H., Landais, C., and Søgaard, J. E. (2019b). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4):181–209. 5, 13, 17, 19, 20, 55, 70

Kuziemko, I., Pan, J., Shen, J., and Washington, E. (2018). The mommy effect: Do women anticipate the employment effects of motherhood? Technical report, National Bureau of Economic Research. 20, 23

Lillard, L. A. and Waite, L. J. (1993). A joint model of marital childbearing and marital disruption. *Demography*, 30(4):653–681. 5, 15

Low, C. (2024). The human capital–reproductive capital trade-off in marriage market matching. *Journal of Political Economy*, 132(2):552–576. 4, 44

Low, H., Meghir, C., Pistaferri, L., and Voena, A. (2018). Marriage, labor supply and the dynamics of the social safety net. Technical report, National Bureau of Economic Research. 4

Lundberg, S. and Pollak, R. A. (2007). The American family and family economics. *Journal of Economic Perspectives*, 21(2):3–26. 1

Lundberg, S. and Rose, E. (2000). Parenthood and the earnings of married men and women. *Labour Economics*, 7(6):689–710. 5

Lundberg, S. J. (2005). The division of labor by new parents: does child gender matter? 5

Norton, R. (1983). Measuring marital quality: A critical look at the dependent variable. *Journal of Marriage and the Family*, pages 141–151. 4

Office for National Statistics (2022). Divorces in England and Wales: 2021. 15, 60

Olivetti, C. and Petrongolo, B. (2017). The economic consequences of family policies: lessons from a century of legislation in high-income countries. *Journal of Economic Perspectives*, 31(1):205–230. 4

Petit, P. (2007). The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector. *Labour Economics*, 14(3):371–391. 23

Piketty, T. (2003). The impact of divorce on school performance: evidence from France, 1968-2002. *Available at SSRN 482863*. 5

Rodríguez-Moro, B. and Román, O. (2024). Measuring relationship quality. *Working Paper*. 2, 7, 34

Roth, J. (2024). Interpreting event-studies from recent difference-in-differences methods. *arXiv preprint arXiv:2401.12309*. 12, 51

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*. 12

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. 5

Siminski, P. and Yetsenga, R. (2022). Specialization, comparative advantage, and the sexual division of labor. *Journal of Labor Economics*, 40(4):851–887. 1, 5, 14, 20, 21

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199. 12

Svarer, M. and Verner, M. (2008). Do children stabilize relationships in denmark? *Journal of Population Economics*, 21:395–417. 5, 15

University of Essex, I. f. S. and Research, E. (2024). Understanding society: Waves 1-14, 2009-2023 and harmonised bhps: Waves 1-18, 1991-2009. DOI: http://doi.org/10.5255/UKDA-SN-6614-20. 19th Edition. 6

Voena, A. (2015). Yours, mine, and ours: Do divorce laws affect the intertemporal behavior of married couples? *American Economic Review*, 105(8):2295–2332. 4

Weiss, Y. and Willis, R. J. (1997). Match quality, new information, and marital dissolution. *Journal of Labor Economics*, 15(1, Part 2):S293–S329. 1, 4, 39, 44

A RQ measure

This appendix outlines the construction and validity of RQ. The companion paper, Rodríguez-Moro and Román (2024), provides details on these exercises, and extends the analysis of the measure to discuss its value to guide theoretical studies.

A.1 Self-Completion Partner Module

Survey text. *For each of the following questions, please indicate which best describes your relationship with your partner at the moment. Please select only one answer per question.*

How often do you have a stimulating exchange of ideas?

1	Never
2	Less than once a month
3	Once or twice a month
4	Once or twice a week
5	Once a day
6	More often

How often do you calmly discuss something?

1	Never
2	Less than once a month
3	Once or twice a month
4	Once or twice a week
5	Once a day
6	More often

How often do you work together on a project?

1	Never
2	Less than once a month
3	Once or twice a month
4	Once or twice a week
5	Once a day
6	More often

How often do you discuss or consider divorce, separation or terminating your relationship?

1	All of the time
2	Most of the time
3	More often than not
4	Occasionally
5	Rarely
6	Never

Do you ever regret that you married or lived together?

1	All of the time
2	Most of the time
3	More often than not
4	Occasionally
5	Rarely
6	Never

How often do you and your partner quarrel?

1	All of the time
2	Most of the time
3	More often than not
4	Occasionally
5	Rarely
6	Never

How often do you and your partner "get on each other's nerves"?

1	All of the time
2	Most of the time
3	More often than not
4	Occasionally
5	Rarely
6	Never

Do you kiss your partner?

1	All of the time
2	Most of the time
3	More often than not
4	Occasionally
5	Rarely
6	Never

Do you and your partner engage in outside interests together?

1	All of them
2	Most of them
3	Some of them
4	Very few of them
5	None of them

The responses below represent different degrees of happiness in your relationship. The middle point, "happy", represents the degree of happiness of most relationships. Please select the number which best describes the degree of happiness, all things considered, of your relationship.

1	Extremely unhappy
2	Fairly unhappy
3	A little unhappy
4	Happy
5	Very happy
6	Extremely happy
7	Perfect

Responding to the module. Due to the personal nature of the questionnaire, preserving the privacy of respondents is a priority of Understanding Society. The questionnaire is part of a self-completion module, meaning that it is not administered by an interviewer but completed individually, either on paper or a computer, regardless of the interview mode.[27]

Although the questionnaire is answered individually, respondents may still be concerned about others being present during the interview. Individuals are given the option to refuse to answer the self-completion part of the survey, and if they do so due to the presence of another person, this reason is recorded separately. However, this is rarely reported: in the most recent wave, no one mentioned it, and in the previous wave, only three out of 363 refusals were due to another person being present.[28] Additionally, when an interviewer is present, they record whether another person is present and whether that person influences the interview. In the majority of cases, no one else is present during the interview.

[27]Understanding Society conducts interviews in three modes: face-to-face, by phone, and online. Until wave six, all interviews were conducted either face-to-face (98%) or by phone. Web mode was introduced in wave seven, and its use has steadily increased, reaching 87% in the most recent wave.

[28]Other reasons for refusal include disliking the computer format, a child needing attention, confidentiality concerns, lack of motivation, or time constraints.

A.2 Factor Analysis

Factor analysis is a statistical dimension reduction method that seeks to disentangle latent associations between different items. The goal is to identify a set of latent variables, or factors, that explain the shared variance among a set of observed items. The variance of each item is assumed to be influenced by these common factors, as well as by a unique error term specific to each item. The factors, which are unobservable, are inferred from their effects on the items. They are identified by solving a system of equations that produces a set of weights, called factor loadings, which indicate how strongly each item is associated with each factor.

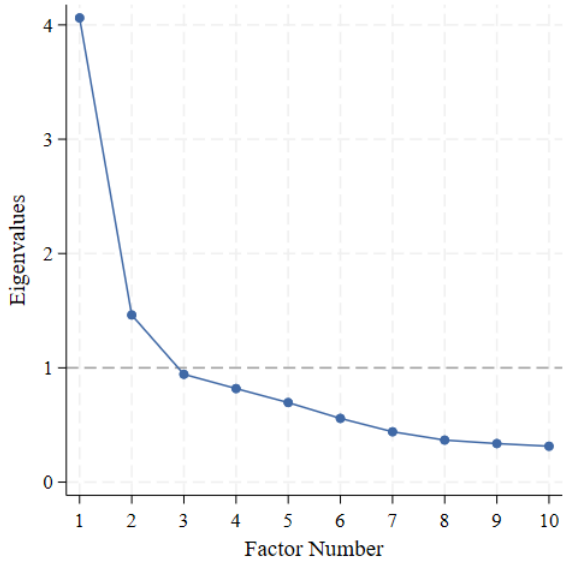
Crucially, factor analysis differs from principal component analysis (PCA), another common data reduction method, in both assumptions and interpretation. PCA aims to reduce dimensionality by finding the linear combinations of items that capture the largest possible variance, driven purely by the data and without assuming an underlying structure. In contrast, factor analysis assumes that observed variables are influenced by latent factors and explicitly models the unique variance of each variable. Its goal is to uncover the latent constructs rather than simply reducing the number of dimensions. In our case, the responses to the questionnaire items in Table 1 are the observed variables. Instead of treating relationship quality as a linear combination of the items, as PCA would, we consider that this variable is a latent factor influencing respondents' answers to each item of the questionnaire.

Before conducting the factor analysis, we recode the responses to the items in Appendix A.1 so that zero represents the least favourable couple behaviours and habits, like constantly quarrelling or never kissing one's partner, to ensure that all item responses are increasing in the quality of the relationship. We then perform factor analysis on all available responses to the ten items, and retain the first factor as the RQ measure. The decision to retain a single factor is based on the eigenvalues plotted in Figure A.1. The first factor has an eigenvalue of 4.06, more than twice that of the second next factor, indicating a substantial drop in the explained variation. All factors from the third onward have eigenvalues below 1. This first factor accounts for 40.61% of the variation in the items.

Table A.1 presents the factor loadings corresponding to RQ. All loadings are positive and greater than 0.5, indicating a strong association between RQ and each questionnaire item. The highest correlations with RQ are found in items related to having calm discussions and regretting getting married, while the lowest correlations are with items related to the frequency of kisses and to the number of joint outside interests.

In Figure A.2, we plot the factor loadings derived from conducting the factor analysis on the items in Table 1 separately for men and women. The factor loadings for men are generally lower than those of the overall RQ measure, indicating weaker correlations between the extracted factor and the individual items. Similarly, the eigenvalue for the male sample (3.83) is lower than that of RQ in the full sample. In contrast, the factor loadings for women are

Figure A.1: Eigenvalues from factor analysis.



Notes: This graph presents the eigenvalues of all factors derived from performing a factor analysis on the responses to the ten questions in the Understanding Society Partner module. The eigenvalues indicate the amount of variation in the sample explained by each factor. The first factor from this analysis is the RQ measure.

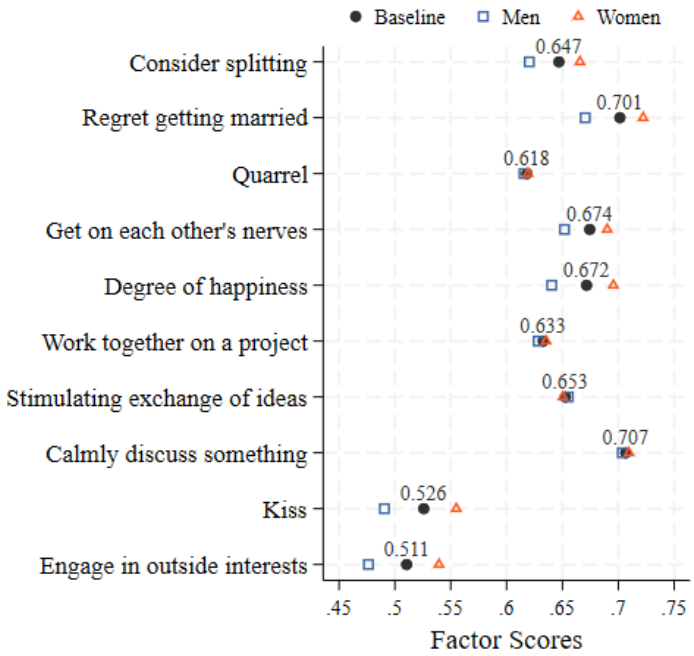
Table A.1: Factor loadings of RQ.

(a) Subjective assessment		(b) Couple time use	
How often do you... ?		*How often do you... ?*	
consider splitting	0.647	work together on a project	0.633
regret getting married	0.701	stimulating exchange of ideas	0.653
quarrel	0.618	calmly discuss something	0.707
get on each others nerves	0.674	kiss partner	0.526
What is the... ?		*Do you and your partner... ?*	
degree of happiness w/ relationship	0.672	engage in outside interests	0.511

Notes: This table reports the factor loadings used to construct the measure of relationship quality (RQ). The measure is based on responses to the ten questions in the Understanding Society Partner module, a self-completion questionnaire administered in waves 1, 5, 7, 9, 11 and 13, covering 2009-2022. We perform a factor analysis on all available responses and retain the first factor as the RQ measure. The reported factor loadings correspond to this first factor, which has an eigenvalue of 4.06 and explains 40.61% of the variation in the data. Panel (a) presents items on subjective assessments of the relationship, and panel (b) contains items about how couples spend their time together.

generally higher than those of RQ, with a correspondingly larger eigenvalue. Despite these differences, the ranking of items remains largely consistent across genders, with women assigning higher loadings and men consistently lower ones. The only exception is the item related to the frequency of stimulating exchanges of ideas.

Figure A.2: Factor loadings of RQ on men and women samples separately.



Notes: This graph displays the factor loadings for the relationship quality (RQ) measure, along with the factor loadings from factor analyses conducted separately for men and women. In all cases, the factor analysis is based on responses to the ten questions in the Understanding Society Partner module. The black circles represent the factor loadings of RQ, obtained from the full sample, with an eigenvalue of 4.06. The blue squares correspond to the factor loadings estimated using only male observations, yielding an eigenvalue of 3.83. The orange triangles represent factor loadings from the female-only sample, with an eigenvalue of 4.24.

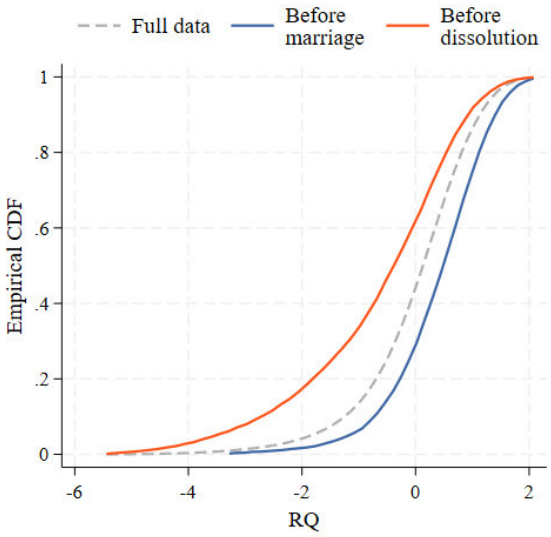
A.3 Validity

To verify that the constructed measure provides sensible information about cohabiting partners' relationships, we conduct a number of validity tests. We follow the life satisfaction literature (Ferrer-i Carbonell and Frijters, 2004), which highlights the importance of two fundamental theoretical assumptions that empirical measures of satisfaction should plausibly satisfy: informativeness and interpersonal comparability. Additionally, we verify that RQ evolves smoothly overtime and does not experience abrupt jumps, and that it correlates with observed characteristics of individuals and partners in expected ways (e.g., Weiss and Willis, 1997; Eckstein et al., 2019).

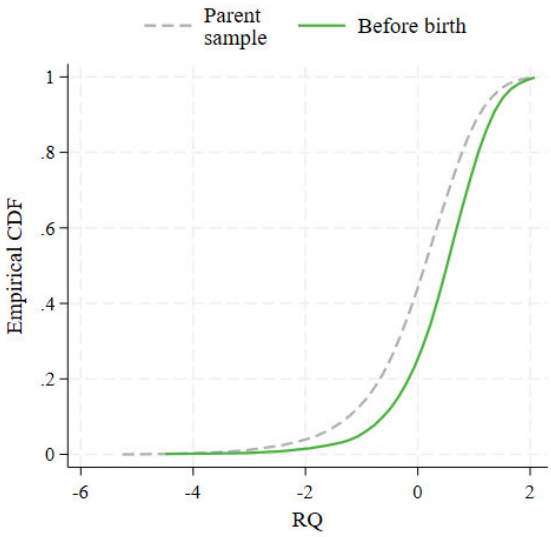
Informativeness. First, we verify that RQ provides meaningful information about the status of a relationship. To do so, we assess the predictive capacity of RQ for couple decisions: (a) marriage and separation and (b) fertility decisions. Marriage increments separation costs, serving as a commitment mechanism. Hence, couples transitioning into marriage are expected to report relatively higher levels of RQ. In contrast, separations typically result from poor-

Figure A.3: Informativeness: behaviour prediction.

(a) Marital transitions



(b) First child birth



Notes: This figure displays the empirical cumulative distribution function (CDF) of RQ for different samples. Panel (a) compares the overall distribution of RQ in the full data (dashed gray line) with its distribution one period before marriage (blue) and one period before couple dissolution (red). Panel (b) plots the overall distribution of RQ in the analysis sample of parents (dashed gray line) and the distribution of the measure in this sample one period before first child birth (green).

quality relationships. Thus, we expect lower than average RQ levels on couples at the brink of dissolution. Finally, we hypothesise that couples deciding to have a child exhibit higher RQ levels.

To assess the predictive power of RQ for couple transitions, we compare its distribution the period before these events with the overall distribution of RQ. Figure A.3 presents the empirical cumulative distribution function (CDF) of RQ for different samples. Panel (a) compares the overall distribution of RQ with its distribution one period before marriage and one period before dissolution. The pre-marriage distribution of RQ is shifted to the right, indicating higher RQ values, whereas the pre-dissolution distribution is significantly shifted to the left, reflecting lower RQ levels before separation across the entire distribution. Notably, the deviation from the overall distribution is larger before dissolution than before marriage, suggesting that negative deviations in RQ have a stronger impact on marital decisions than increases do. Panel (b) plots the overall distribution of RQ in the analysis sample of parents and the distribution of the measure in this sample one period before first child birth, at the time of conception. This distribution is slightly shifted to the right in comparison to the benchmark.

To formally test for differences between these distributions, we conduct a two-sample Kolmogorov-Smirnov equality-of-distributions test. This test assesses whether two samples are drawn from the same population by computing the distance between their empirical CDFs. Table A.2 reports the D-statistics and p-values for the samples considered, where the null hy-

pothesis is that both samples are drawn from the same distribution. We find that the pre-divorce and pre-marital samples contain significantly lower and higher RQ values than the full data, respectively. Similarly, the pre-child sample exhibits significantly higher values than the complete sample of first-time parents. Overall, the test results confirm that all three samples differ significantly from their respective benchmarks.

Table A.2: Two-sample Kolmogorov-Smirnov tests on RQ preceding marital decisions.

	d_0 = Full data		d_0 = Parent sample
	d_1 = Before separation	d_1 = Before marriage	d_1 = Before first child
$d_0 > d_1$	0.000	0.1741	0.2491
	(1.000)	(0.000)	(0.000)
$d_0 < d_1$	-0.2192	0.0000	-0.0000
	(0.000)	(1.000)	(1.000)
Combined	0.2192	0.1741	0.2489
	(0.000)	(0.000)	(0.000)

Notes: This table displays the results of two-sample Kolmogorov-Smirnov tests on different samples. The first two columns compare the distribution of RQ in the full data with the distribution the period before couple dissolution and marriage. The third column compares the overall distribution of the sample of parents with the distribution before first child birth within this sample. The null hypothesis is that the two samples compared are drawn from the same distribution. The reported coefficients are the resulting D-statistics and p-values, in parentheses.

We further study the correlation between RQ and couple dissolution, regressing a binary variable indicating whether the period precedes a separation on RQ. We perform this analysis separately in the full data and in the main analysis sample of parents. Controls include age, sex, college education, employment status, log monthly income, presence of children, relationship tenure, marital status, area of residence, and period. The estimated coefficients are reported in Table A.3. All else equal, a one standard deviation higher RQ is associated with approximately a 0.7 percentage point lower probability of separation.

The periods preceding 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 finding is consistent with the theoretical assumption of informativeness.

Interpersonal comparability. Second, RQ should exhibit some degree of commonality across individuals, meaning that partners' perceptions of relationship quality should be correlated. We assess this by examining the correlation between RQ levels within couples. Table A.4 reports the results from regressing women's RQ on their male partners' RQ. Men's RQ is a highly significant predictor of women's RQ and has the largest coefficient in magnitude, nearly five times larger than the second-largest predictor, being married.

Table A.3: Correlation between RQ and couple dissolution.

	Full data		Parent sample	
	(1)	(2)	(3)	(4)
RQ	-0.00931***	-0.00862***	-0.00833***	-0.00701*
	(0.001)	(0.001)	(0.002)	(0.003)
Controls	✓	✓	✓	✓
Individual FE		✓		✓
R-squared	0.037	0.024	0.054	0.050
Observations	106826	106826	15563	15563

Notes: This table presents the results from regressing a binary variable indicating whether the period precedes couple dissolution on RQ. The coefficients in columns (1) and (2) are estimated using the full data, while columns (3) and (4) present estimation results from the analysis sample of first-time parents. Controls include age, sex, college education, employment status, log monthly income, presence of children, relationship tenure, marital status, area of residence and period. Standard errors, clustered at the couple level, are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.4: Regression of woman RQ on man RQ.

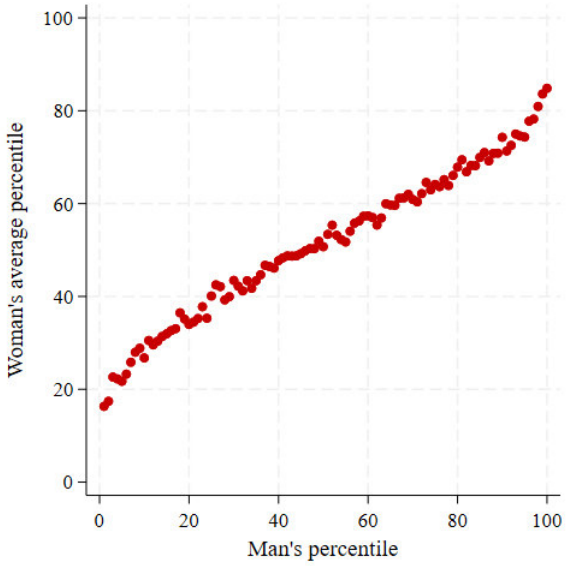
	Woman RQ		
	(1)	(2)	(3)
Man RQ	0.613***	0.603***	0.594***
	(0.007)	(0.007)	(0.008)
Age × Tenure × Wave		✓	✓
Controls			✓
R-squared	0.320	0.330	0.334
Observations	42889	42889	39525

Notes: This table presents the results from regressing women's RQ on their (male) partners' RQ. Column (2) includes age, relationship tenure and survey wave dummies. Column (3) additionally controls for age, sex, college education, employment status, log monthly income, presence of children, relationship tenure, marital status, area of residence and period. Standard errors, clustered at the couple level, are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We study the non-linear relation between partners' RQ using a rank-rank plot. Figure A.4 displays the average RQ percentile rank of women per their male partners' percentile rank. While the correlation is not perfect, which would produce a 45-degree line, there is a clear positive relation between the two. The slope is steepest at the top and bottom percentiles, approaching one, but flattens by nearly half at the centre of the distribution. This suggests that extreme assessments of the quality of the relationship are more strongly shared between partners than intermediate ones.

Smooth evolution. We next verify that the measure evolves smoothly over time and does not experience abrupt jumps. If systematic jumps occurred at specific ages or relationship tenures, the effects captured in the main analysis could be driven by the passage of time rather than childbirth. To examine this, we analyse how RQ progresses with age and relationship tenure.

Figure A.4: Rank-rank correlation of RQ between couple members.



Notes: This figure presents the rank-rank correlation of RQ between couple members in the full data. We plot the average RQ percentile rank of women against their male partners' RQ percentile rank. Each dot represents a percentile.

We estimate the following regression:

$$y_{i,t} = \alpha_i + \mu_t + \sum_a \mathbb{1}\{a = \text{age}_{i,t}\}\gamma_a + \sum_d \mathbb{1}\{d = \text{tenure}_{i,t}\}\lambda_d + u_{i,t} \tag{3}$$

where $y_{i,t}$ denotes RQ of individual i at time t. We include full sets of age ($\mathbb{1}\{a = \text{age}_{i,t}\}$) and relationship tenure ($\mathbb{1}\{d = \text{tenure}_{i,t}\}$) dummies to trace the evolution of RQ along these two time dimensions. We use age 25 and tenure 1 as the reference categories. Additionally, we include individual (α_i) and period (μ_t) dummies to eliminate unobserved individual heterogeneity and isolate within-individual changes in RQ. By including both age and tenure non-parametrically, the estimated coefficients provide the respective age and tenure profiles of RQ.

Figure A.5 presents the results from this estimation. Panel (a) plots the coefficient estimates of γ_a, representing the age profile of RQ, for ages ranging from 25 to 70, using age 25 as the baseline. Each point represents the difference in RQ between individuals aged 25 with tenure 1 and individuals of other ages with tenure 1. While these results are purely observational, ageing is positively associated with RQ. Additional years of age correspond to progressively higher RQ levels. These increments are highly smooth and almost linear. An increase of approximately 15 years in age is associated with an increase of 1 standard deviation in RQ.

Panel (b) presents the relationship tenure profile of RQ by plotting the coefficient estimates of λ_d, using new relationships as the baseline. Each point reflects the difference in RQ between individuals aged 25 with tenure 1 and individuals of other tenures at age 25. RQ declines

sharply with tenure during the first ten to fifteen years of a relationship, although this decrease is slower than the increase associated with age. After this initial period, the decrease in RQ stabilizes for longer relationships. As with age, additional years of tenure reduce RQ smoothly, without abrupt jumps.

Figure A.5: Age and tenure profiles of RQ.

Notes: This figure plots the (a) age and (b) relationship tenure profiles of RQ. We estimate Equation 3 on the full data. Panel (a) presents the coefficient estimates of γ_a for ages a ranging from 25 to 70, using age 25 as the baseline. Each point represents the difference in RQ between individuals aged 25 with tenure 1 and individuals of other ages with tenure 1. Panel (b) displays the coefficient estimates of λ_d for relationship tenures d from 1 to 35, using one-year-old relationships as the baseline. Each point reflects the difference in RQ between individuals aged 25 with tenure 1 and individuals of other tenures at age 25.

Observable characteristics. Past literature has linked match quality to observable characteristics of individuals and couples (Weiss and Willis, 1997; Eckstein et al., 2019; Low, 2024). To investigate how RQ relates to these traits, we estimate the following regression:

$$y_{i,t} = \mathbf{X}_{i,t}\beta + \mu_t + \sum_a \mathbb{1}\{a = \text{age}_{i,t}\}\gamma_a + \sum_d \mathbb{1}\{d = \text{tenure}_{i,t}\}\lambda_d + u_{i,t} \tag{4}$$

As in Equation 3, $y_{i,t}$ denotes RQ for individual i at wave t, $\mathbb{1}\{a = \text{age}_{i,t}\}$ and $\mathbb{1}\{d = \text{tenure}_{i,t}\}$ represent age and tenure dummies, respectively, and μ_t captures period effects. $\mathbf{X}_{i,t}$ is a vector of individual and couple characteristics. Individual characteristics include sex, college education, employment status, and log personal monthly income of both the individual and their partner. Couple characteristics include relationship marital status, presence of children, urban residence, and the female shares of both labour market work and housework, as defined in Subsection 2.2. Standard errors are clustered at the couple level.

Table A.5 columns (1) and (3) present the results for the full data and the parent sample, respectively. The magnitudes and signs of the coefficients remain stable across samples. Women consistently report lower RQ than men. College education, both own and partner's, is positively associated with RQ. Notably, couple characteristics show the strongest associations. Marriage is linked to higher RQ, whereas having children is associated with lower RQ. Additionally, a higher female share of housework correlates with lower RQ.

While these estimates capture level differences in RQ across groups, unobserved individual heterogeneity may drive some of these correlations. To address this, we re-estimate the model including individual fixed effects, which remove time-invariant unobserved heterogeneity and provide within-individual estimates. The results, shown in Table A.5 columns (2) and (4), indicate that individual characteristics lose significance, suggesting that compositional differences drive the initial correlations. However, the coefficients of marriage and children remain significant, although marriage is only marginally so, implying that RQ tends to increase after marriage and decline with the arrival of children.

Table A.5: Regression of RQ on individual, couple and partner characteristics.

	Full data		Parent sample	
	(1)	(2)	(3)	(4)
Female	-0.100***		-0.104***	
	(0.009)		(0.023)	
College	0.068***	-0.017	0.070*	-0.097
	(0.014)	(0.043)	(0.030)	(0.065)
Employed	0.038*	0.024	0.100	0.014
	(0.018)	(0.014)	(0.051)	(0.040)
Log Personal Income	0.017**	-0.002	0.014	0.006
	(0.006)	(0.004)	(0.013)	(0.010)
Married	0.256***	0.087*	0.333***	0.133*
	(0.029)	(0.035)	(0.060)	(0.057)
At least one child	-0.271***	-0.055**	-0.265***	-0.100**
	(0.021)	(0.019)	(0.043)	(0.034)
Urban	-0.057**	-0.009	-0.078*	-0.060
	(0.018)	(0.032)	(0.040)	(0.062)
Female share: labour market	-0.044	-0.036	-0.068	-0.029
	(0.024)	(0.026)	(0.075)	(0.058)
Female share: housework	-0.105**	-0.012	-0.183*	0.004
	(0.032)	(0.025)	(0.072)	(0.061)
Partner college	0.083***	0.050	0.071*	0.042
	(0.014)	(0.042)	(0.029)	(0.065)
Partner employed	0.007	-0.005	0.035	-0.070
	(0.016)	(0.013)	(0.048)	(0.042)
Partner income	0.017**	0.000	0.027*	-0.004
	(0.006)	(0.004)	(0.014)	(0.009)
Constant	0.106	-0.262	0.100	-0.243
	(0.118)	(0.541)	(0.263)	(0.776)
Age × Tenure × Wave	✓	✓	✓	✓
Individual FE		✓		✓
R-squared	0.047	0.053	0.078	0.110
Observations	54011	54011	10494	10494

Notes: This table presents the results from regressing RQ on a set of individual, couple and partner characteristics (Equation 4). Columns (1) and (2) show the results for the full data, and columns (3) and (4) display the results in the main analysis sample of parents. Individual and partner characteristics include sex, college education, employment and log monthly personal income. Couple characteristics comprise marital status, presence of children, urban residence, and the female shares of labour market work and in housework, as defined in Subsection 2.2. We control for age, relationship tenure and period non-parametrically. Columns (1) and (3) report pooled OLS estimates, while columns (2) and (4) present fixed effects results. Standard errors, clustered at the couple level, are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B Empirical Strategy

We take a dynamic difference-in-differences (DiD) approach to study the causal impact of children on RQ. This methodology exploits sharp changes in individual outcomes after the birth of their first child and allows for treatment heterogeneity with time relative to birth. In essence, it allows us to estimate the impact of having a child at each period relative to birth. Furthermore, it exploits differences in the timing of child birth to dynamically construct a control group.

In the main analysis, we use a sample of individuals who become parents to study the impact of the first birth (treatment) on RQ (outcome). This has two important implications. First, the treatment is staggered because different individuals have their first child in different periods. This divides the sample into different treatment adoption cohorts, depending on the calendar year when individuals become parents. Second, having a sample of new parents implies that everyone in the sample is treated at some point. This means that, in each period, we compare individuals who just had a child with individuals who have not become parents yet and with individuals who have been parents for more than one period.

The usual approach is to estimate the dynamic two-way fixed effects (TWFE) regression in Equation 1. The econometrics literature has pointed out several problems derived from estimating such specifications through OLS in contexts of staggered treatment adoption (see Goodman-Bacon, 2021, for a review). There are two main issues that threaten the causal interpretation of the obtained estimates. First, it performs forbidden comparisons between individuals changing status from control to treated and cohorts that have already been treated for more than one period. Using already-treated as controls mechanically generates negative or non-convex weights. Second, although such weights exist even under homogeneous treatment effects, heterogeneity across treatment cohorts makes them bias-inducing, as the estimator combines effects from different periods in a non-representative way. As we explain below, we take an alternative approach in our estimation.

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

The method proposed by Callaway and Sant'Anna (2021) overcomes the issues related to TWFE estimation by clearly separating identification, estimation and aggregation. This method is applicable in settings where a panel dataset is available and the binary and absorbing treatment is adopted in a staggered manner. We feel comfortable assuming that having a child is a binary (having a child or not) and absorbing treatment (having a child forever). The only situation where this treatment would not be absorbing is in the case of child death, which is a very rare occurrence in the data. Additionally, the treatment is adopted in a staggered manner, since individuals have children in different periods, i.e., there are different treatment adoption

cohorts.

Identification. The building blocks of this method are the average treatment effects on the treated (ATT) for each treatment adoption cohort g and at each period t, denoted as ATT(g,t). Identifying these 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. Note that "limited" anticipation implies that the method allows for some preceding reaction, but an assumption needs to be made on the amount of anticipation periods before treatment.

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. Importantly, the method allows for the looser assumption of parallel trends conditional on covariates.

Those two assumptions allow to construct the counterfactual for each ATT(g,t) using (i) the period before treatment as the baseline period, and (ii) all treatment cohorts that have not been treated by t as controls. Therefore, the control group at each t for the same cohort g varies because at each subsequent period new cohorts enter treatment status.

Estimation. Each ATT(g,t) is estimated as a 2×2 difference-in-differences coefficient using the baseline period and the control group described. The estimation can be done in three ways, which use information from different parts of the data generating process: using outcome regression, inverse probability weighting, or doubly robust estimands. Refer to Callaway and Sant'Anna (2021) for detailed information on the estimation methods. We use the last method, since it constitutes the combination of the other two.

Aggregation. The ATT(g,t) estimates are the building blocks used to summarize the treatment effects across treatment adoption cohorts, periods or time relative to treatment. In each case, aggregation involves carefully chosen and estimable weights for the different ATT(g,t). Our main aggregation involves obtaining the ATT at each event-time. Crucially, because each event-time estimate is a weighted average of the cohorts present at that event-time, it mechanically reflects the cohort composition prevailing in that period. This may prevent the causal interpretation of the aggregated coefficients, unless homogeneity across treatment adoption cohorts is assumed. In this scenario, we need to assume that the impact of first birth is the same regardless of the year of birth.

B.2 Identification

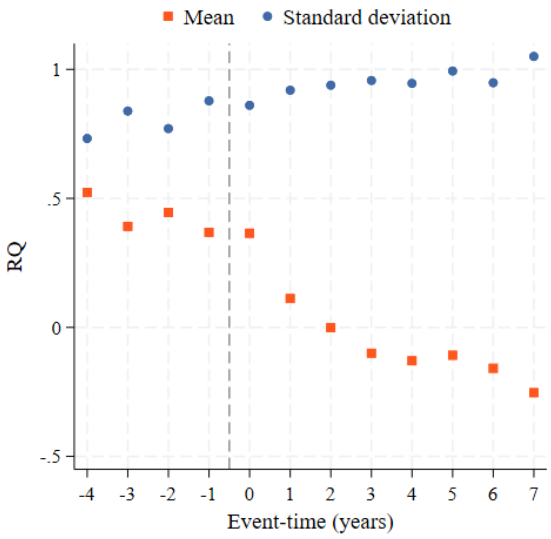
Although we cannot directly test whether the identification assumptions above are satisfied, we can provide evidence in favour of their plausibility.

The **limited anticipation** assumption would be violated if individuals decided to have children in response to either negative or positive shocks to RQ. Figure 2 shows no evidence that the birth of a first child may be preceded by deviations in RQ, on average, given that 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.

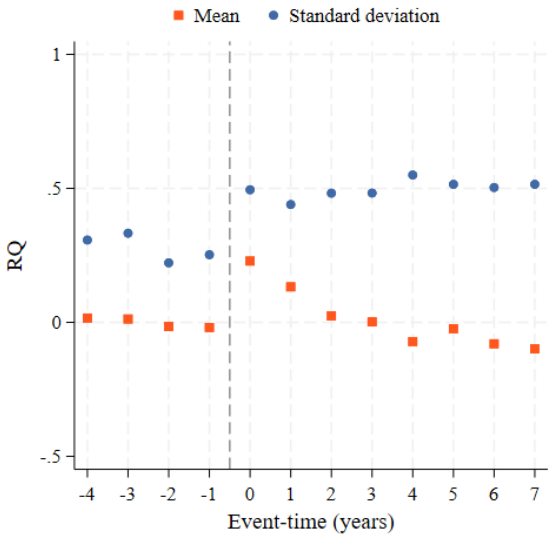
Figure B.1 (a) plots the mean and standard deviation of RQ at each event time. Consistent with Figure 2, average RQ remains stable before childbirth and declines sharply afterward. The standard deviation is lowest before childbirth, supporting the absence of anticipation, and steadily increases post-birth. Additionally, there are no significant deviations from individual RQ trajectories before birth that could predict its timing. Figure B.1 (b) shows the mean and the standard deviation of deviations from individual-specific RQ means, computed separately for pre- and post-birth periods. Before childbirth, deviations are nearly zero on average, indicating no systematic pre-birth RQ changes. The low standard deviation further suggests minimal heterogeneity in pre-birth RQ changes.

Figure B.1: Evidence on no anticipation assumption.

(a) RQ levels



(b) Deviations from individual means



Notes: This figure presents evidence supporting the plausibility of the no anticipation assumption. Panel (a) displays the mean and standard deviation of RQ at each period around first child birth in the main sample of parents. Panel (b) presents the mean and standard deviation of deviations from individual-specific RQ means, computed separately for pre- and post-birth periods.

In Table B.1, we examine which observable characteristics predict the timing of first child birth. The outcome is a dummy variable equal to one if the individual has their first child in the subsequent period. The specification includes RQ, dummies for college education, labour market activity and employment, log monthly labour income, weekly hours worked in the labour market, a marriage indicator and period dummies. We also include individual fixed effects, so the estimates reflect within-individual variation and indicate how changes in characteristics are associated with the timing of birth. The analysis is restricted to the period before first birth. The results show that changes in RQ prior to birth do not predict the timing of childbirth, which provides additional support for the no-anticipation assumption. Among the included covariates, only being married emerges as a significant predictor of first birth.

Table B.1: Observable characteristics and timing of birth.

	(1)
RQ	-0.0169
	(0.028)
College educated	-0.113
	(0.124)
Active in labour mkt	0.0317
	(0.123)
Employed	0.0705
	(0.112)
Log monthly labour income	-0.000398
	(0.038)
Weekly work hours	-0.00128
	(0.002)
Married	0.129**
	(0.043)
Individual FE	✓
Period FE	✓
R-squared	0.624
Observations	1116

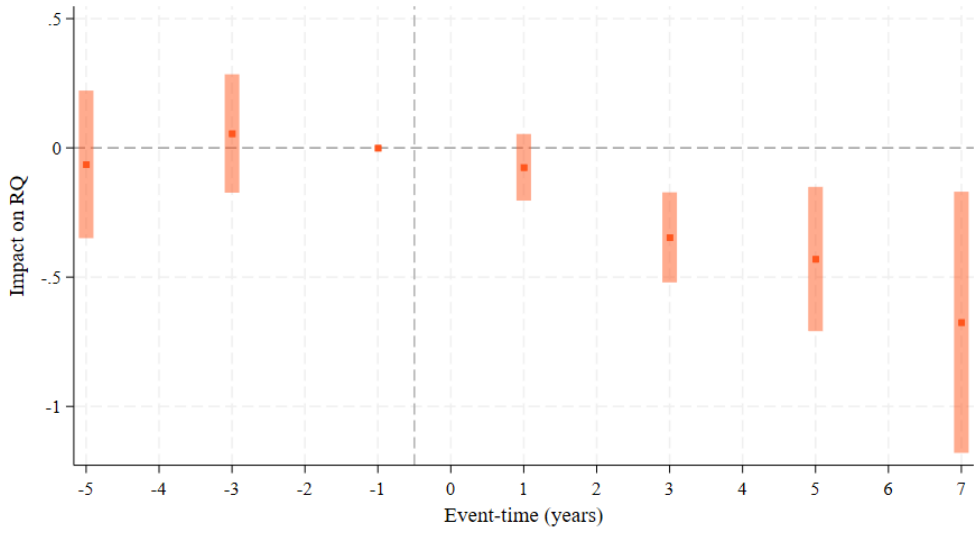
Notes: This table presents the results from regressing an indicator variable equal to one the individual's first child is born in the subsequent period (and 0 otherwise) on a set of observable characteristics. Regressors include RQ, dummies for college education, labour market activity and employment, log monthly labour income, weekly hours worked in the labour market, a marriage indicator, as well as period and individual fixed effects. The regression is estimated on pre-birth observations. Standard errors are reported in parentheses.

Conditional parallel trends cannot be directly tested, since it is an assumption about unrealized counterfactual scenarios. However, the flat pre-trends in Figure 2 provide suggestive evidence in its favour. We formally test whether the coefficients for all pre-treatment peri-

ods are statistically equal to zero and fail to reject this hypothesis. By default, Callaway and Sant'Anna (2021) conducts "short" pre-birth comparisons, where consecutive periods are compared, whereas post-birth comparisons are "long", or relative to the baseline period. To address this asymmetry, Roth (2024) suggests using "long" comparisons for pre-birth periods as well, where all pre-birth periods are compared to a single baseline. While this does not affect post-birth coefficient estimates (as long as the baseline period is kept fixed), it alters pre-birth comparisons.

Implementing this is not straightforward in our context, given that RQ is never observed in two consecutive periods. To overcome this, we normalize the last observed period before birth as the baseline, pooling event-time periods $t - G_i = -2$ and $t - G_i = -1$. The remaining pre-birth periods are then redefined relative to this baseline. Figure B.2 presents the results of this approach. The slight differences in post-birth coefficients stem from this redefinition of the baseline period. Importantly, pre-birth estimates remain close to zero.

Figure B.2: Impact of first child birth on RQ using "long" pre-birth comparisons.



Notes: This figure presents the results from a dynamic difference-in-differences estimation of the effect of first child birth on RQ, using the method proposed by Callaway and Sant'Anna (2021). The estimation conducts "long" comparisons for both pre- and post-birth periods, where all periods are compared to the baseline period. The baseline is defined as the last period when RQ is observed before first birth. Given the structure of the data, this may be one or two years prior to the event. All other periods are defined relative to this baseline, with gaps of one period due to RQ being observed only every other survey collection wave. We plot 95% confidence intervals.

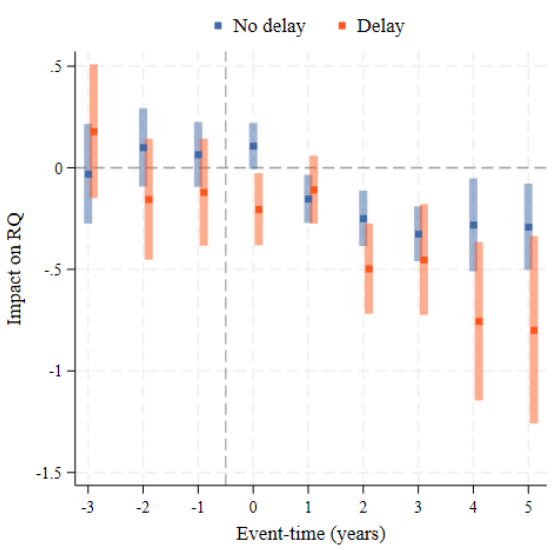
A potential limitation of this approach is that we observe only realized fertility, not intended fertility. To address this, we separately examine individuals who experience a delay in the birth of their first child, identified as those who report a miscarriage or the use of fertility treatments.[29] While we cannot precisely determine when these individuals initially intended to

[29]Fertility treatment information is available only if the treatment was successful and led to a pregnancy since the previous interview. In our sample, approximately 9% of each treatment cohort experiences a delay in fertility.

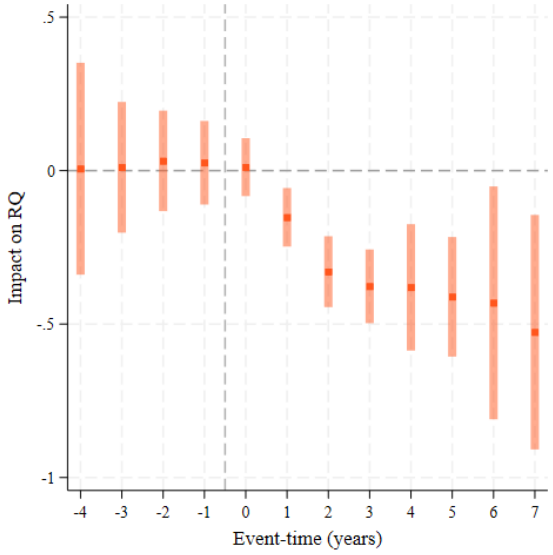
have a child, we do observe that they faced a delay.[30] Figure B.3 (a) presents separate estimation results for individuals who experience a delay in fertility and those who do not. The impact differs mainly on the birth period: parents without delays experience a slight increase in RQ, while those with delays show a decline. Although from event-time one onwards point estimates are similar, estimates for delayed parents are less precise. Importantly, before childbirth, there are no significant differences in RQ between those who experience delays, regardless of when they eventually have a child.

Figure B.3: Evidence on parallel trends assumption.

(a) Delayed pregnancy



(b) Similar baseline RQ



Notes: This graph provides evidence supporting the parallel trends assumption. Panel (a) presents the results from a dynamic difference-in-differences estimation of the effect of first child birth on RQ using the method proposed by Callaway and Sant'Anna (2021), separately for individuals who experienced a delay in childbirth and those who did not. We identify individuals who reported using fertility treatments or experiencing miscarriages before birth as subject to delayed fertility. Panel (b) replicates the main dynamic difference-in-differences estimation, restricting comparisons to individuals within the same RQ quintile before birth. We plot 95% confidence intervals.

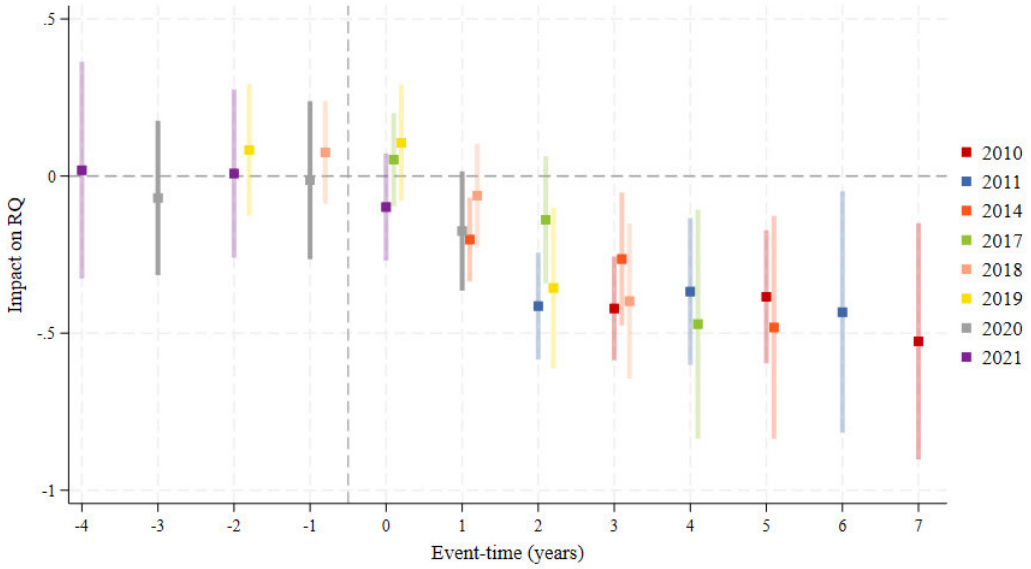
Another way to assess the plausibility of parallel trends is to replicate our main analysis restricting comparisons to individuals with similar RQ levels before the birth of their first child but who had children in different periods. This specification makes the assumption more plausible. Specifically, we compare individuals within the same RQ quintile prior to birth. The results, shown in Figure B.3 (b), closely align with our main findings, further reinforcing the plausibility of the parallel trends assumption.

Although not an identifying assumption, a final condition required for the time aggregation in our main results is the **homogeneity of treatment effects across cohorts**. Figure B.4 presents the dynamic impact of first childbirth separately for each treatment cohort. Due to

[30] On average, couples try to conceive for about a year before turning to assisted reproduction technologies (Bögl et al., 2024).

sample size limitations, the estimates at this level of disaggregation are somewhat noisy. However, the point estimates remain similar across cohorts.

Figure B.4: Impact by treatment cohort.



Notes: This graph presents the results from a dynamic difference-in-differences estimation of the effect of first child birth on RQ using the method proposed by Callaway and Sant'Anna (2021), separately for each treatment cohort. A treatment cohort is defined by the calendar year when an individual's first child was born. We plot 95% confidence intervals.

B.3 Alternative Counterfactuals

Our main analysis sample is formed by individuals observed becoming parents at some point within our observation window. Coefficients are estimated by carrying out comparisons between individuals who become parents in a given period with individuals who have a child at a later date. Using this sample we focus on individuals who are exposed to the transition into parenthood and isolate its impact without conflating it with differences between parents and non-parents, who may differ systematically in unobservable ways. Moreover, restricting the sample to individuals who eventually become parents ensures that our estimates are directly relevant to study the consequences of family-oriented policies, which are designed to support individuals during and after the transition to parenthood.

In this section, we assess the robustness of the results obtained using the parent sample by considering two alternative counterfactuals or control groups: individuals who want to have children but do not manage to (*infertile individuals*), and individuals who do not want nor have children (*never parents*).

B.3.1 Infertile Individuals

First, we consider individuals who want to have children but do not manage to. This counterfactual could approximate how RQ would evolve in the case where individuals in our treatment sample did not manage to conceive. The causal interpretation of the estimates obtained using this control group relies on the assumption that the individuals identified as infertile are comparable to the treated sample.

Ideally, we would have data on fertility intentions, which could either be given by given by self-declared intentions or by reported usage of fertility treatments. Unfortunately, information on fertility is only available when a couple manages to conceive. We identify *infertile* individuals as those who manage to become pregnant but who do not have a live birth by the end of the observation window. While we are aware that this is only a subsample of the population of infertile individuals, since some couples may never conceive, they are the only cases which we are able to identify.

The control sample of infertile individuals is formed by women having a pregnancy between ages 18 and 45 who never become mothers, and their partners. We observe around 280 such women, resulting in a sample of 2,615 observations, including their partners. Within this sample, the average number of failed pregnancies is 1.239, where about 82% of individuals have a single failed pregnancy. Out of the 1,760 individuals in our main analysis sample, only 93 report having had a failed pregnancy before first child birth, and the average number of failed pregnancies is 0.185 in the entire observation window. Thus, we still focus on first birth in the treatment group, and we use time of first pregnancy in the control group.

Table B.2 presents summary statistics for the main treatment sample in column (1) and the control sample of infertile individuals in column (2), during the period before first birth in the case of parents and before first pregnancy in the case of infertile individuals. The control group is formed by older and more educated individuals. They are more often active in the labour market and employed, and they also earn more and work more hours. Finally, infertile couples are in longer relationships and likelier to be married. This could indicate that women in this sample are more career-oriented and attempt motherhood at later ages, experiencing more difficulties.

We repeat the main analysis using the method proposed by Callaway and Sant'Anna (2021) and infertile individuals as the control group. We exploit the information on the year of individuals' first failed pregnancy as a proxy for fertility intentions. We acknowledge that this is an imperfect proxy, since individuals may have unsuccessfully tried conceiving for before first pregnancy.

The causal interpretation of the results relies on the usual assumptions of no anticipation and conditional parallel trends. The interpretation of no anticipation remains largely the same as in the main analysis: the time when individuals decide to get pregnant cannot be predicted by

changes in RQ prior to pregnancy. To exploit fertility intentions, the parallel trends assumption is now conditional on the year when the pregnancy occurs. In the treatment group pregnancy results in live birth, whereas in the control group it ends without child. Thus, the underlying assumption is that, if the pregnancy had not been successful for treated individuals, their RQ would have evolved like that of infertile individuals.

Figure B.5 displays the distribution of RQ for parents in the treatment group and infertile individuals in the control group separately (a) before and (b) after first birth/pregnancy. The distribution plotted in panel (a) indicates that the infertile sample has lower RQ ex ante. This could be due to prior difficulties to get pregnant, inducing RQ losses. Panel (b) indicates that the average RQ of infertile individuals also drops after the failed pregnancy. However, this drop is smaller than the drop documented for parents.

The results from estimating the effect of first child birth on RQ using the infertile sample as a control group and conditioning on the calendar year of the first pregnancy are presented in Figure B.6. The documented dynamics are similar to the main results presented in Figure 2. RQ decreases over the first four years after childbirth, but the magnitude of the coefficients is halved using this control group. This is due to the fact that individuals in the control group undergo a decrease in RQ after a failed pregnancy. The persistence of the effect is also reduced in comparison to the main result, experiencing a slight increase six years post-birth.

B.3.2 Never Parents

The second control group we consider are individuals who do not wish to become parents and never become parents. We define *never parents* as those individuals never observed becoming pregnant nor cohabiting with their children. We restrict the sample to respondents observed at least once during fertility ages, between 18 and 45. Since we do not observe fertility intentions until pregnancy, this group could contain individuals who want children but never manage to become pregnant, which we cannot distinguish.

We follow Kleven et al. (2019b) and use a design based on assigning placebo births to never parents. There are two types of individuals never observed becoming parents: those who have not had children by the end of their fertility cycle, and those who do not have children yet, but may have them before their fertile ages have passed, at age 40. We refer to cohorts 1951-1980 as old, non-truncated cohorts, and to cohorts 1981-2001 as young, truncated cohorts. This is a crucial distinction when assigning placebo births.

The first step before assigning placebo births to never parents is removing from the sample those individuals from young cohorts who are likely to have children in the future. We start by dividing individuals into ten-year cohorts, distinguishing between old (1951, 1961 and 1971) and young (1981, 1991 and 2001), and computing young cohort women's motherhood

Table B.2: Summary statistics of the alternative counterfactual samples.

	(1) Parents	(2) Infertile	(3) Never parents
(a) Individual characteristics			
Age	30.03	33.58	32.63
	(5.972)	(6.356)	(6.382)
Women (%)	54.46	65.81	66.54
	(49.81)	(47.64)	(47.23)
College educated (%)	35.92	49.54	49.72
	(47.98)	(50.23)	(50.05)
Active in labour mkt (%)	86.34	93.16	88.57
	(34.33)	(25.35)	(31.77)
Employed (%)	81.11	89.74	86.38
	(39.12)	(30.47)	(34.26)
Monthly labour income (£)	1982.2	2079.3	2122.4
	(1521.7)	(1365.2)	(1774.0)
Weekly work hours	29.50	33.38	30.95
	(16.40)	(13.06)	(14.38)
Weekly housework hours	7.095	7.833	7.229
	(5.638)	(6.068)	(4.985)
RQ	0.371	0.303	0.214
	(0.875)	(0.769)	(0.979)
Observations	5821	117	547
(b) Couple characteristics			
Tenure	4.157	5.675	5.919
	(3.256)	(4.525)	(5.103)
Married (%)	43.48	44.87	38.87
	(49.33)	(50.06)	(48.32)
Monthly household income (£)	4082.0	4714.3	4548.4
	(3011.8)	(2705.3)	(2983.9)
Female share of paid work	0.472	0.499	0.486
	(0.209)	(0.248)	(0.237)
Female share of housework	0.630	0.680	0.586
	(0.204)	(0.166)	(0.229)
Observations	4010	78	364

Notes: This table presents summary statistics for three different samples: (1) the main analysis sample of parents (*Parents*); (2) the sample of individuals becoming pregnant but never having a live birth (*Infertile*); (3) the sample of individuals never getting pregnant and never having children (*Never parents*). All individuals considered are above 18. The period considered includes from four years before (1) first birth, (2) first pregnancy and (3) placebo birth, until one year before. Income measures are given in gross terms. Standard errors are displayed in parentheses.

probability. This will then be extrapolated to their partners.

Using observations from old cohorts, we estimate a linear probability model of becoming a mother on a set of observable characteristics: cohort, within-cohort income quartile, within-cohort income quartile of the partner, educational level (college or not) and country (England, Wales, Scotland or Northern Ireland). We then use the obtained coefficient estimates, to compute the fitted values for the young cohorts. This gives us the probability that each woman becomes a mother before the end of their fertility period. We sort young cohort women by cohort based on that probability.

We next want to determine which women will have children before the end of their fertility cycle. To do that, we compute the share of mothers in old cohorts by cohort and impute births for young women based on their motherhood probability such that the share of women who become mothers is the same across old and young cohorts. The individuals that get a birth imputed stop forming part of our control sample of never parent individuals.

The final control group is formed by old cohort individuals who never had children and young cohort individuals who do not have children and are unlikely to have them in the future. This is the sample of individuals to which we assign placebo births. We do this by assigning placebo birth ages. We approximate the distribution of age at first birth with a cohort and education level specific log normal distribution. This is straightforward for old cohorts, since we observe the entire distribution of age at first birth. For young cohorts, we use the predicted average from estimating a linear trend on old cohorts as the average of the distribution. We then assign a random draw from the corresponding distribution to each individual and compute the placebo event-time periods.

Table B.2 (3) presents summary statistics for the control sample of never parents, during the period before the placebo birth. Compared to the treatment group, in column (1), never parents are older, more often college educated, have higher earnings, have been together for longer and are less often married. Average RQ is also below the treatment sample mean before birth. However, as seen in Figure B.5 (b), average RQ becomes higher than parents' after the birth of their first child.

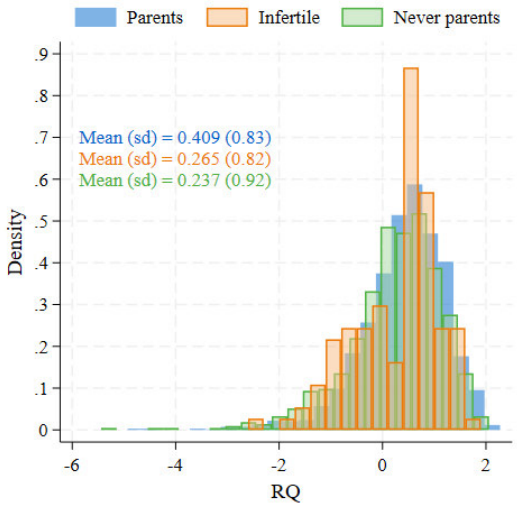
We repeat the main analysis using never parents as the control group and exploiting placebo births. However, the identification assumptions are somewhat less robust in this approach, particularly the parallel trends assumption. A key difference between the treatment and control groups is that never parents do not wish to have children. For the parallel trends assumption to hold, wanting children but being unable to conceive should have no impact on RQ. Recall that infertile individuals experience a decline in RQ following a failed pregnancy. Arguably, if the treated group had been unable to become parents, a similar drop in RQ might have occurred.

We plot the estimation results in Figure B.6. The dynamics and magnitudes of the estimated coefficients are very similar to the baseline estimates in Figure 2 – they are virtually the same

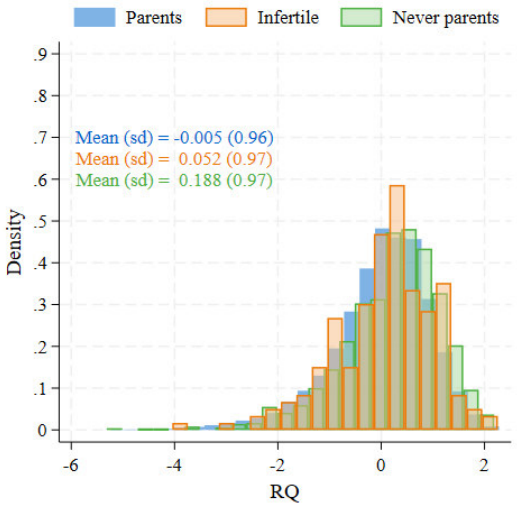
until five years after birth, but more precisely estimated. There is a further RQ decrease in year six and a recovery in the last estimated period.

Figure B.5: Distribution of RQ in different samples.

(a) Before

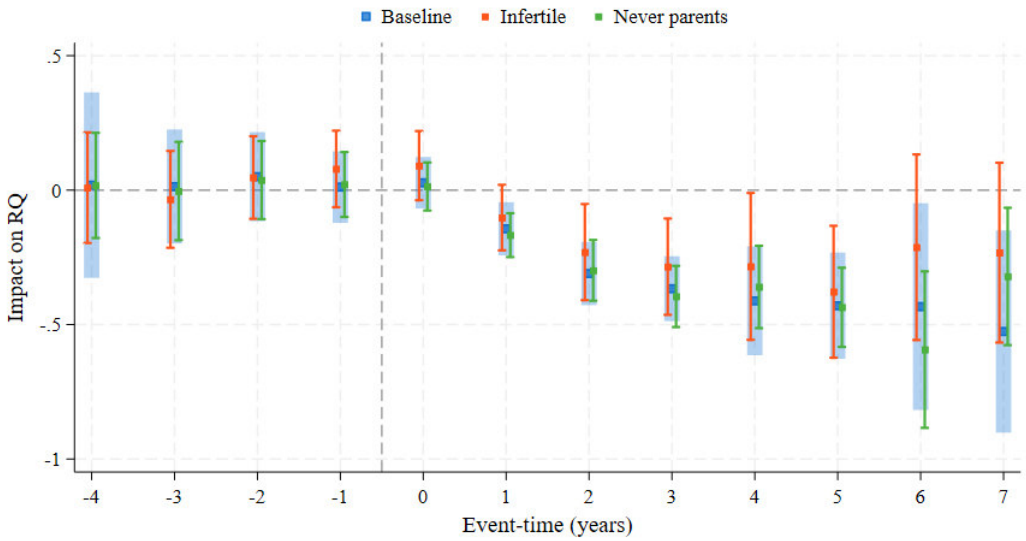


(b) After



Notes: This figure presents the distribution of RQ in three different samples: the main analysis sample of parents (*Parents*, in blue), the sample of individuals getting pregnant but never having a live birth (*Infertile*, orange) and the sample of individuals never having children (*Never parents*, green). Panel (a) contains the periods up to 4 years before birth for the treatment sample and up to 4 years before the first pregnancy attempt for the infertile sample. Panel (b) covers the period from birth up to 7 years after birth for the treatment sample and up to 7 years after the first pregnancy attempt for the infertile sample. Both panels display the full distribution for never parents.

Figure B.6: Impact of first child birth on RQ, using alternative control groups.

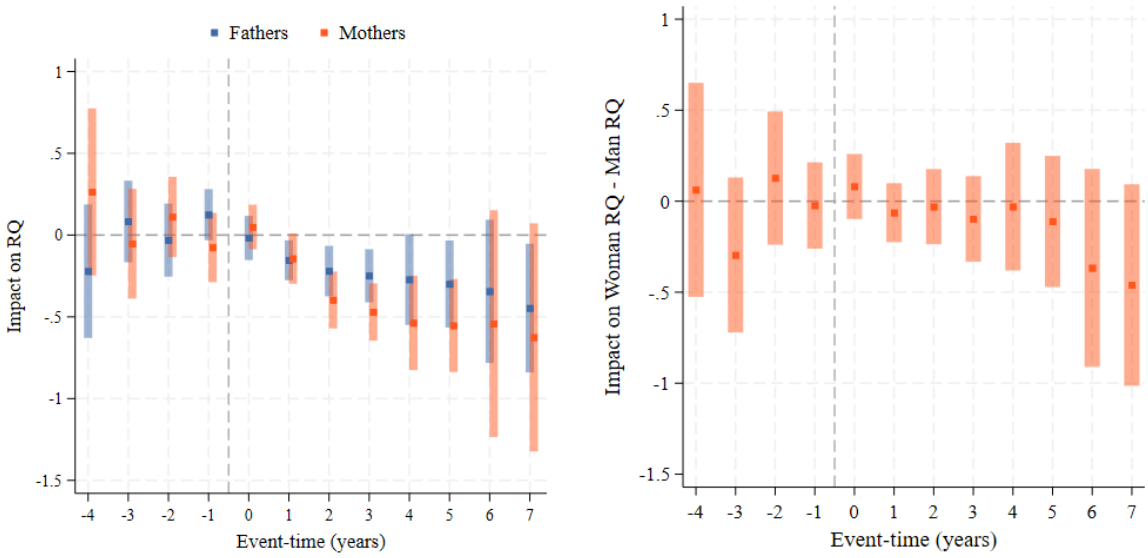


Notes: This figure presents the results from dynamic difference-in-differences estimations of the effect of first child birth on RQ, using the method proposed by Callaway and Sant'Anna (2021). The main treatment sample includes individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18–45 (women) or 18–50 (men) at the time of birth, and were not cohabiting with children from previous relationships. The baseline analysis (blue) uses individuals becoming parents at later dates as controls. The *infertile* analysis (orange) uses as the control group the set of individuals having a pregnancy between ages 18 and 45 who never become parents. We control for the period of first pregnancy. The *never parent* analysis (green) uses as controls individuals never becoming pregnant. We assign placebo births to the control group and control for placebo birth dates. We also plot 95% confidence intervals.

C Related Results

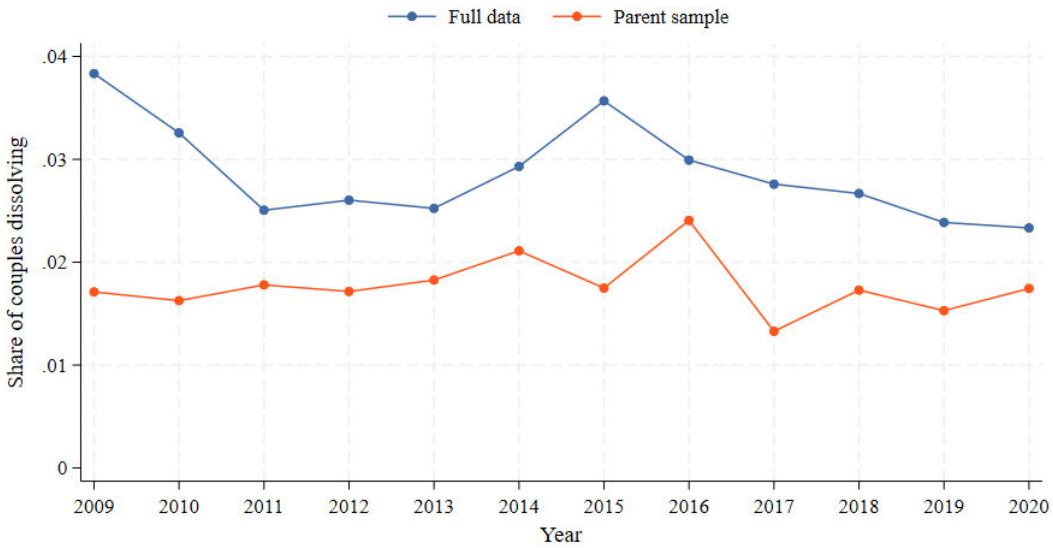
Figure C.1: Impact of first child birth on RQ for mothers and fathers.

(a) Separately (b) Difference



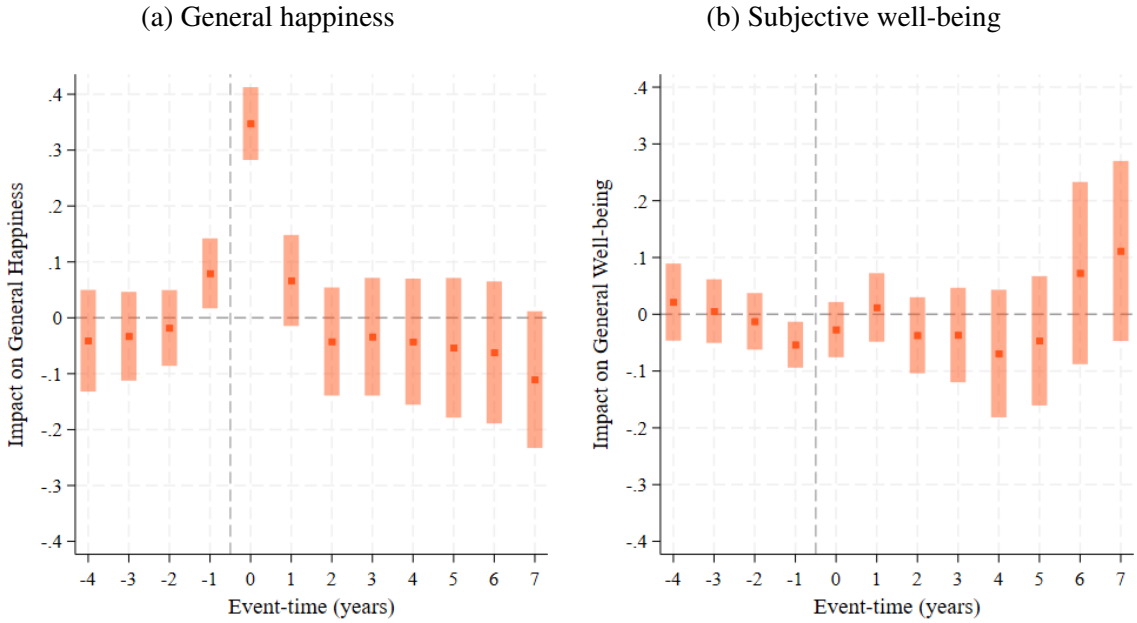
Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of first child birth on RQ, using the method proposed by Callaway and Sant'Anna (2021). Panel (a) reports separate estimates for mothers and fathers in the main sample of parents. Panel (b) uses the difference between women's RQ and their male partners' RQ as the outcome. We plot 95% confidence intervals.

Figure C.2: Yearly couple dissolution rate.



Notes: This figure plots the number of couples that dissolve each year as a share of all observed couples in the full data (blue) and in the main sample of parents (orange). Office for National Statistics (2022) reports that, on average, 1.44% of married couples of reproductive age (20–45 years old) in England and Wales divorced annually between 2009 and 2021.

Figure C.3: Impact of children on general happiness and subjective well-being.



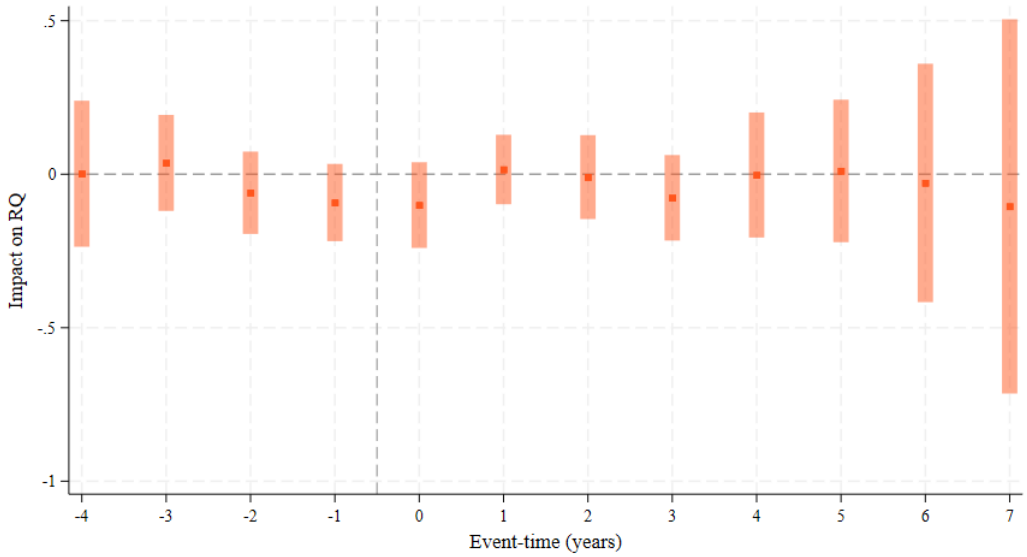
Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of first child birth on different outcomes, using the method proposed by Callaway and Sant'Anna (2021). Panel (a) reports results for a standardized measure of general happiness, based on responses to the question "Have you recently been feeling reasonably happy, all things considered?". Responses are recorded on a four-point Likert scale. Panel (b) displays results for a standardized measure of subjective well-being, constructed by Understanding Society using responses to the 12-item General Health Questionnaire (GHQ-12). We plot 95% confidence intervals.

Table C.1: Regression of general happiness on RQ in the sample of never parents.

	General Happiness		
	(1)	(2)	(3)
RQ	0.239***	0.229***	0.209***
	(0.013)	(0.014)	(0.024)
Controls		✓	✓
Individual FE			✓
R-squared	0.051	0.079	0.043
Observations	12147	9692	9692

Notes: This table presents the results from regressing a standardized measure of general happiness on RQ in the sample of individuals who never become parents. The sample consists of individuals first observed between ages 18 and 45 who are never observed cohabiting with children. The general happiness measure is based on responses to the question "Have you recently been feeling reasonably happy, all things considered?". Responses are recorded on a four-point Likert scale. Controls include age, relationship tenure, period, sex, college education, employment status, marital status, and urban residence. Standard errors clustered at the couple level are reported in parentheses.

Figure C.4: Impact of unemployment on RQ.



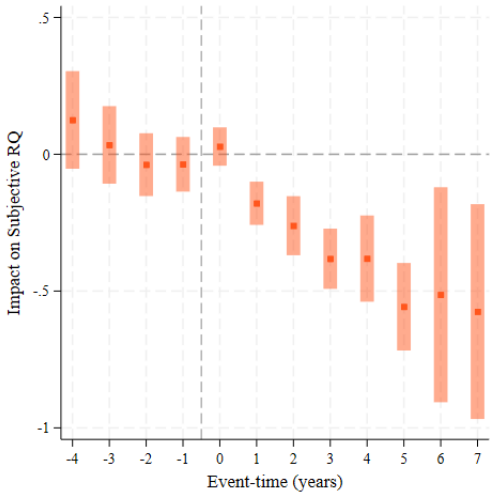
Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of unemployment events on RQ, using the method proposed by Callaway and Sant'Anna (2021). The sample consists of parents from the main analysis sample, restricted to individuals of working age (20-55). An unemployment event is defined as being unemployed at the time of the survey but not in the previous period. We only consider the first instance of unemployment events and the control group is formed only by individuals who experience unemployment at some point ("not yet unemployed"). We plot 95% confidence intervals.

D Robustness

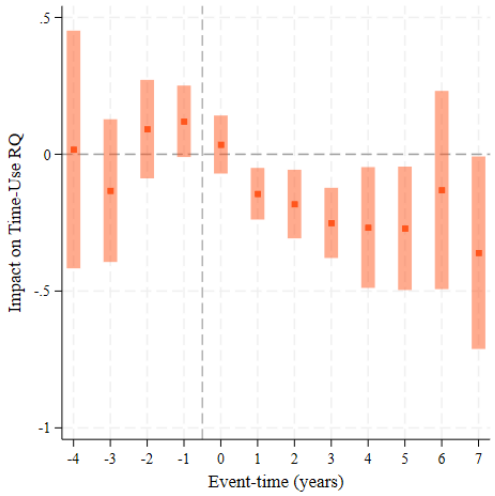
D.1 Time invariance of RQ

Figure D.1: RQ by item block.

(a) Subjective assessment

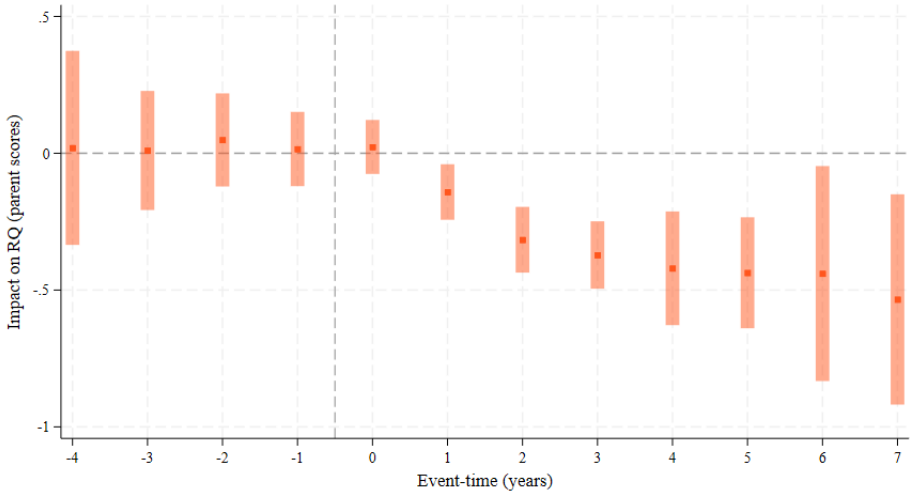


(b) Couple time use



Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of first child birth on alternative RQ measures, using the method proposed by Callaway and Sant'Anna (2021). The outcome in panel (a) is constructed computing a factor analysis on the subjective assessment items in Table 1, retaining the first factor. The outcome in panel (b) is based on a factor analysis on the couple time use items in Table 1, also retaining the first factor. We plot 95% confidence intervals.

Figure D.2: Impact of children on RQ using factor scores after birth.

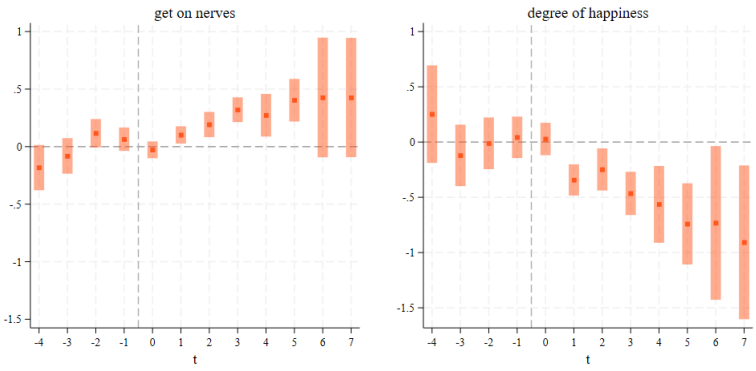


Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of first child birth on RQ, using the method proposed by Callaway and Sant'Anna (2021). The outcome is derived from a factor analysis of the items in Table 1, retaining the first factor, and is constructed using only observations from individuals who have already become parents. We plot 95% confidence intervals.

Figure D.3: Impact on subjective assessment items.

(a) Consider splitting (b) Regret marriage (c) Quarrel

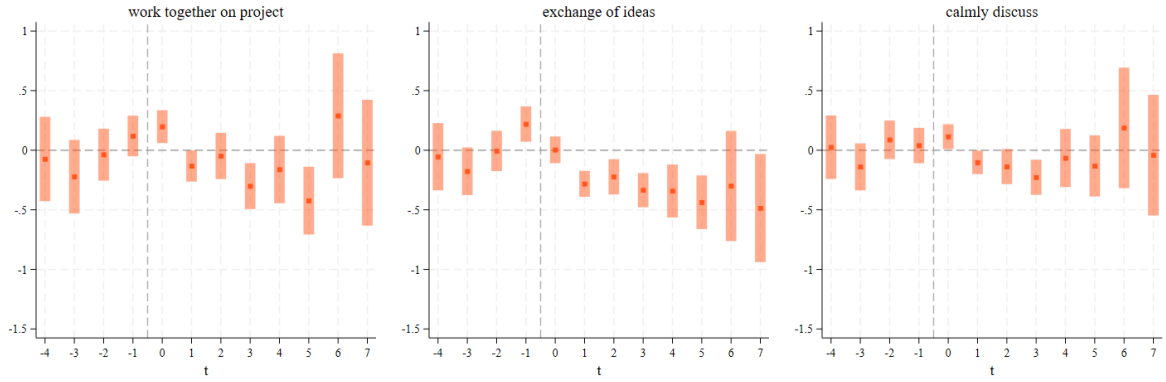
(d) Get on nerves (e) Degree of happiness



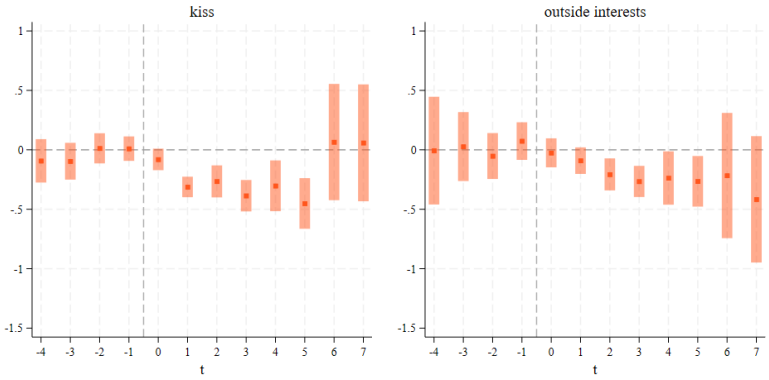
Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of first child birth on different measures, using the method proposed by Callaway and Sant'Anna (2021). The outcomes are the standardized responses to the subjective assessment items in Table 1. We plot 95% confidence intervals.

Figure D.4: Impact on couple time use items.

(a) Work together (b) Exchange ideas (c) Calmly discuss



(d) Kiss (e) Outside interests

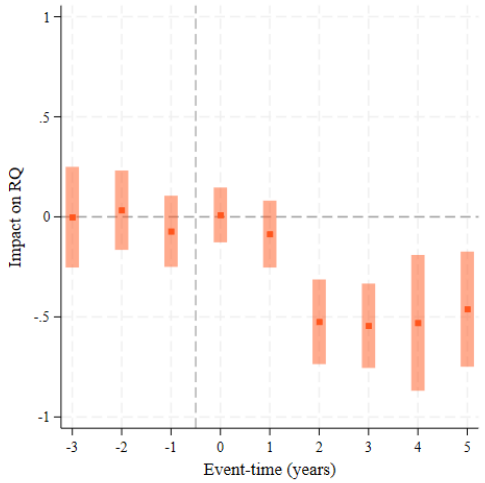


Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of first child birth on different measures, using the method proposed by Callaway and Sant'Anna (2021). The outcomes are the standardized responses to the couple time use items in Table 1. We plot 95% confidence intervals.

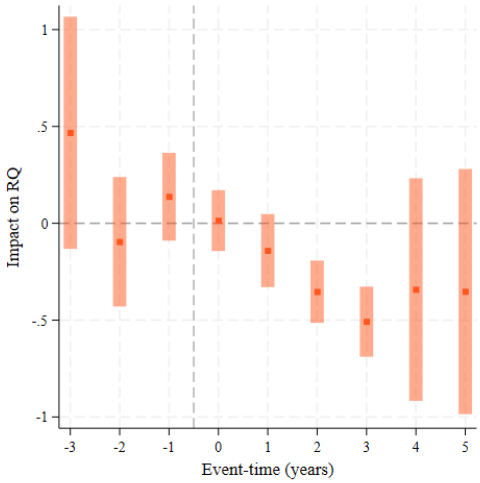
D.2 Total realized fertility

Figure D.5: Impact by final number of children.

(a) One child



(b) Two children



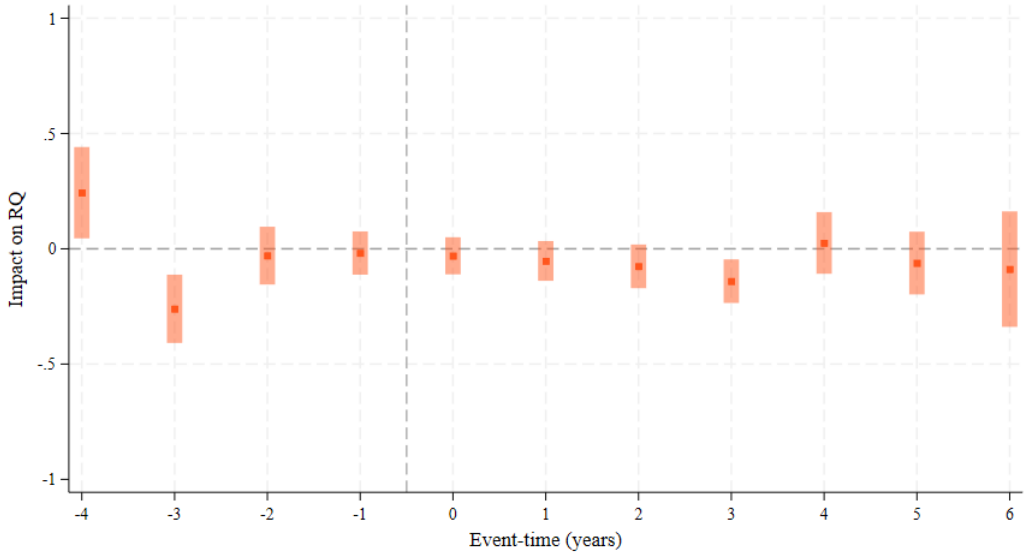
Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of first child birth on RQ in different samples, using the method proposed by Callaway and Sant'Anna (2021). Panel (a) shows results for individuals who have only one child, and panel (b) displays results for individuals with two or more children by the end of the observation period. We plot 95% confidence intervals.

Table D.1: Impact by total realized fertility.

	(1) One child	(2) More than one
ATT	-0.360***	-0.212*
	(0.102)	(0.101)
Observations	693	1041

Notes: This table presents results from a static difference-in-differences estimation of the effect of first child birth on RQ, using the method proposed by Callaway and Sant'Anna (2021). Column (1) reports results for individuals who have only one child by the end of their fertility cycle (age 40), and column (2) presents results for those with more than one child. Standard errors clustered at the couple level are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure D.6: Impact of second child birth on RQ.

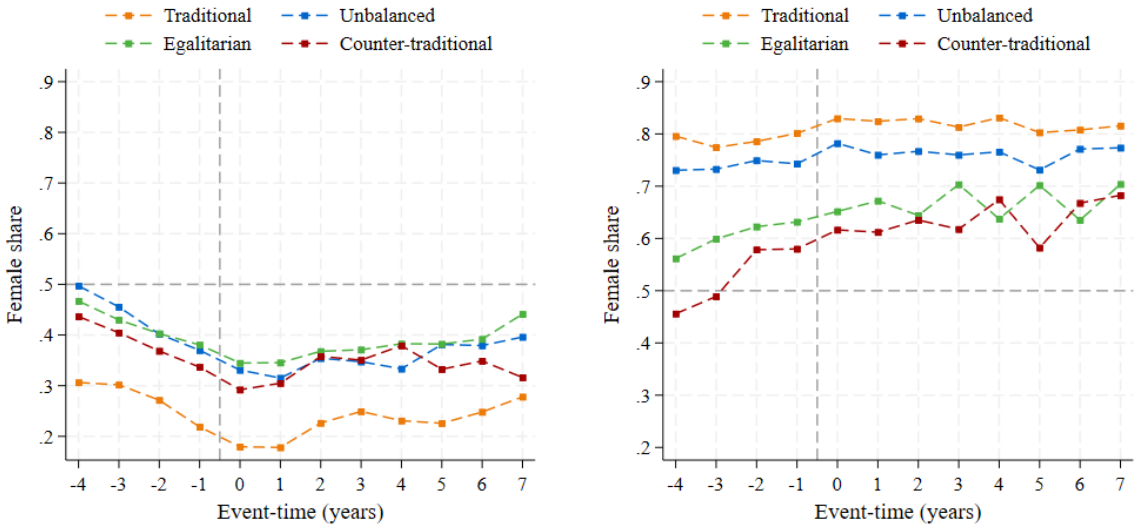


Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of second child birth on RQ in different samples, using the method proposed by Callaway and Sant'Anna (2021). The sample includes individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18–45 (women) or 18–50 (men) at the time of birth, were not cohabiting with children from previous relationships, and became second-time parents. We plot 95% confidence intervals.

Figure D.7: Household specialization around second child birth by couple type.

(a) Paid work share

(b) Unpaid work share

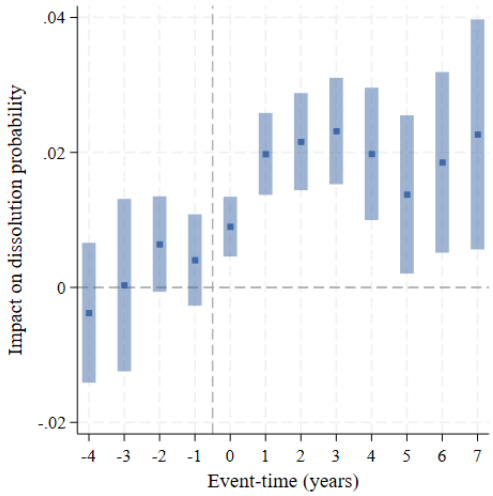


Notes: This figure displays the average the female shares of (a) paid and (b) unpaid work — defined as the share of total weekly household paid labour market hours and unpaid housework hours performed by women — at each time around the birth of the second child. These averages are plotted separately by couple type, defined based on the distribution of tasks before first childbirth. The sample includes individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18–45 (women) or 18–50 (men) at the time of birth, were not cohabiting with children from previous relationships, and became second-time parents.

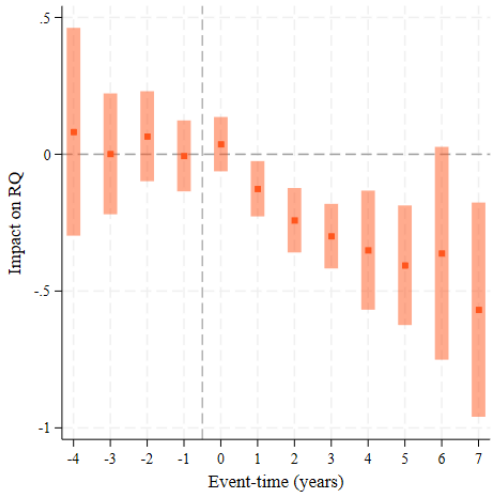
D.3 Selected sample

Figure D.8: Sample selection on separating couples.

(a) Impact on couple divorce.



(b) Impact on non-separating couples.

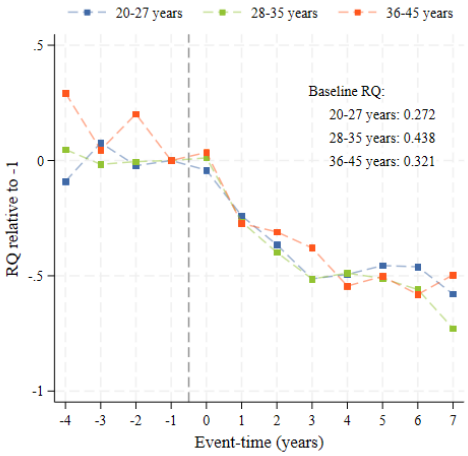


Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of first child birth, using the method proposed by Callaway and Sant'Anna (2021). Panel (a) uses couple separation as the outcome in the main sample of parents. Note that, by sample construction, individuals cannot be observed separating before the child is conceived. Panel (b) uses RQ as the outcome in the sample of first-time parents who have not separated by the end of the observation period. We plot 95% confidence intervals.

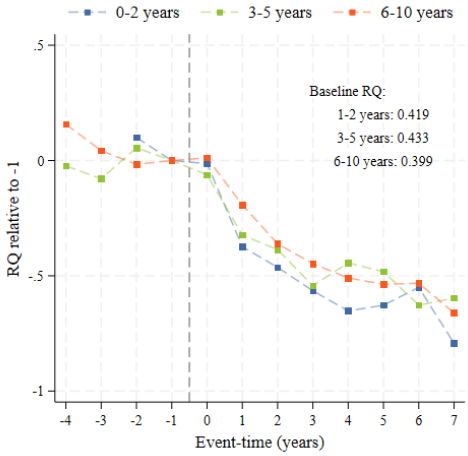
D.4 Age and tenure at birth

Figure D.9: Average RQ per event-time period.

(a) Age bins

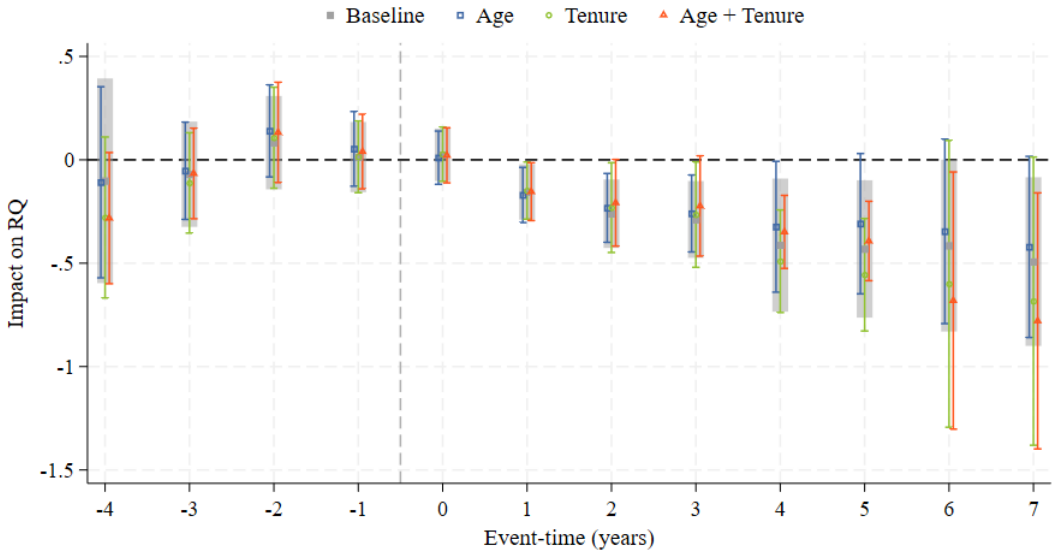


(b) Tenure bins



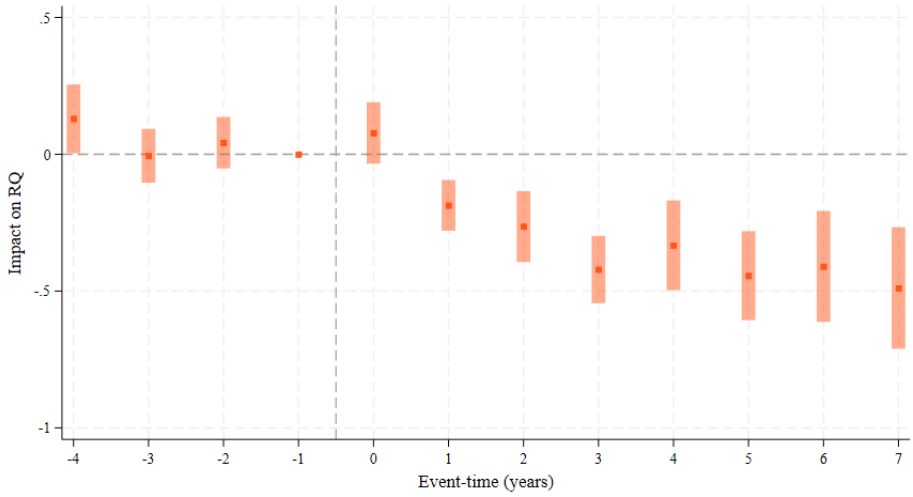
Notes: This figure plots the average RQ at each event-time period. The average in the period before birth ($t - G_i = -1$) is normalized to zero. Panel (a) displays these means separately for individuals who had their first child at different ages: 20-27 years, 28-35 years and 36-45 years. Panel (b) displays RQ means separately by relationship tenure at first birth: 1-2 years, 3-5 years and 6-10 years.

Figure D.10: Impact on RQ using different sets of controls.



Notes: This figure presents results from a dynamic difference-in-differences estimation of the effect of first child birth on RQ, using the method proposed by Callaway and Sant'Anna (2021) and including different sets of controls. The method accounts for age and tenure by comparing individuals with the same value of the covariates in the period before birth. We plot 95% confidence intervals.

Figure D.11: Impact of first child birth on RQ using Kleven et al. (2019b).



Notes: This figure presents the result from estimating the following regression:

$$y_{i,t} = \mu_t + \sum_{j \neq -1} \mathbb{1}\{j = t - G_i\}\delta_j + \sum_a \mathbb{1}\{a = \text{age}_{i,t}\}\alpha_a + \sum_d \mathbb{1}\{d = \text{tenure}_{i,t}\}\gamma_d + u_{i,t}$$

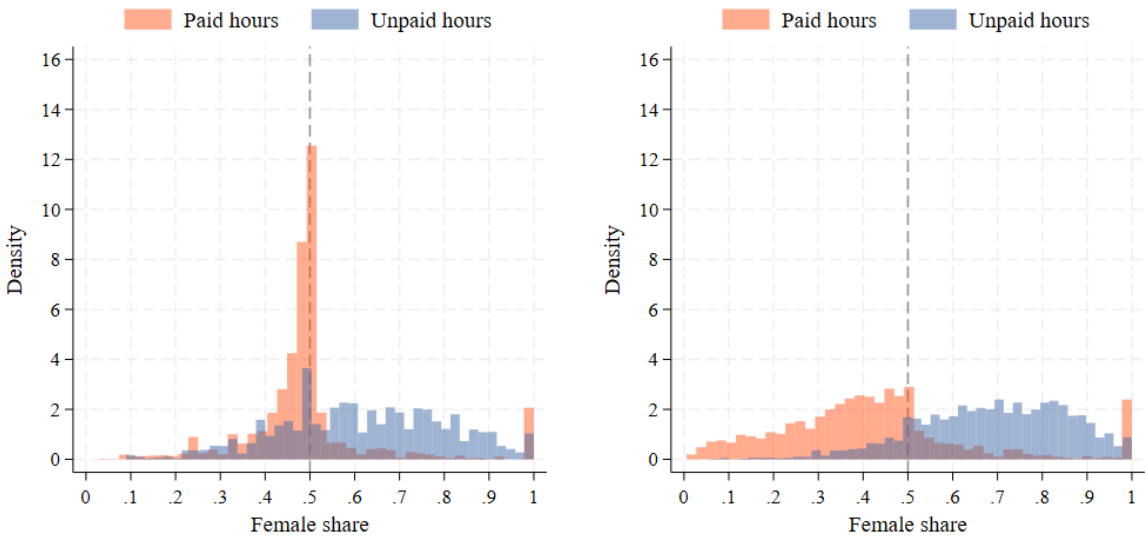
This regression is the event-study specification popularized by Kleven et al. (2019b), and it accounts for the dynamic effects of age and tenure. The plotted coefficients $\hat{\delta}_j$ correspond to the event-time dummies. Standard errors clustered at the couple level are used to construct the plotted 95% confidence intervals.

E Household Specialization

Figure E.1: Distribution of the female share of paid labour market and unpaid housework hours.

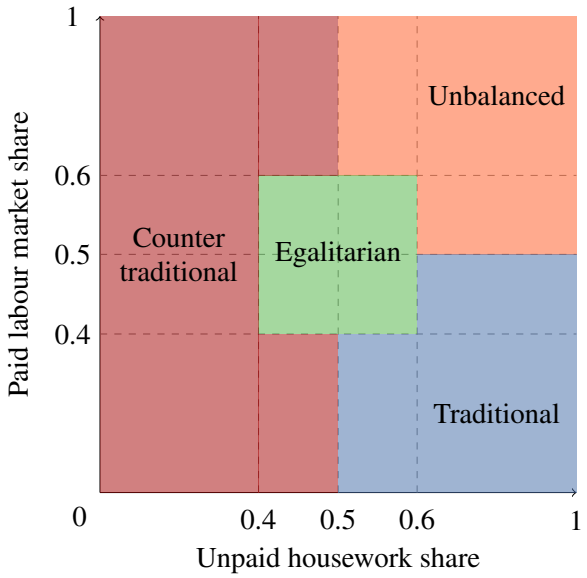
(a) Before birth

(b) After birth



Notes: This figure displays the distribution of the female shares of paid and unpaid work — defined as the share of total weekly household paid labour market hours and unpaid housework hours performed by women — in the sample of individuals who become parents. Panel (a) presents data for up to four years before the birth of the first child. Panel (b) covers the period from the year of birth to seven years after. These time frames correspond to the main analysis period.

Figure E.2: Classification of couples based on the female shares.



Notes: This figure presents the classification of couples based on the female share of paid labour market work and unpaid housework. These shares represent the proportion of total weekly household paid labour market hours and unpaid housework hours performed by women, each ranging from zero to one. Classification is based on average shares before the birth of the first child. We distinguish four couple types: (i) Traditional couples (blue), where women perform more than 50% of housework and men take a larger share of paid work; (ii) Unbalanced couples (orange), where women take on more than half of both types of work; (iii) Egalitarian couples (green), where men and women share both types of work roughly equally, with female shares between 40% and 60%; and (iv) Counter-traditional couples (red), where men contribute more than 50% to housework.

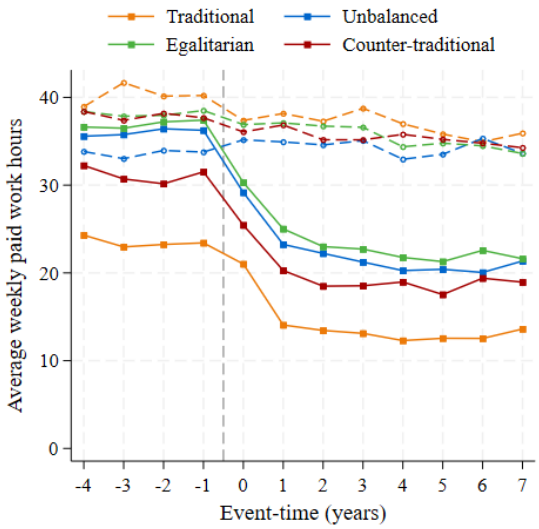
Table E.1: Summary statistics before the birth of the first child, by couple type.

	Traditional		Unbalanced		Egalitarian		Counter-tradit.	
	Fathers (1)	Mothers (2)	Fathers (3)	Mothers (4)	Fathers (5)	Mothers (6)	Fathers (7)	Mothers (8)
(a) Individual characteristics								
Age	31.96	29.48	32.59	30.38	32.45	30.45	32.73	29.75
	(5.749)	(4.942)	(5.549)	(4.779)	(5.312)	(4.284)	(5.344)	(4.376)
College educated (%)	0.323	0.391	0.406	0.421	0.413	0.518	0.421	0.549
	(0.468)	(0.488)	(0.492)	(0.495)	(0.493)	(0.500)	(0.495)	(0.499)
Active in labour mkt (%)	0.988	0.923	0.898	0.981	0.989	0.992	0.929	0.923
	(0.109)	(0.267)	(0.303)	(0.136)	(0.105)	(0.0910)	(0.258)	(0.267)
Employed (%)	0.980	0.884	0.857	0.974	0.989	0.978	0.898	0.908
	(0.140)	(0.319)	(0.351)	(0.160)	(0.105)	(0.148)	(0.303)	(0.290)
Monthly labour income (£)	2775.2	1715.2	2534.7	2353.1	2757.2	2253.5	2482.8	2369.6
	(1703.2)	(1180.4)	(1847.6)	(1256.6)	(1275.6)	(964.6)	(1473.8)	(1307.1)
Weekly work hours	39.70	29.42	30.59	36.65	38.66	36.56	33.93	32.79
	(8.868)	(13.35)	(14.43)	(7.521)	(5.767)	(6.609)	(14.76)	(12.22)
Weekly housework hours	3.553	10.65	4.013	10.24	6.789	7.059	9.051	5.110
	(2.696)	(5.715)	(3.078)	(5.002)	(3.347)	(3.390)	(4.354)	(4.102)
RQ	0.180	0.381	-0.00236	0.0110	0.493	0.511	0.446	0.392
	(0.944)	(0.830)	(0.664)	(0.725)	(0.635)	(0.633)	(0.968)	(0.778)
Observations	500	497	266	266	359	360	196	195
(b) Couple characteristics								
Tenure	4.616		4.914		4.810		4.604	
	(3.092)		(3.234)		(2.896)		(2.907)	
Married (%)	0.698		0.664		0.652		0.713	
	(0.453)		(0.468)		(0.473)		(0.449)	
Monthly household income (£)	4049.0		4226.7		4627.7		4483.8	
	(2175.3)		(2446.8)		(2311.1)		(2390.9)	
Female share of paid work	0.398		0.586		0.484		0.509	
	(0.165)		(0.180)		(0.0668)		(0.222)	
Female share of housework	0.747		0.726		0.511		0.338	
	(0.148)		(0.138)		(0.105)		(0.150)	
Observations	502		269		361		197	

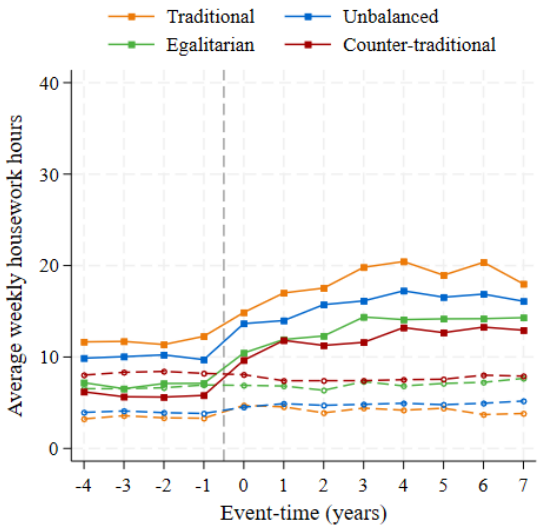
Notes: This table presents summary statistics for the main analysis sample of parents in the period before the birth of the first child, disaggregated by couple type based on their division of labour market and housework time prior to birth. The classification is illustrated in Figure E.2. Panel (a) reports mean values of individual characteristics separately for women and men, and panel (b) summarizes couple characteristics. Female shares of paid and unpaid work represent the share of total weekly household paid labour market hours and unpaid housework hours performed by women. Income measures are presented in gross terms. Standard deviations are reported in parentheses.

Figure E.3: Time use around first child birth by couple type.

(a) Paid work share



(b) Unpaid work share



Notes: This figure displays the average weekly (a) paid labour market hours and (b) unpaid housework hours at each time around the birth of the first child. These averages are plotted separately by couple type, defined based on the distribution of tasks before first childbirth, and by sex. The sample includes individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18–45 (women) or 18–50 (men) at the time of birth, and were not cohabiting with children from previous relationships.

F Alternative Mechanisms

The main mechanism we study are changes in household specialization after the birth of the first child. In this section, we consider the possibility that other channels may mediate the effect of childbirth on RQ. Specifically, we study changes in gender norms, household income and the within-couple earnings gap as potential alternative mechanisms.

Gender norms. To measure gender norms, we use survey questions on gender role attitudes. We employ a subset of the standard questionnaire developed by the International Social Survey Programme (ISSP), widely used in the economics literature on gender norms (e.g., Bertrand et al., 2021; Farré et al., 2022). These questions are part of the Adult Gender Attitudes module of Understanding Society, which is available in waves 2, 4 and 10. Specifically, we use a set of items on a 5-point Likert scale on agreement with a set of statements related to the role of women and men at home and in the labour market, listed in Table F.1.[31]

Following Farré et al. (2022), we recode the items such that higher values indicate more gender-equal views: a score of 0 corresponds to the most conservative position (agreement with (a) and disagreement with (b) in Table F.1) and 4 corresponds to the most egalitarian views. We construct the measure of gender role attitudes by running a factor analysis on the 5 items and retaining the first factor. All items' loadings are positive and the factor explains 43.91% of the variation in the data. The resulting is standardized to zero mean and a standard deviation of one. Figure F.1(a) presents the distribution of this measure before the birth of the first child. The distribution is skewed to the right, indicating relatively more progressive views on average, with considerable heterogeneity.

Table F.1: Items in the gender role attitude questionnaire.

Do you personally agree or disagree...	
(a) Conservative statements	(b) Progressive statements
A pre-school child is likely to suffer if his or her mother works.	Both the husband and wife should contribute to the household income.
All in all, family life suffers when the woman has a full-time job.	Employers should make special arrangements to help mothers combine jobs and childcare.
A husband's job is to earn money, a wife's job is to look after the home and family.	

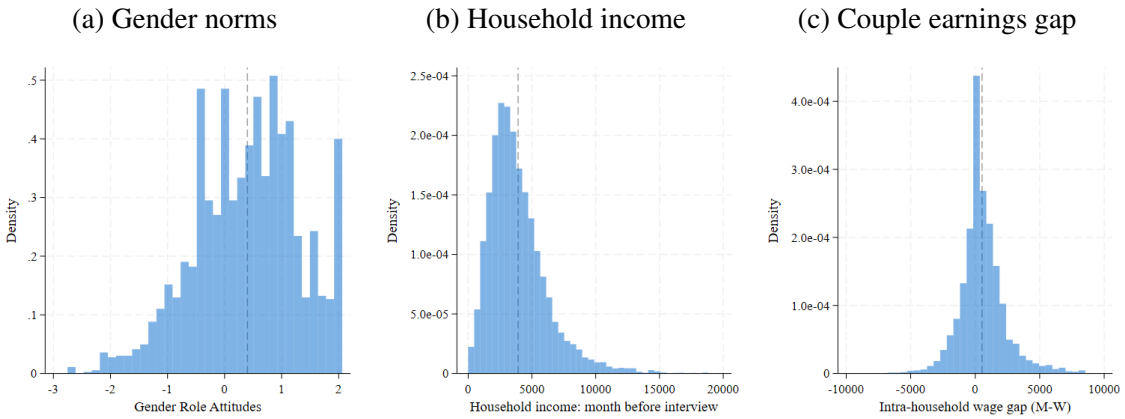
Notes: This table lists all the questions used to construct the gender role attitudes, drawn from the Understanding Society Adult Gender Attitudes module. This self-completion questionnaire is administered in waves 2, 4 and 10. We group these items in two blocks based on whether the statements are of (conservative) or (b) progressive nature. The responses are given in a five-point Likert scale of agreement, where one means complete agreement and five complete disagreement.

[31] It is missing the following items from the usual questionnaire: "A job is all right, but what most women really want is a home" and "Being a housewife is just as fulfilling as working for pay".

Household income. Our measure of household resources is gross monthly household income, reported for the month prior to the interview, in pounds. Since household income is reported at the household level, both members of the couple have the same value. Figure F.1 (b) presents the pre-birth distribution of this variable, which averages around £4,000 per month but exhibits a long right tail.

Couple earnings gap. Finally, we construct the within-couple earnings difference using gross monthly individual labour income. We compute the gap as the male partner's monthly earnings minus the female partner's. Positive values indicate that men have higher earnings, while negative values capture cases where women are the primary earners. Figure F.1 (c) presents the pre-birth distribution. It is centred slightly above zero, consistent with an average gender gap in favour of men, but there is substantial variation on both sides of the distribution.

Figure F.1: Distribution of alternative mechanisms.



Notes: This figure plots the distribution, within the main analysis sample of parents and before childbirth, of (a) gender role attitudes, (b) household income and (c) the within-couple earnings gap. The sample consists of individuals in heterosexual cohabiting relationships who became first-time parents between 2009 and 2022, were aged 18-45 (women) or 18-50 (men) at the time of birth, and were not cohabiting with children from previous relationships.

Table F.2 reports the correlations between our main mechanism, female shares of paid and unpaid work as defined in Subsection 2.2, and the alternative mechanisms discussed in this section. The strongest correlation arises between the couple earnings gap and the female share of paid work: as women contribute more hours in the labour market, the earnings gap shifts in favour of them being the primary earners. Conversely, a higher female share of unpaid work is associated with larger gaps in favour of men. The female share of paid work is also positively correlated with more egalitarian gender role attitudes, suggesting an alignment between labour market participation and progressive views. The weakest correlation is between our main mechanism and household income.

Table F.2: Correlation between alternative mechanisms.

	Share of paid work	Share of unpaid work	Gender role attitudes	Household income	Couple earnings gap
Share of paid work	1				
Share of unpaid work	-0.310	1			
Gender role attitudes	0.275	-0.162	1		
Household income	0.126	-0.028	0.143	1	
Couple earnings gap	-0.551	0.269	-0.217	0.226	1

Notes: This table presents the correlations between our main mechanism (the female shares of paid and unpaid work) and the alternative mechanisms considered in this section (gender norms, household income and the intra-household earnings gap). The correlations are computed in the main analysis sample of first time parents across all available periods.

F.1 Impact of First Child Birth

We begin by examining how the different mechanisms change around the birth of the first child. Mechanisms that undergo larger shifts have greater potential to mediate the effect of childbirth on RQ. We consider both our main mechanism and the alternative mechanisms discussed above. We standardize all measures to facilitate the comparison of coefficient magnitudes. We estimate a static difference-in-differences specification in the main analysis sample of first time parents, using the method proposed by Callaway and Sant'Anna (2021), and use each mechanism as the outcome in separate regressions.

Table F.3 reports the results. The largest changes occur in the household specialization variables. After first birth, women's share of paid work falls by 0.40 standard deviations and their share of housework rises by 0.33 standard deviations. The next largest coefficient is the one corresponding to the within-couple earnings gap, which increases after birth. This indicates that women lose relative economic power within the couple in favour of men. Finally, we find a marginally significant decline in the gender norms index, suggesting a shift toward more conservative attitudes after birth, and a reduction of household income.

Table F.3: Impact of first childbirth on potential mechanisms.

	(1) Paid share	(2) Unpaid share	(3) Gender norms	(4) HH income	(5) Couple gap
ATT	-0.404***	0.331***	-0.137*	-0.110**	0.187***
	(0.028)	(0.036)	(0.061)	(0.037)	(0.055)
Observations	24107	16846	1682	46254	7838

Notes: This table presents the results from a static difference-in-differences estimation of the effect of first child birth on alternative mechanisms, using the method by Callaway and Sant'Anna (2021): (1) the female share of paid work, (2) the female share of unpaid work, (3) gender role attitudes, (4) gross monthly household income, (5) within-couple monthly labour earnings gap. All measures are standardized. We use the main analysis sample of first time parents, as defined in Subsection 2.3. Standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

F.2 Children, RQ and Alternative Mechanisms

We study how RQ is associated with the different mechanisms considered, both overall and around childbirth. We start by estimating a linear regression of RQ on all potential mechanisms, standardized to facilitate the comparison of coefficient magnitudes.[32] As controls, we include second-order polynomials of age and tenure, dummies for employment status, labour market activity and college education, a marriage indicator, and the full set of time dummies. The sample is restricted to the usual observation window, covering four years before and seven years after first birth. Table F.4 (1) presents the results, showing significant associations only with gender norms and household income. In Column (2), we add individual fixed effects to capture within-individual variation, which mechanically excludes individuals observed only once and thus reduces the sample size. Under this specification, none of the mechanisms remain significant.

We then examine whether the impact of childbirth on RQ varies systematically with pre-birth levels of the different mechanisms. To do this, we compute the individual-specific average of each mechanism before birth and include them jointly in a static difference-in-differences regression, interacting each with the treatment (post-birth) indicator. We do not use the Callaway and Sant'Anna (2021) estimator, as it does not allow for interaction terms. Since the mechanisms are considered in a time-invariant way and we include individual fixed effects, we only obtain estimates of the baseline post-treatment dummy and the interaction terms. The results, presented in Table F.4 (3), indicate that none of the alternative mechanisms significantly mediate the impact of childbirth on RQ when considered jointly. The only marginally significant interaction is with the female share of paid hours, but this effect is not robust to alternative specifications.

Overall, the evidence is consistent with household specialization being the mechanism with the greatest potential to mediate the impact of childbirth on RQ. Compared to alternative factors, it undergoes the largest shifts around first birth and shows the most systematic associations with RQ. By contrast, changes in gender norms, household income or the within-couple earnings gap appear smaller and are in any case closely connected to household specialization.

[32] Since housework time and gender norms are not collected in the same survey waves as RQ, we use their lagged values.

Table F.4: Association between RQ and alternative mechanisms.

	(1)	(2)	(3)
Female share of paid hours	0.013	-0.081	
	(0.052)	(0.130)	
Female share of unpaid hours	-0.036	0.026	
	(0.030)	(0.072)	
Gender role attitudes	0.128***	0.144	
	(0.030)	(0.084)	
Monthly gross household income	0.110**	0.116	
	(0.037)	(0.084)	
Within-couple labour earnings gap	-0.019	-0.078	
	(0.041)	(0.119)	
Post			-0.011
			(0.071)
Post × Female share of paid hours			0.145*
			(0.065)
Post × Female share of unpaid hours			0.025
			(0.042)
Post × Gender role attitudes			0.024
			(0.042)
Post × Monthly gross household income			-0.033
			(0.054)
Post × Within-couple labour earnings gap			-0.003
			(0.056)
Controls	✓	✓	✓
Individual FE		✓	✓
R-squared	0.073	0.787	0.715
Observations	1058	258	1563

Notes: This table reports the associations between RQ and the alternative mechanisms in the main analysis sample of first time parents, as defined in Subsection 2.3. All mechanism measures are standardized. Columns (1) and (2) present results from estimating a linear regression of RQ on all mechanisms jointly. Column (3) reports results from a static difference-in-differences regression, where the treatment variable is interacted with individual-specific pre-birth averages of all mechanisms. All specifications control for second-order polynomials of age and tenure, dummies for employment status, labour market activity and college education, a marriage indicator, and the full set of time dummies. Columns (2) and (3) additionally include individual fixed effects. Standard errors are reported in parentheses.