The Effects of Family Instability on Children's Outcomes

Rebecca Lessem* and Carl Sanders†

July 8, 2019

Abstract

Children of married couples in the United States have better outcomes than those of unmarried cohabiting couples. It has been suggested in the economics and sociology literature that this correlation results from a causal difference in the ability of married versus cohabiting parents to commit to joint investments in children. In this paper, we quantify the effects of family relationship instability relationship on the outcomes of children. We create a structural household-level model of relationship and labor market choice and investment in children, and estimate it using measurements on children's outcomes drawn from the Fragile Families panel. Our counterfactuals consider the effects on children of increasing the desirability and stability of marriage, and we find that even large increases in marriage rates (from 25% to 50% in our sample) are associated with only small increases in child outcomes, with an average effect size that moves a child up 2 percentage points in the quality distribution relative to the baseline. This suggests that observed differences between child outcomes across relationship statuses are driven by selection on unobserved child quality, and that policies targeting parental relationships may not have large returns.

JEL Codes: J12, J13

1 Introduction

The share of children born to unmarried women in the US rose from 30% to 44% between 2000 and 2014.[1] Policymakers on both the left and the right sides of the political spectrum have pushed for programs that promote two-parent households, believing that this is best for children. There are plausible reasons to believe that family instability has a negative effect on children: for one, it reduces the resources available to the child, both in terms of time

*Carnegie Mellon University, rlessem@andrew.cmu.edu

†Washington University in St Louis, carlsanders@wustl.edu

[1]Statistics calculated by the Pew Research Center. See http://www.pewsocialtrends.org/2015/12/17/1-the-american-family-today/

and money, to only have one parent for extended periods of time. Additionally, children may have a difficult time adjusting psychologically to the uncertainty associated with having an unstable family. However, there are also reasons to think this instability may be a relatively minor problem relative to other issues such as income inequality: worse outcomes for children of unstable families are strongly confounded by the large socioeconomic difference between these families, and family instability may simply be another symptom of poverty rather than a cause of it.

In this paper, we study how family instability causally affects child outcomes, focusing on cognitive and physical development.[2] To do this, we develop and estimate a dynamic model of marriage and child development, where parental marital status affects the evolution of child outcomes over time. We use our results to quantify how much child outcomes would improve if there were policies put in place to encourage parents to stay together: for example, transfers that were tied to the couple remaining as joint parents, or changes to the legal treatment of cohabiting couples making it more costly to separate.

Estimating the causal effects of relationship status on the development of children is a difficult problem, and our approach relies on using an economic choice model to correct for selection into relationships. The most closely related work to ours is Tartari (2015), who develops and estimates a dynamic model to analyze how divorce affects child outcomes. Our work builds on this research in three dimensions. First, we analyze marriage and cohabitation separately, in contrast to previous literature that would have considered unmarried couples living together to be the same as singles. This is important because marriage may create a more binding commitment to stay together, which could potentially create more incentives for couples to make long-term joint investments in their children. Second, in contrast to Tartari (2015), we use a sample that includes children whose parents were not married when the child is born. Therefore, instead of solely analyzing the impact of divorce, we can also learn about the impact of marital instability on children born to unmarried parents, which is the population that likely faces the most adverse outcomes. And lastly, instead of using simple averages of academic test scores as a measure of child outcomes, we use a latent factor method, following Cunha et al. (2010), Agostinelli and Wiswall (2016), and Agostinelli (2018). This allows us to aggregate a variety of measurements at the child level, and our estimation strategy will reveal the relative informativeness of each measure.

The model is setup as follows. In each period, the mother chooses her relationship status with her child's father, her labor market status, and how much to invest in the child. Child outcomes evolve over time, as a function of marital and labor market status and parental

[2]We also have data on mental health outcomes, which we will include in future work to allow us to also study emotional development.

investment. The mother considers the dynamic implications of her choices on future child outcomes, creating differing incentives across relationships over time and depending on how her child is doing. We estimate the model using data from the Fragile Family and Child Well-Being Study, a survey that starts from a sample of mothers and children at birth. The survey follows the mother and child up to age 15, allowing us to track parental investments and marriage decisions at each age, as well as the child's academic and health outcomes. Because investment and outcomes are not directly observed in the data, in estimation we use the fact that the problem is entirely standard conditional on the child's latent quality, and we can recover information on child quality using our test scores and other measures.

Our results show that relationship status and labor market decisions substantially affect children's outcomes. In particular, we see very different path for children depending on their parent's relationship status. On average, children of married parents where the mother does not work have the highest outcomes, but there is considerable heterogeneity depending on individual demographic characteristics and the child's unobserved (although partially measured) quality. Our estimation strategy incorporates a variety of outcome measures, including health status, test scores, and grades; our results show that the academic test scores are the most informative about child latent quality. In a counterfactual, we raise the utility from marriage so that there is a 10 percentage point increase in the share of parents that are married in period 1. We find small effects of the counterfactual: increasing the desirability of marriage increases the average quality of children at age 15 by about 0.04 standard deviations. Our results are consistent with most of the observed differences between married and cohabiting parents being driven by selection on unobservable child quality: parents of children with higher test scores tend to select into marriage rather than the state of marriage itself being primarily responsible for those high scores.

The existing literature has indeed documented that family instability is negatively correlated with children's academic test scores and psychological well-being.[3] Craigie et al. (2012) find that many differences in children's outcomes across parents' current relationship status (married, cohabiting without marriage, or separated) are reduced when controlling for socioeconomic characteristics such as parents' education and financial resources. However, the actual causal impact of instability remains an open question. One exception is Lang and Zagorsky (2001), who use parental death as an arguably exogenous shock that causes a child to have one parent, and find that this does not affect economic well-being in adulthood. As mentioned above, Tartari (2015) quantifies the negative effects of parents' divorce on the

[3] There is also some evidence that this persists to adulthood. For example, see Ross and Mirowsky (1999), Amato and Sobolewski (2001), Maier and Lachman (2000), Tucker et al. (1997), and Hayward and Gorman (2004).

cognitive test scores of children using as structural model. Our work builds on these papers in a number of ways. First, we use a variety of measures to understand the child's outcomes. Our latent factor model allows us to learn about the relative importance of each factor, instead of simply taking the average. This methodology has been used in previous work, but not to study this research question. Second, we study the whole time path of relationship status to learn how this affects child outcomes. This differs from past work; for example, Tartari (2015) looks at the impact of divorce, whereas we look more broadly at relationship status in each period, including respondents who were not married when the child was born. Third, we study cohabitation separately from marriage, which is important given that institutional features make it easier to exit cohabiting relationships versus marriage. Given the current rates of cohabitation, it is important to understand how this household structure is affecting children's outcomes. There is some work that studies cohabitation and marriage, but does not examine how it affects child outcomes.[4]

2 Data and descriptive statistics

Our sample is drawn from the Fragile Family and Child Well-being Study (hereafter Fragile Families), a longitudinal survey of parents and children administered jointly by research groups at Princeton and Columbia Universities. Parents were surveyed at the time of the birth of the child, responding to a range of questions on relationship status, household structure, investments in children, and labor market status, and were re-surveyed at repeated intervals up to when the child turned 15. Importantly, at each wave the child was given tests to assess cognitive development and psychological evaluations, and there is information on the child's physical health. We will use the panel data structure to link together relationship status and investments in children with changes in children's outcomes over time. The surveys were done at birth (year 0), as well as at ages 1, 3, 5, 9, and 15. In the remainder of the paper, we refer to the birth year as period 0, and the remaining years as periods 1-5.

Table 1 shows some basic descriptive statistics. First looking at education, we see that this sample has relatively low levels of education. We split the sample into two groups, low (high school degree or less) and high (at least attended college) education. About a third of the sample is in the high education group. We next look at the racial composition of the sample; about a third of the respondents are white, half are black, and the remainder are other races. One unique feature of our data is that we have information on relationship quality, including, for example, information on physical or emotional abuse of the mother, as well as drug and alcohol use by the father. We create a variable that we call "bad relationship", which equals 1 if the father exhibits any of these behaviors and 0 otherwise. In

[4] For example, see Brien et al. (2006).

particular, utility from a marriage with a partner who displays any of these characteristics is likely lower, and it may also affect the child's outcomes. Almost half of the families have a value of 1 for the bad relationship variable.[5]

2.1 Relationship and employment transitions

In our model, we allow for the mothers to pick their marital status in each period, choosing from married, cohabiting, or other (which includes being single or being in a relationship but not living together, and we will refer to this state as "single" throughout for simplicity). In this section, we look at the transitions over these states in the data. We will show that movement between states is fairly common, illustrating that it is important to use a dynamic model. In addition, in comparison to previous work which does not separately look at cohabitation, we will show that this is fairly common in the data. Therefore it is important to look at how this outcome affects children separately.

In Table 2, we show the distributions of current period relationship status, conditioning on prior period relationship status. First, we see that almost 75% of our sample had parents that were unmarried when the child was born. Relationship states change significantly for mothers over time; for example, of the single households at period 0, almost 20% of the parents are cohabiting with the birth father in period 1. We continue to see transitions between states each period, meaning that children in these households experience transitions in their household structure over the first 15 years of their life. This demonstrates that our approach of using the whole history of marital status to understand evolution of child outcomes is important, because just the static relationship status is missing a significant part of differences across the experiences of children. Marriage and cohabitation seem fairly persistent; however we see that married parents are less likely to transition into the single state. For example, 5% of the couples who are married in period 0 are single by period 1, as compared to 30% of couples who are cohabiting in period 0. [6]

Next, Table 3 shows the results of a probit regression where the dependent variable equals 1 if a person is single in a period and is 0 otherwise. We control for the previous period relationship status, and see that people who are married or cohabiting are less likely to be single. Comparing the coefficients for married and cohabiting, we also see that people who were previously married are less likely to be single than those who were previously cohabiting. Mother's education decreases the likelihood of being single, and women are more likely to

[5]We defined this fairly loosely, especially by combining drug/alcohol abuse with physical/emotional abuse.

[6]We are only considering marriage between the child's birth parents. If a parent is married to someone else, this is part of the single category.

exit bad relationships.

Our model also includes employment decisions. We believe that this is a joint decision with relationship status, since whether or not you are married should affect your decision on whether or not to work. This could also impact child outcomes, for example by making it harder to spend time with your child. Table 4 looks at the percent of mothers who work each period, splitting the sample by relationship status. We do not see strong differences across relationship status. To look at this in more detail, in Table 5 we show the results of a probit regression where the dependent variable equals 1 if a mother works in a given period. We see evidence that married and cohabiting mothers work less than single mothers. As expected, employment decisions are strongly persistent, so mothers who worked in the previous period are more likely to work today. Mothers who are in bad relationships are less likely to work. Education increases the likelihood that a mother works, and we do not see a difference across races.

2.2 Investments in children

In our model, we allow for parents to invest in their children each period. This affects the child's quality in future periods. The Fragile Families data give information on how much time parents spend with their children doing different activities, which we use as a measure of investment. Here we look at how marital status correlates with parental investments in children. We run regressions for various measures of parental investments; the results are shown in Tables 14-17 in Appendix A. We do not see much of an effect of marital status on child investments. The strongest trend seems to be that mother's with more education spend more time with their children. We also see some evidence that mothers who work spend more time with their children.

2.3 Health outcomes

In our model, child quality evolves over time, as a function of marital status and parental investment. To measure quality, we need data on child outcomes. One component of this is the health of the child. The Fragile Families data report a measure of self-reported health a 1-5 scale. We run a regression where the dependent variable is self-reported health, and the results are in Table 18 in Appendix B. We see some evidence that children in married households have higher health measures, and mother's education also seems to increase the health outcomes. We also see differences across different racial groups. Being in a bad relationship is also associated with lower child's health outcomes.[7]

2.4 Academic outcomes

We also use academic outcomes to measure child quality. Over the different years of the survey, various academic tests were administered to the children. In particular, we use information on the Peabody Picture Vocabulary Test (PPVT) at ages 3, 5, and 9; Woodcock Johnson vocabulary tests at ages 5 and 9 (two tests at age 9); and the Spanish-language version of the PPVT that was administered to some non-native English speakers at age 3 (called the TVIP). For each child there may be a number of missing test scores. At age 15, children self-report their grades in Math, English, Science, and History high school classes. We run a series of regressions to understand the relationship between academic outcomes and relationship status. We use the child's percentile rank in the sample as the dependent variable. These results are shown in Tables 19-22 in Appendix C. We see that being married, as well as mother's education, are associated with significantly higher test scores and grades, and for most of the test scores children of cohabiting parents seem to only do as well as children of single parents. In period 5, where we see grades, we also see some evidence that children of cohabiting parents do better than those who are not living together.

3 Model

In each period, a mother decides on her relationship status and whether or not to work. We call this the RLM status in the remainder of the paper, and we denote it by s_t.[8] Her relationship options are to be married, cohabit with her partner, or be single, and her labor market choices are working and not working. The full choice set then has 6 elements: $S = \{\text{Marriage, Cohabitation, Single}\} \times \{\text{Work, Don't Work}\}$. We analyze this as a joint decision since it is likely that relationship and labor market decisions are affected by one another. Child outcomes evolve over time as a function of the mother's RLM status choice.

In each period $t \in \{1...T\}$, the state variables are her previous period status s_{t-1} and the child's quality at the start of the period $q_{t-1} \in \mathbb{R}$. The mother has a vector of exogenous

[7] We have looked at more objective measures of health (BMI, height, and weight), but did not see any interesting trends in the data so chose to not include them in the paper.

[8] For now, we focus on the mother's decisions and assume that she can unilaterally decide on relationship status. In future work we will expand this to allow for the father's preference/characteristics to affect outcomes.

observable characteristics denoted by X. The value function of the mother can be written recursively as:

$$\begin{aligned} V_t\left(s_{t-1}, q_{t-1}, X, \{\varepsilon_{st}\}\right) &= \max_{s_t \in S, I_t \in \mathbb{R}_+} U\left(s_{t-1}, s_t, q_{t-1}, X\right) - c\left(I_t\right) + \varepsilon_{st} \\ &+ \beta E\left[V_{t+1}\left(s_t, q_t, X, \left\{\varepsilon_{s(t+1)}\right\}\right)\right] \end{aligned} \tag{1}$$

$$q_t = H\left(I_t, s_t, q_{t-1}, X\right) \tag{2}$$

The first component in equation (1) is today's utility, which is a function of RLM status, child quality, and the mother's characteristics. In particular, we calculate a utility from each of the possible RLM transitions.[9] For example, we consider the net utility from going from married and working to married and not working. This setup combines the utility of being in a given state with the cost of transitioning between states. There is a cost of child investment, denoted as $c(I_t)$. There are extreme value shocks to the utility of different RLM statuses, denoted as ε_{st}, which are unobserved by the econometrican. Mothers also consider the future valuations with discount factor β. As stated above, investment has a cost today, but a benefit from increased child quality tomorrow. We assume that child quality evolves deterministically as a function of current child quality, RLM status, investment, and characteristics, given by the function $H(\cdot)$ in equation (2).

For each RLM status choice in each period, we can calculate the optimal investment level by solving the following optimization problem:

$$I^*\left(s_t, q_{t-1}, X\right) = \arg\max_{I_t} -c\left(I_t\right) + \beta E\left[V_{t+1}\left(s_t, H(I_t, s_t, q_{t-1}, X), X, \left\{\varepsilon_{s(t+1)}\right\}\right)\right] \tag{3}$$

In equation (3), we write next period child quality as $H(\cdot)$, to make clear how child investment impacts future valuations. The optimal investment decision comes from the first order condition. We can rewrite the value function, denoting the optimal investment as $I^*(\cdot)$.

$$\begin{aligned} V_t\left(s_{t-1}, q_{t-1}, X, \{\varepsilon_{st}\}\right) &= \max_{s_t \in S} U\left(s_{t-1}, s_t, q_{t-1}, X\right) - c\left(I^*\left(s_t, q_{t-1}, X\right)\right) + \varepsilon_{st} \\ &+ \beta E\left[V_{t+1}\left(s_t, q_t, X, \left\{\varepsilon_{s(t+1)}\right\}\right)\right], \end{aligned} \tag{4}$$

$$q_t = H\left(I^*\left(s_t, q_{t-1}, X\right), s_t, q_{t-1}, X\right) \tag{5}$$

To simplify notation, we rewrite the value function as follows:

[9] Since all but one of these parameters are identified, so we set the utility of staying married and working to 0.

$$
\begin{aligned}
V_t\left(s_{t-1}, q_{t-1}, X, \{\varepsilon_{st}\}\right) &= \max_{s_t \in S} v_t\left(s_{t-1}, s_t, q_{t-1}, X\right) + \varepsilon_{st} && (6)\\
v_t\left(s_{t-1}, s_t, q_{t-1}, X\right) &= U\left(s_{t-1}, s_t, q_{t-1}, X\right) - c\left(I^*\left(s_t, q_{t-1}, X\right)\right) && (7)\\
&+ \beta E\left[V_{t+1}\left(s_t, H\left(I^*\left(s_t, q_{t-1}, X\right), s_t, q_{t-1}, X\right), X, \left\{\varepsilon_{s(t+1)}\right\}\right)\right] && (8)
\end{aligned}
$$

Given the optimal level of investment in each RLM state, the problem becomes a dynamic discrete choice problem. Because the shocks to RLM status ε_{st} follow the extreme value distribution, we can solve for the closed form solution for the continuation values.

$$
E\left[V_t\left(s_{t-1}, q_{t-1}, X, \{\varepsilon_{st}\}\right)\right] = \log\left(\sum_{s_t \in S} \exp\left(v_t\left(s_{t-1}, s_t, q_{t-1}, X\right)\right)\right) + \gamma \qquad (9)
$$

where γ is Euler's constant, and we use the terminal conditional that $V_{T+1} \equiv 0$. We can also use the properties of the extreme value distribution to compute the probability that a person chooses a given RLM status in a period. This takes a logit form as follows:

$$
P\left(s_t | s_{t-1}, q_{t-1}, X\right) = \frac{\exp\left(v_t\left(s_{t-1}, s_t, q_{t-1}, X\right)\right)}{\sum_{z \in S} \exp\left(v_t\left(s_{t-1}, z, q_{t-1}, X\right)\right)} \qquad (10)
$$

For initial conditions, the mother has exogenous RLM status s_0, which we take from the data as the marital status and labor market status at the child's birth. We also need to know the initial level of child quality, which we assume is drawn from the normal distribution where the mean depends on the mother's characteristics. In time 0, while s_0 is exogenous, I_0 is chosen optimally given the assigned relationship status. We observe RLM decisions starting at time $t = 1, ..., T$.

4 Estimation

We estimate the model developed in section 3 using maximum likelihood. In this section, we first explain the parameterization of the model. In addition, the model assumes quality and investment are observed, whereas this is not something we can explicitly measure in the data. In section 4.2, we explain how we use data on different measurements on child's outcomes to learn about child quality. We do the same for investment in the child. We also derive the likelihood function in this section. Appendix D explains the identification of the choice parameters.

Each period, a mother picks her relationship status and labor market status. There are 3 relationship states (married, cohabiting, single) and 2 labor market outcomes (working, not working). We denote $s_t \in \{1, 2\}$ as the states for married, $s_t \in \{3, 4\}$ as the cohabiting states, and $s_t \in \{3, 4\}$ as the single states. For each of these pairs, the first number indicates the state where she is working, and the second is when she is not working. For example, $s_t = 1$ is a married woman who works, and $s_t = 2$ is a married woman who does not work.

4.1 Parameterization

Utility: the utility function is written as $U(s_{t-1}, s_t, q_{t-1}, X)$. We assume that utility is linear in relationship status and child quality at the start of the period, so

$$U(s_{t-1}, s_t, q_{t-1}, X) = k_{s_{t-1}, s_t} + q_{t-1} + \gamma_{s_{t-1}} I\{s_t \geq 5, br = 1\}.$$

The k_{s_{t-1}, s_t} are R^2 constants giving the base level of utility from one's current RLM status, as a function of your previous RLM status. Since there are six RLM options each period, this means we estimate 36 parameters that capture the net utility of moving between each pair of states. Since these are net utility parameters, they represent the utility of being in a state minus the moving cost of moving between states. In actuality, there are only 35 parameters because we normalize the utility of moving from being married and working in period $t-1$ to remaining married and working in period t to be 0. Utility also depends linearly on child quality. The last part of the utility function allows for there to be a utility gain to being single if you are in what we call a "bad relationship", which we include as part of the observables X. The variable $br = 1$ if a person is in a bad relationship. The term $\gamma_{s_{t-1}}$is the utility gain from being single if you are in a bad relationship. This term has a subscript s_{t-1} which reflects that we allow this parameter to depend on your previous marital status.[10] The inclusion of this parameter allows for different transition rates across relationship status depending on the quality of your relationship.

We assume a discount factor $\beta = 0.95$.

[10]Even though there are technically 6 RLM statuses, we just estimate 3 parameters: one for each relationship status (allowing the effect to be the same for those who are working and those who are not working.

Evolution of child quality: We assume that child quality evolves as follows

$$q_{t+1} = H\left(q_t, s_t, I_t, X\right) = q_t + X\beta_s + \delta_{1s} + \delta_{2s}\log\left(1+\exp\left(q_t\right)\right) + \delta_{3s}I_t. \tag{11}$$

First note that all of the parameters in equation (11) have an s subscript, meaning that we allow for the quality transitions to depend on RLM status. Child quality first evolves depending on demographics. In addition, there is a constant shifter δ_{1s}, as well as a parameter δ_{2s} that allows the change in quality to vary by the child's current quality. If $\delta_{2s} > 0$, higher-quality children have faster quality growth, while $\delta_{2s} < 0$ would give a "catch-up" effect where lower quality children have faster quality growth.

Cost of investment: We parameterize the cost of investment as

$$c\left(I_t\right) = I_t^2. \tag{12}$$

We cannot clearly identify whether heterogeneity across relationship states in investments is driven by differential costs or differential benefits, so we normalize the costs to be uniform across states and let the differences operate solely in benefits.

4.2 Measurement and likelihood

We observe relationship states without error in the data. However, we do not have direct observations of child quality q and investment I. Instead, we observe them with error. For quality, the data report test scores and health outcomes. For investment, we see information on how much time the mom spends with the child doing different activities.[11] We use this information to learn about total investment and quality.

We have M_t^q measures of quality at time t. For each measure j, we write

$$m_{jt}^q = \alpha_{jt}^1 + \alpha_{jt}^2 q_t + \alpha_{jt}^3\varepsilon,\ j = 1...M_t^q, t = 1...T,\ \varepsilon \sim \mathrm{N}\left(0,1\right). \tag{13}$$

In equation 13, we write that outcome j (for example, a test score) is a function of the child's quality. We can do this for each outcome measure we have in the data, where the parameters α differ by both test j and time t. There may be different numbers of tests per time period, and we do not assume that even same-name tests (e.g. Woodcock-Johnson academic tests) in different periods measure the same skills or measure skills with the same parameters.

[11] We have similar information on the dad, but am not using this in the current estimation.

We can do the same for investment

$$m_{jt}^{I}=\eta_{jt}^{1}+\eta_{jt}^{2}I_t+\eta_{jt}^{3}\varepsilon,\ j=1...M_t^I,t=1...T,\,\varepsilon\sim\mathrm{N}\left(0,1\right).$$

The simplest example of a measurement model is if we know a priori that $\alpha_{jt}^2=1$ for all j and t. This would be the case, for example, if we knew that our measurements were simply white noise around a true value; it may make sense to treat multiple measurements of height like this. Then, as is well known, the best way to combine measurements would be to take a weighted average of the different measurements with weights inversely related to the variance of the difference measures. The more general measurement model uses the same idea, but uses information about both the variances of measurements as well as the covariances across measurements to recover the weights the tests put on the underlying latent factor. Our approach is a full-information one, using the normality assumptions explicitly within a Maximum Likelihood framework.

Our full data of child characteristics consists of M_t^q measures of quality and M_t^I measures of investment for each period t. Given the parameters, relationship decisions, and known initial q_0, we can predict quality and investment for each individual. In particular, child investment is a deterministic function of one's relationship state and previous child quality, and quality also evolves deterministically as a function of investment, relationship status, and previous quality. This also allows us to predict each quality and investment measure, which we denote as $\hat{m}_{jt}^q$ and $\hat{m}_{jt}^I$. In particular,

$$\begin{aligned}\hat{m}_{jt}^{q} &= \alpha_{jt}^{1}+\alpha_{jt}^{2}q_t\\ \hat{m}_{jt}^{I} &= \eta_{jt}^{1}+\eta_{jt}^{2}I_t\end{aligned}$$

We also know that the standard deviation of each quality measure is α_{jt}^3 and the standard deviation of each investment measure is η_{jt}^3.

For each person, we write the time t likelihood as

$$\begin{aligned}&L_t\left(s_t,\left\{m_{jt}^q\right\}_{j=1}^{M_t^q},\left\{m_{jt}^I\right\}_{j=1}^{M_t^I}|s_{t-1},q_{t-1},X\right)= \\ &\Pr\left(s_t|s_{t-1},q_{t-1},X\right)\cdot\prod_{j=1}^{M_t^q}\frac{1}{\alpha_{jt}^3}\phi\left(\frac{m_{jt}^q-\hat{m}_{jt}^q}{\alpha_{jt}^3}\right)\cdot\prod_{j=1}^{M_t^I}\frac{1}{\eta_{jt}^3}\phi\left(\frac{m_{jt}^I-\hat{m}_{jt}^I}{\eta_{jt}^3}\right)\end{aligned}\tag{14}$$

Equation (14) gives the likelihood of seeing a given relationship choice and outcome measures in the data. The first component is the probability the mother chooses a given relationship. This has a logit form and was derived in equation (10). In the second part of the likelihood function, we use the fact that with known quality q, the measurements are independent of everything else. Furthermore, given an initial q_0 and a path of relationship choices, q is known. Therefore, we can write the likelihood of seeing each quality measure given the known quality, since the measurement outcomes follow the normal distribution with standard deviation α^3_{jt} (where we write the pdf of the standard normal distribution as $\phi(\cdot)$). Since investment also is given deterministically by the model, we can do the same thing for investment measures.

We can then derive the likelihood over all periods for each person, initial on q_0, which we write as $\tilde{L}\left(s, m^q, m^I | q_0\right)$. We are using s to denote the series of relationship statuses, m^q to denote the series of quality measures, and m^I to denote the series of investment measures. This equation is written as follows:

$$\tilde{L}\left(s, m^q, m^I | q_0, X\right) = \prod_{t=1}^{T} L_t\left(s_t, \left\{m^q_{jt}\right\}_{j=1}^{M^q_t}, \left\{m^I_{jt}\right\}_{j=1}^{M^I_t} | s_{t-1}, q_{t-1}, X\right) \tag{15}$$

Equation (15) is conditional on both initial quality and initial relationship status (the latter to derive the choice probability for the first period). The initial relationship quality is taken from the parent's marital status when the child is born. We assume that initial quality is drawn from the normal distribution, with a mean that is a function of the mother's characteristics and a standard deviation of 1. We can then integrate out initial quality q_0 to find the log likelihood for each person.

$$\mathcal{LL}\left(s, m^q, m^I\right) = \log\left[\int_{q_0=-\infty}^{\infty} \tilde{L}\left(s, m^q, m^I | q_0\right) f\left(q_0\right) dq_0\right]$$

We numerically evaluate the integral with a Gaussian quadrature rule.

5 Results

In this section, we explain our results. In the current version of this paper, we do not use the data on investments in children.[12] Therefore mothers only decide on marital status each period, and child quality evolves as a function of previous quality and RLM status.

Table 6 shows the utility transition parameters over marital and labor market. Recall

[12]We are currently working to include investment in the estimation, and it will be included in future versions of the paper.

that there are 6 states, which include choices over marital status and working. We report the net utility from transitioning between each pair of states. Because these are only identified relative to each other, we normalize the utility from starting the period as married and working and then staying in that state at 0. Therefore each of the other utilities are relative to that state. Utility also depends on whether or not one is in a relationship with a partner who demonstrates bad behaviors, including abuse or heavy alcohol or drug use. We allow for the utility of being single to vary if you are in what we call a "bad relationship." We allow for this utility shift to depend on your prior relationship status. These parameter estimates are shown in Table 7. In all cases, the utility from being single is higher for people in a bad relationship. This means, that conditional on all other factors, these people are more likely to leave a marriage or exit from cohabitation.

Table 8 shows the parameter estimates for initial child quality, which we assume is drawn from a normal distribution with a mean that depends on one's characteristics and has a standard deviation of 1. For race, the excluded group is whites, showing that white children have higher initial quality draws. Mother's education also increases initial quality.

Next, we look at the evolution of child quality, as given in equation (11). Quality shifts by a constant, as well as a term that depends on your previous quality. These parameters differ based on previous marital and working status. Because these terms are only defined relative to one another, we assume that the parameters for quality for children whose mothers are single and not working are set to 0. The evolution of quality also depends on demographics. The only demographics we use are education and relationship status (each of which have 2 possible options), so we divide the sample into 4 groups to capture each possible combination. We then estimate 3 parameters to allow for different growth rates for each demographic group. To demonstrate the magnitude of these parameters, in Figure 1 we show the path of quality, conditional on demographics, assuming that a person starts at the mean value given by their characteristics. We show this for each RLM state (assuming it stays constant in each period) to compare the effect of marital status and mother's employment on child outcomes. Interestingly, the relative rankings of each relationship status vary based on demographics. For example, for moms with a low education and a bad relationship, we see that the child outcomes are always higher when the mom is not working. On the other hand, when the mom has a high education, the two highest outcomes are when the mom is married, regardless of her labor market status. To understand the average effects, in Figure 2 we plot the average path (over demographics) for each RLM state. This shows that children of married and not working mothers seem to do the best, and the children of single working mothers do the worst over time.

In Figure 1, we just looked at the transition path for a given initial value of quality.

Remember that initial quality values are drawn from the normal distribution. In Figure 3, we show the distribution of initial qualities for each set of demographic characteristics. In this figure, all of the plots are the same across RLM status because the initial quality only depends on demographics. Comparing across demographics, we see that relationship quality has a small effect on the initial quality distribution, but education shifts the distribution to the right. To show how RLM status affects quality over the lifetime, Figure 4 shows the distribution at period 5.

The last set of results shows the measurement parameters. Remember that quality is unobserved; we use different outcomes to infer about one's quality. We have self-reported health in each period (reported on a 1-5 scale), test scores in periods 2-4, and grades in period 5. A strength of our approach is that we can combine all of these factors to learn about quality, and let the variation in the data tell us the relative importance of each one (as compared to simply taking an average, which would be difficult given that these are all measured on different scales). These results are shown in Table 10. To interpret these parameters, we plot the signal to noise ratio for each measure in Figure 5. The signal to noise ratio is defined as α^2/α^3 for each measure. To understand this, assume q follows the standard normal distribution. If there were no noise, α^2 would be the variance of m. The term α^3 is the variance of the error term. Therefore the ratio of these 2 terms tells us how much of the variance in the test score comes from quality versus noise. Higher values indicate a more informative test. We see that test scores are far more informative than health outcomes. Grades do slightly better than health, but not nearly as well as test scores. Furthermore, there is a fair amount of heterogeneity within the different tests, suggesting there is value in our methodology instead of just taking an average (which would put equal weight on each test score).

5.1 Model fit

To assess the model fit, we compare the choice probabilities in the data and the model. In particular, for each period, we show the probability of transitioning from a given state into another. This is shown in Figures 6-10, which demonstrate that our model does a good job of fitting the data.

6 Counterfactuals

In this section, we perform counterfactuals to try to understand the implications of policies that help to support family formation and stability. The counterfactuals are compared with of a baseline set of RLM choices and child qualities, simulated from the model at our estimated parameters. To generate the counterfactuals, we simply adjust the estimated parameters and re-simulate, holding the unobserved shocks to utility and measurement errors constant.

One important thing to note for the counterfactuals is that we cannot identify the absolute level of child qualities, but only relative to some reference group, since all our child quality measurements have no objective location or scale. In particular, our reference group at time 0 is low education white women who have a "bad relationship" (as defined above) with the birth father. This group is normalized to have initial child quality distributed Normal(0,1), and other groups have mean qualities determined by their demographics. In later periods, the reference group is a women with the aforementioned demographics who has been single and not working every period. All other quality changes are with reference to this group, e.g., saying a particular child's quality went from 0 to 1 is equivalent to saying that the child went from the 50th percentile to the (approximately) 67th percentile relative to the children of the mothers in the reference group. In the counterfactuals, we will use the baseline distribution of child qualities in each period as a comparison group for the counterfactual children.

In the first counterfactual (hereafter referred to as counterfactual 1), we increase the utility from being married, conditional on being married, by 0.5. This change can be interpreted as any policy that would lead to lower divorce rates but does not directly affect the ability of parents to raise their children. Lower rates of exiting marriage has a secondary effect on marriage formation, since individuals know that the lifetime value of being married will increase if they choose that path. Table 11 shows the percent of households that are married, cohabiting, and single in the baseline and counterfactual scenarios. We see that this leads to a large increase in marriage: in period 5, 45% of the couples are married, as compared to 25% in the baseline. Our second counterfactual (counterfactual 2) increases the utility when you transition from a non-married to a married state by 0.5. Potential real-world interpretations of this counterfactual include increased societal pressure for couples with children to marry, or financial incentives for marriage (to the extend the additional income wouldn't directly affect child quality.) Counterfactual 2 has a direct impact on the cohabiting to marriage and the single to marriage transitions, with a minor indirect effect of increasing transitions out of marriage since it is easier to re-enter in the future. This again has large increases on marriage rates: 25% and 50% of the sample is married in the baseline and counterfactual, respectively.

The increased utility from marriage in the counterfactuals does not have a direct effect on child quality evolution, holding choices constant, but instead induced changes in child quality via changes in optimal relationship/labor market choices. As more mothers choose to be married, our estimated parameters indicate that average child quality will increase relative to their previous choices. In Figure 11, we plot the difference between average child quality in the baseline and in each counterfactual. The counterfactual raises child quality, and the effects increase over time, which is expected. Counterfactual 2, making marriage more desirable, has a marginally larger effect than counterfactual 1, but both lead to similar levels of child quality growth. To interpret the numbers, we calculate the standard deviation of child quality in period 5 (age 15) to be 1.64. Assuming the distribution in of quality in period 5 is still approximately normal, an increase of 0.08 means the median child would move up in the baseline quality distribution by 0.05 standard deviations if he or she was the only one affected by the counterfactual. This corresponds to moving up about 2% in the quality distribution at the median, although the effects would not be uniform across the distribution. Regardless, given the large effect of the counterfactual on marriage rates, this effect on child quality seems an order of magnitude smaller.

Additionally, while child quality increases in the counterfactuals, we can use our data and model to also see one potential downside of such a policy. Recall that in our data, we created a variable called "bad relationship" that indicated whether or not one's partner demonstrated any dangerous behavior, including physical or emotional abuse of the mother, or heavy alcohol of drug use. Our counterfactuals lead to increased marriage rates and higher child outcomes. An added effect of this increase in marriage rates is that the counterfactuals also lead to a greater share of marriage or cohabiting relationships that exhibit one of these bad indicators. In the terminal period, 23% of the married or cohabiting relationships are defined as bad in the baseline, compared to 35% in the counterfactual 1 and 38% in counterfactual 2. While mothers are still choosing optimally within the logic of the model so we cannot evaluate any sort of direct welfare loss, it seems plausible for a policymaker to be concerned about increasing the incidence of dangerous relationships.

7 Conclusion

In this paper, we study how family instability causally affects child outcomes, focusing on cognitive and physical development. We develop and estimate a dynamic model of marriage and child development, where parental marital status affects the evolution of child outcomes over time. We use our results to quantify how much child outcomes would improve if there were policies put in place to encourage parents to stay together. In a counterfactual, we raise the utility from marriage so that there is a 10 percentage point increase in the share of parents that are married in period 1. This increases our measure of child outcomes at age 15 by about 0.05 standard deviations.

References

Agostinelli, F. (2018). Investing in children's skills: an equilibrium analysis of social interactions and parental investments.

Agostinelli, F. and Wiswall, M. (2016). Estimating the technology of children's skill formation. NBER working paper n. 22442.

Amato, P. R. and Sobolewski, J. M. (2001). The effects of divorce and marital discord on adult children's psychological well-being. *American Sociological Review*, 66 (6):900–921.

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.

Craigie, T.-A., Brooks-Gun, J., and Waldfogel, J. (2012). Family structure, family stability and outcomes of five-year-old children. *Families, Relationships, and Societies*, 1 (1):43–61.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78 (3):883–931.

Hayward, M. D. and Gorman, B. K. (2004). The long arm of childhood: the influence of early-life social conditions on men's mortality. *Demography*, 41(1):87–107.

Lang, K. and Zagorsky, J. L. (2001). Does growing up with a parent absent really hurt? *The Journal of Human Resources*, 36 (2):253–273.

Maier, E. H. and Lachman, M. E. (2000). Consequences of early parental loss and separation for health and well-being in midlife. *International Journal of Behavioral Development*, 24 (2):183–189.

Ross, C. E. and Mirowsky, J. (1999). Parental divorce, life-course disruption, and adult depression. *Journal of Marriage and the Family*, 61 (4):1034–1045.

Tartari, M. (2015). Divorce and the cognitive achievement of children. *International Economic Review*, 56 (2):597–645.

Tucker, J. S., Friedman, H. S., Schwartz, J. E., Criqui, M. H., Tomlinson-Keasey, C., and Wingard, D. L. (1997). Parental divorce: Effects on individual behavior and longevity. *Journal of Personality and Social Psychology*, 73(3):381–391.

Tables and figures

Table 1: Descriptive statistics

Percent of mothers with high education	36.09%
Race: white	31.20%
Race: black	51.35%
Race: other	17.45%
Bad relationship	45.18%

High education defined as some college or more.

Table 2: Relationship transitions

		Current relationship status			
	Prior relationship status	Married	Cohabiting	Other	Number of observations
Period 0 to period 1	Married	94.21%	1.02%	4.77%	985
	Cohabiting	14.52%	54.59%	30.89%	1460
	Other	3.98%	17.76%	78.26%	1610
Period 1 to period 3	Married	90.20	1.53	8.27	1112
	Cohabiting	13.84	53.78	32.37	1004
	Other	3.19	10.19	86.63	1600
Period 3 to period 5	Married	85.92%	1.75%	12.33%	1087
	Cohabiting	15.89%	48.56%	35.55%	661
	Other	2.12%	5.04%	92.84%	1648
Period 5 to period 9	Married	81.44%	1.71%	16.86%	878
	Cohabiting	19.76%	45.13%	35.10%	339
	Other	2.14%	4.68%	93.18%	1495
Period 9 to period 15	Married	78.97%	0.97%	20.06%	718
	Cohabiting	11.44%	45.77%	42.79%	201
	Other	1.37%	1.30%	97.32%	1382

Table 3: Probit regression on relationship status

	(1)
	Dependent variable = 1 if single in current period
Married in prior period	-0.915***
	(0.0133)
Cohabiting in prior period	-0.614***
	(0.0115)
Employment status	0.0438***
	(0.0104)
Bad relationship	0.130***
	(0.0100)
High education (mom)	-0.0586***
	(0.0110)
White	0.0152
	(0.0149)
Black	0.124***
	(0.0137)
Period	0.0214***
	(0.00117)
Observations	16180

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report marginal effects from a probit regression. High education for the mother means at least some college.

Table 4: Relationship and employment outcomes

	Married		Cohabiting		Single	
Period	Percent working	N	Percent working	N	Percent working	N
1	56.40%	1204	54.16%	1093	52.96%	1758
2	56.92%	1193	54.58%	720	58.68%	1803
3	60.06%	1074	58.16%	423	61.14%	1899
4	67.58%	814	59.66%	238	62.77%	1660
5	77.18%	609	68.38%	117	70.16%	1575

Table 5: Probit regression on employment decisions

	(1)
Married	-0.0177*
	(0.0103)
Cohabiting	-0.0192*
	(0.0116)
Employed in previous period	0.332***
	(0.00849)
Bad relationship	-0.0401***
	(0.00833)
High education (mom)	0.152***
	(0.00897)
White	0.00488
	(0.0121)
Black	0.0182
	(0.0113)
Period	0.0170***
	(0.000921)
Observations	16180

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the marginal effects from a probit regression. High education means the mom at least attended college.

Table 6: Utility parameters

		Choices					
		Married		Cohabiting		Single	
Prior state		Working	Not working	Working	Not working	Working	Not working
Married	Working	0	-2.29	-4.46	-6.21	-1.92	-3.66
			(0.12)	(0.22)	(0.40)	(0.10)	(0.16)
	Not working	0.02	-0.08	-4.60	-3.87	-1.80	-2.24
		(0.15)	(0.07)	(0.43)	(0.33)	(0.18)	(0.19)
Cohabiting	Working	-1.55	-3.37	-0.70	-2.05	-0.39	-1.64
		(0.20)	(0.23)	(0.08)	(0.23)	(0.18)	(0.19)
	Not working	-1.33	-1.84	-0.61	-0.75	-0.33	-0.55
		(0.27)	(0.27)	(0.25)	(0.09)	(0.24)	(0.24)
Single	Working	-4.14	-5.87	-3.39	-4.65	-0.11	-1.45
		(0.16)	(0.22)	(0.16)	(0.22)	(0.06)	(0.09)
	Not working	-4.22	-4.74	-3.9027254	-3.83	-0.34	-0.43
		(0.23)	(0.22)	(0.21)	(0.22)	(0.10)	(0.07)

Table 7: Bad relationship utility parameters

Prior relationship status	Utility increase when single if in "bad relationship"
Married	1.33
	(0.11)
Cohabiting	0.70
	(0.09)
Single	0.18
	(0.05)

Table 8: Determinants of initial quality

	Estimate
Bad relationship	0.015
	(0.08)
Black	-0.55
	(0.05)
Other race	-0.57
	(0.06)
High education	1.10
	(0.11)

We assume the initial quality is drawn from a normal distribution, where the mean is a function of the characteristics in this table and the standard deviation is equal to 1.

Table 9: Child quality transition parameters

	Married		Cohabiting		Single	
	Working	Not working	Working	Not working	Working	Not working
Constant	-0.24	-0.068	-0.17	0.065	-0.29	0
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	
Previous quality	0.18	0.15	0.13	0.01	0.17	0
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	
Bad relationship, high education	0.00	-0.12	-0.01	-0.23	-0.09	-0.22
	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)
Good relationship, low education	0.07	0.06	0.10	0.01	0.05	0.01
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Good relationship, high education	0.08	-0.03	0.07	-0.25	0.00094	-0.18
	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)

We assume the constant and previous quality terms are equal to 0 for single mothers who do not work as a normalization.

Table 10: Measurement parameters

Period	Measurement	Constant term	Coefficient on quality	Standard deviation
1	Health	4.50	0.15	0.78
		(0.02)	(0.01)	(0.01)
2	TVIP	99.06	2.60	11.91
		(1.45)	(1.04)	(0.83)
	PPVT	86.26	7.44	13.65
		(0.57)	(0.31)	(0.21)
	Health	4.48	0.15	0.74
		(0.02)	(0.01)	(0.01)
3	PPVT	94.71	8.52	10.71
		(0.64)	(0.26)	(0.16)
	WJ22	51.08	11.18	24.12
		(0.94)	(0.58)	(0.53)
	Health	4.51	0.14	0.71
		(0.02)	(0.01)	(0.01)
4	PPVT	94.08	8.40	9.08
		(0.66)	(0.23)	(0.17)
	WJ9	38.16	12.83	16.51
		(1.05)	(0.43)	(0.27)
	WJ10	49.42	14.18	20.16
		(1.15)	(0.52)	(0.39)
	Health	4.40	0.12	0.79
		(0.03)	(0.01)	(0.01)
5	English grades	2.95	0.17	0.82
		(0.03)	(0.02)	(0.02)
	Math grades	2.79	0.17	0.91
		(0.03)	(0.02)	(0.02)
	History grades	2.96	0.23	0.85
		(0.03)	(0.02)	(0.02)
	Science grades	2.89	0.18	0.88
		(0.03)	(0.02)	(0.02)
	Health	4.35	0.11	0.81
		(0.03)	(0.01)	(0.01)

Table 11: Counterfactual 1: Increased utility from staying in marriage

	Married		Cohabiting		Single	
Period	Baseline	Counterfactual	Baseline	Counterfactual	Baseline	Counterfactual
0	23.7%	23.7%	36.6%	36.6%	39.7%	39.7%
1	30.1%	42.3%	26.4%	22.1%	43.6%	35.6%
2	32.8%	50.4%	19.0%	14.6%	48.3%	35.0%
3	33.1%	53.4%	13.2%	9.6%	53.7%	62.4%
4	30.5%	51.5%	9.8%	6.8%	59.8%	42.2%
5	26.1%	45.5%	5.1%	3.9%	68.9%	50.6%

We increase the utility from remaining into marriage by 0.5. Each cell reports the fraction of the sample in a given relationship status.

Table 12: Counterfactual 2: Increased utility from transitioning to marriage

	Married		Cohabiting		Single	
Period	Baseline	Counterfactual	Baseline	Counterfactual	Baseline	Counterfactual
0	23.7%	23.7%	36.6%	36.6%	39.7%	39.7%
1	30.1%	46.1%	26.4%	21.2%	43.6%	32.7%
2	32.8%	56.5%	19.0%	12.2%	48.3%	31.4%
3	33.1%	59.7%	13.2%	7.8%	53.7%	32.5%
4	30.4%	58.3%	9.8%	5.2%	59.8%	36.5%
5	26.1%	51.5%	5.1%	2.9%	68.9%	45.6%

We increase the utility from transitioning into marriage by 0.5. Each cell reports the fraction of the sample in a given relationship status.

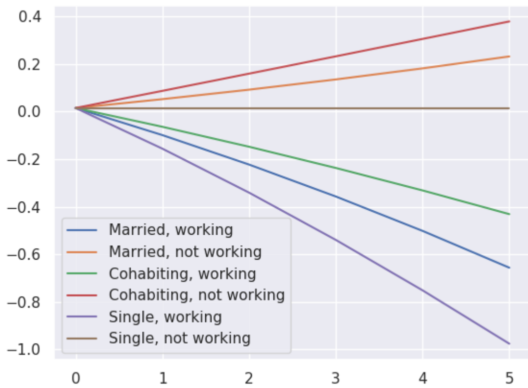
Table 13: Counterfactual effect on bad relationships

Period	Baseline	Counterfactual 1	Counterfactual 2
0	35.3%	35.3%	35.3%
1	35.0%	39.4%	40.1%
2	33.3%	40.9%	42.3%
3	30.5%	40.5%	42.4%
4	27.8%	38.8%	41.2%
5	23.2%	35.0%	38.0%

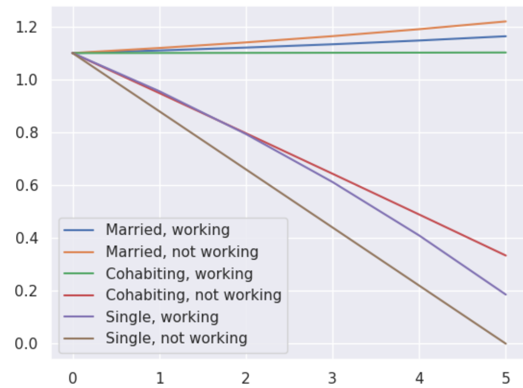
For each counterfactual, we report the fraction of married or cohabiting “bad relationships.

Figure 1: Evolution of child quality

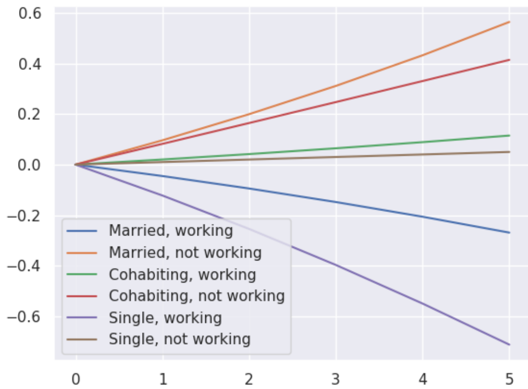
(a) Bad relationship, low education



(b) Bad relationship, high education



(c) Good relationship, low education



(d) Good relationship, high education

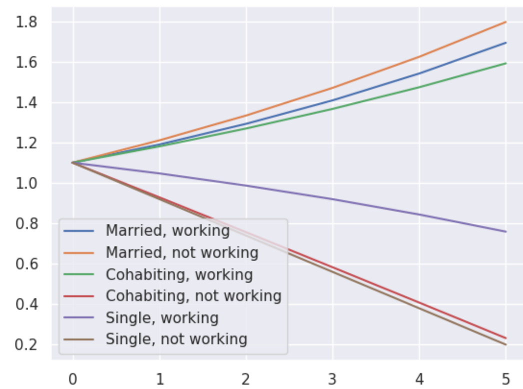


Figure 2: Average evolution of child quality

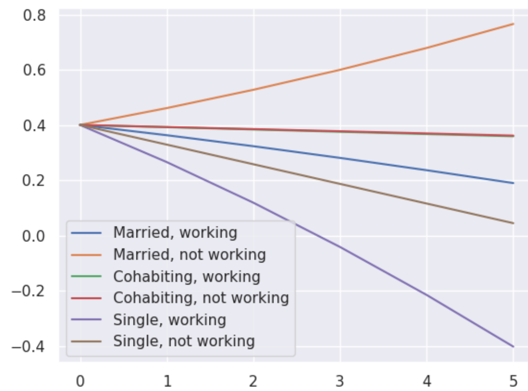
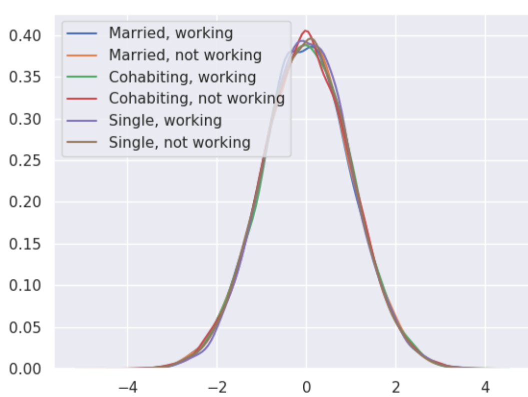
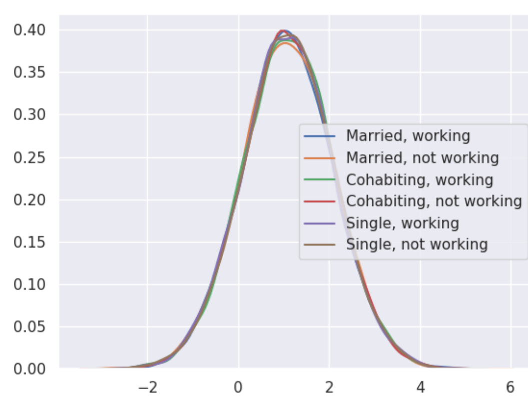


Figure 3: Distribution of initial quality

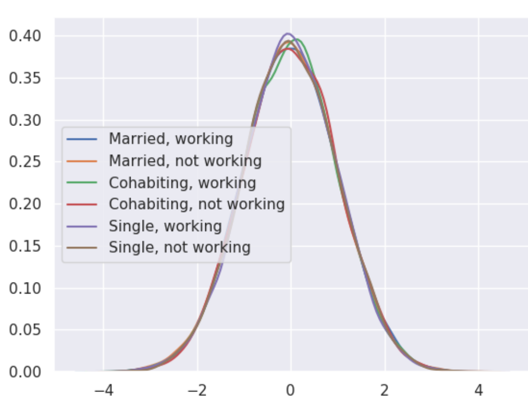
(a) Bad relationship, low education



(b) Bad relationship, high education



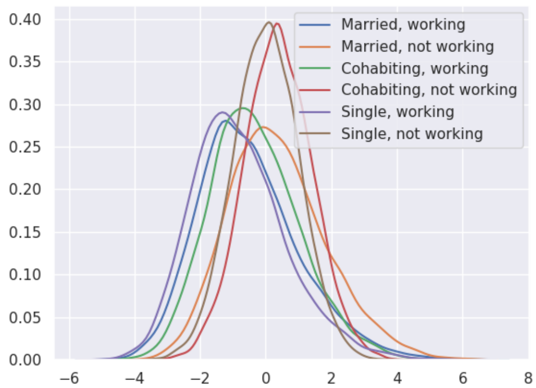
(c) Good relationship, low education



(d) Good relationship, high education

Figure 4: Distribution of quality in period 5

(a) Bad relationship, low education



(b) Bad relationship, high education

(c) Good relationship, low education

(d) Good relationship, high education

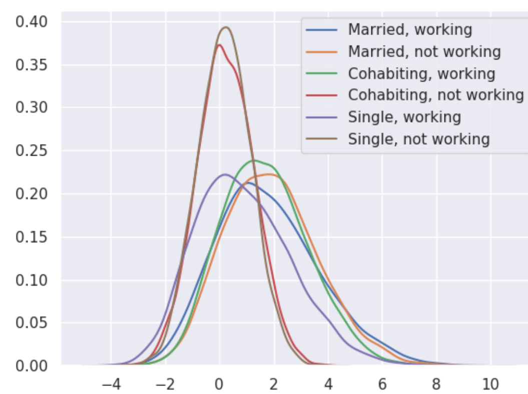


Figure 5: Signal to noise ratio

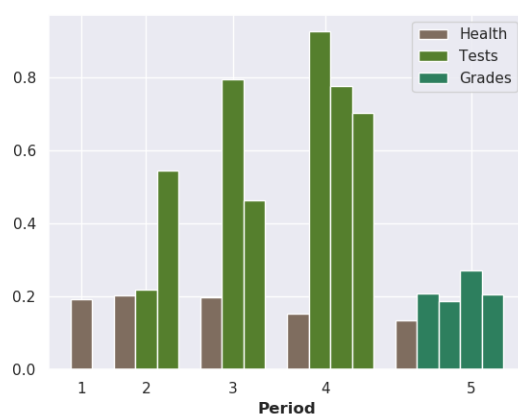


Figure 6: Model fit, period 1

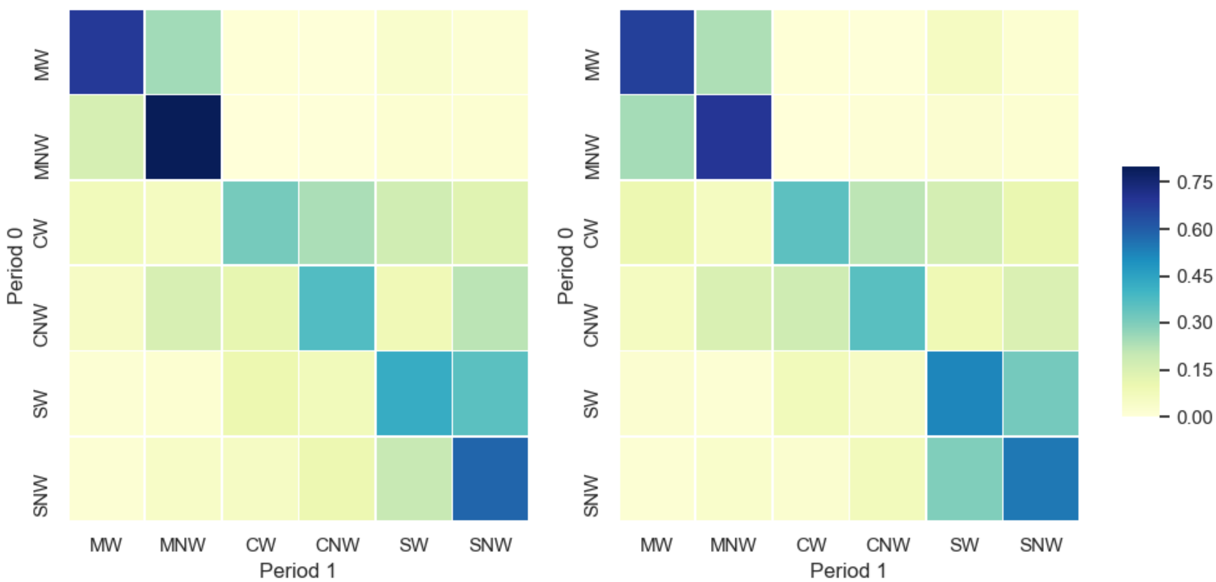


Figure 7: Model fit, period 2

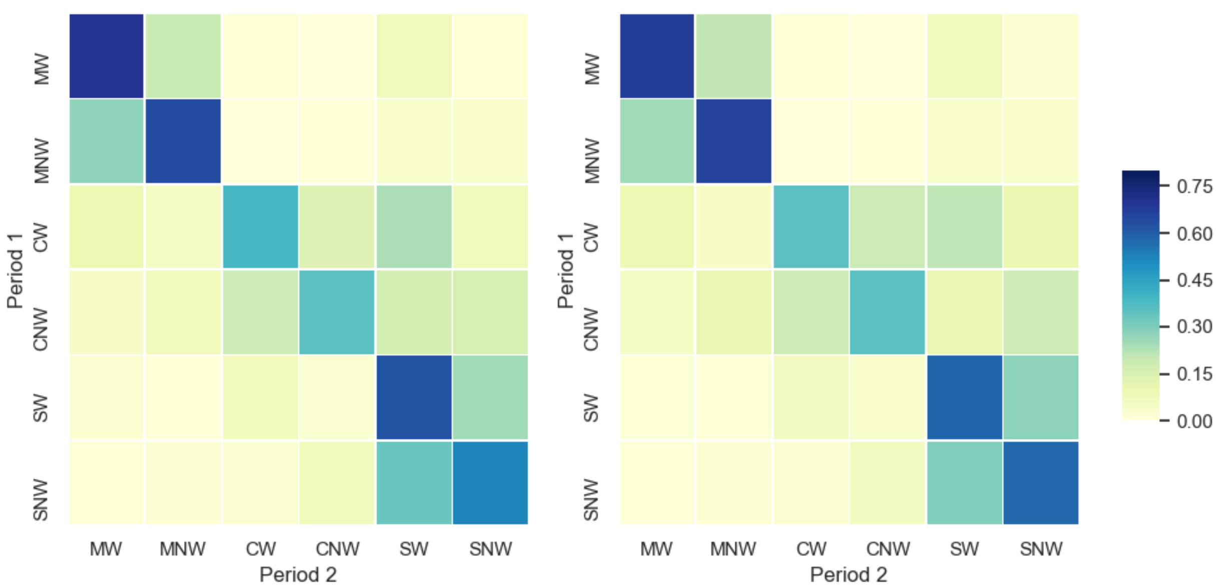


Figure 8: Model fit, period 3

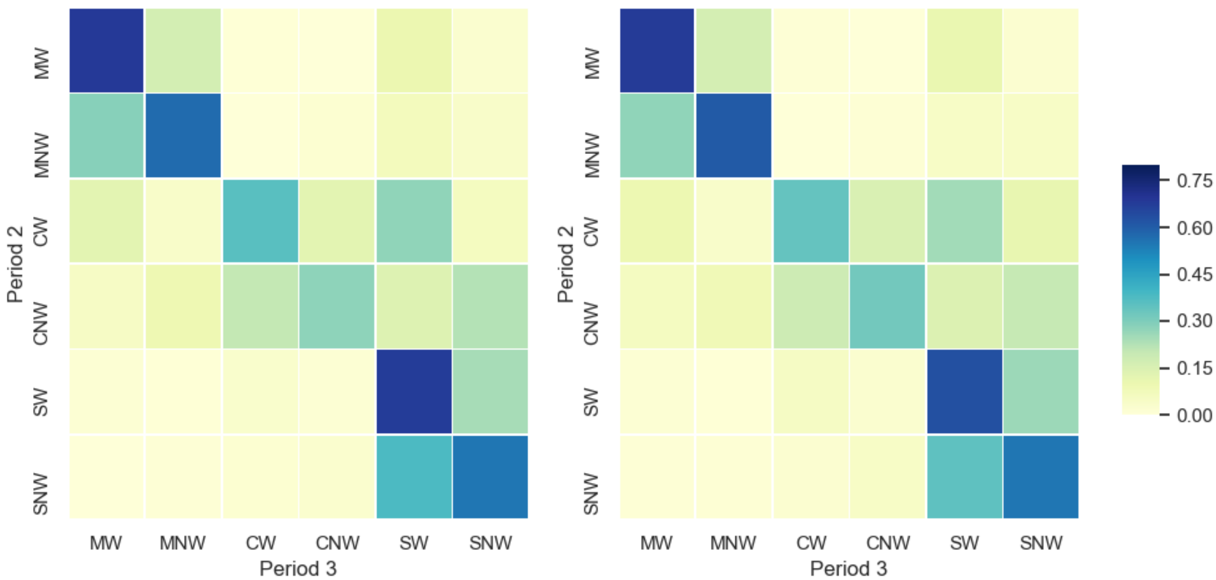


Figure 9: Model fit, period 4

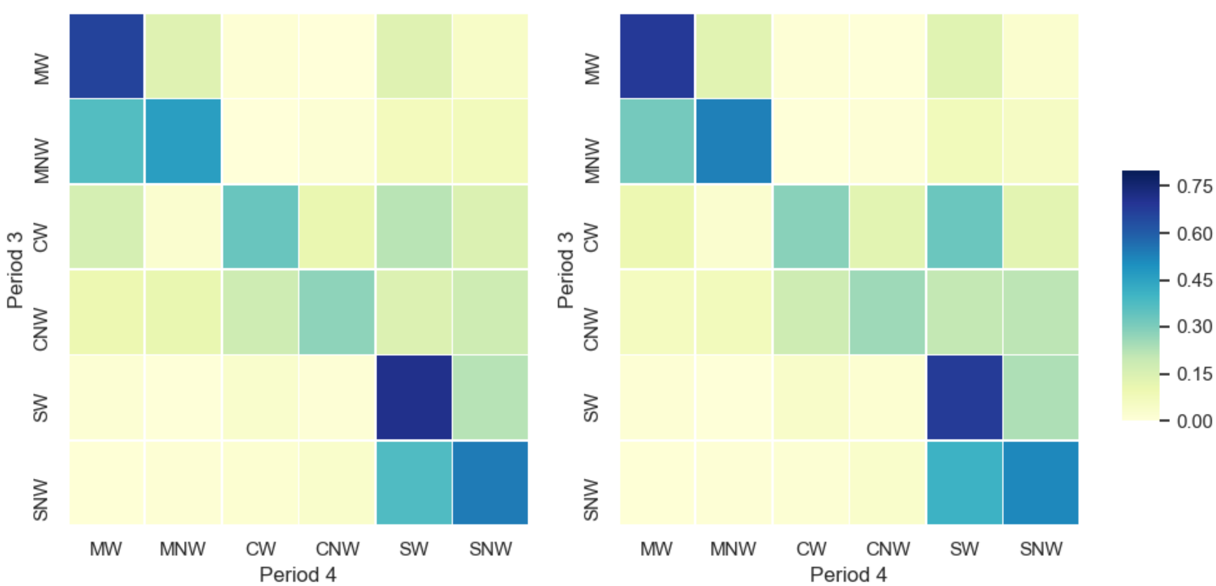


Figure 10: Model fit, period 5

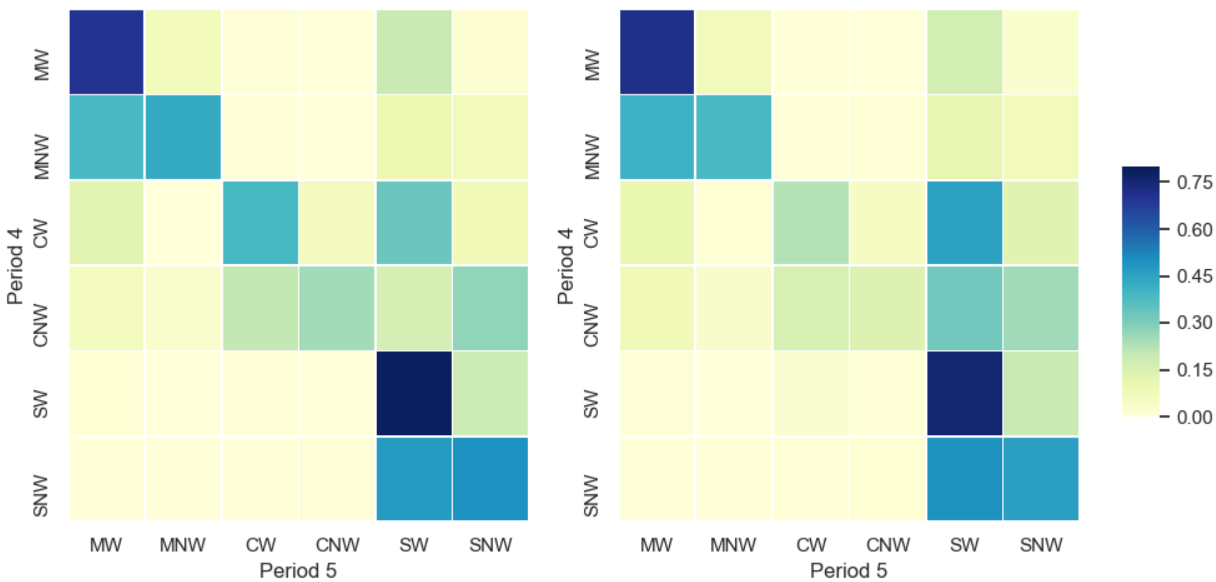
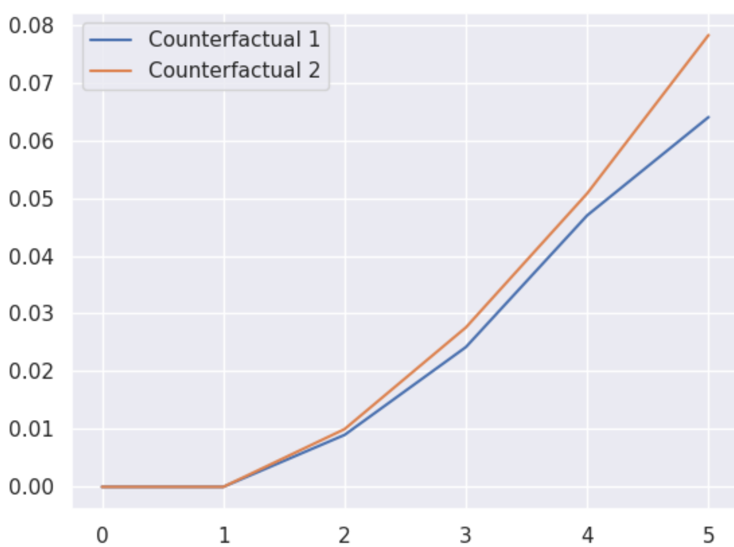


Figure 11: Counterfactual effects on child quality



Appendix A: Reduced form evidence- child investments

Table 14: Child investments, period 1

	(1) Games	(2) Sings	(3) Read stories	(4) Tell stories	(5) Play
Married	-0.0475	0.121	0.175*	-0.100	-0.0963
	(0.0703)	(0.0849)	(0.0975)	(0.113)	(0.0831)
Cohabiting	-0.0440	0.0347	0.0141	-0.0651	0.0814
	(0.0654)	(0.0790)	(0.0907)	(0.105)	(0.0769)
Employment status	-0.214***	-0.148**	-0.228***	-0.314***	-0.230***
	(0.0544)	(0.0657)	(0.0754)	(0.0873)	(0.0642)
Bad relationship	-0.111**	-0.198***	-0.285***	-0.312***	-0.0474
	(0.0546)	(0.0659)	(0.0757)	(0.0875)	(0.0644)
High education (mom)	0.365***	0.442***	0.515***	0.455***	0.264***
	(0.0598)	(0.0722)	(0.0829)	(0.0956)	(0.0703)
White	0.326***	0.267***	0.893***	0.528***	0.386***
	(0.0785)	(0.0948)	(0.109)	(0.128)	(0.0939)
Black	0.0647	-0.0584	0.741***	0.297**	-0.0838
	(0.0736)	(0.0888)	(0.102)	(0.119)	(0.0873)
Constant	5.972***	5.485***	3.536***	3.659***	5.867***
	(0.0831)	(0.100)	(0.115)	(0.133)	(0.0982)
Observations	4021	4026	4025	3564	3565

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are the number of days per week the mother spends doing a given activity with her child when the child is 1 year old. High education means the mother attended at least some college.

Table 15: Child investments, period 2

	(1)	(2)	(3)	(4)	(5)
	Sings	Games	Read stories	Tell stories	Play
Married	-0.00481	-0.0453	-0.00735	-0.0655	-0.169*
	(0.0888)	(0.100)	(0.0869)	(0.105)	(0.0880)
Cohabiting	-0.0766	0.205*	-0.121	0.101	0.0604
	(0.0946)	(0.107)	(0.0925)	(0.112)	(0.0937)
Employment status	-0.0615	-0.210***	-0.192***	-0.211**	-0.345***
	(0.0719)	(0.0813)	(0.0703)	(0.0852)	(0.0712)
Bad relationship	-0.190***	-0.187**	-0.263***	-0.204**	-0.0733
	(0.0723)	(0.0818)	(0.0707)	(0.0857)	(0.0716)
High education (mom)	0.506***	0.435***	0.502***	0.247***	0.174**
	(0.0778)	(0.0880)	(0.0761)	(0.0922)	(0.0771)
White	0.382***	0.490***	0.696***	0.372***	0.447***
	(0.104)	(0.118)	(0.102)	(0.124)	(0.103)
Black	0.161	-0.0387	0.342***	0.0183	0.165*
	(0.0984)	(0.111)	(0.0963)	(0.117)	(0.0975)
Constant	5.018***	4.548***	4.953***	4.523***	5.514***
	(0.113)	(0.128)	(0.111)	(0.134)	(0.112)
Observations	3666	3655	3666	3660	3664

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the number of days per week a mother spends doing a given activity with her child when the child is 3 years old. High education means the mother at least attended college.

Table 16: Child investments, period 3

	(1) Sings	(2) Read stories	(3) Tell stories	(4) Play inside	(5) Play outside
Married	0.0426	0.0968	0.193*	-0.0535	0.240***
	(0.0967)	(0.0886)	(0.103)	(0.0991)	(0.0922)
Cohabiting	0.0576	-0.0691	0.124	0.0775	0.219*
	(0.123)	(0.112)	(0.131)	(0.126)	(0.117)
Employment status	-0.206**	-0.332***	-0.255***	-0.226***	-0.379***
	(0.0812)	(0.0744)	(0.0865)	(0.0832)	(0.0775)
Bad relationship	0.0890	-0.244***	0.0830	0.133	0.0687
	(0.0812)	(0.0744)	(0.0865)	(0.0833)	(0.0775)
High education (mom)	0.202**	0.355***	0.152*	-0.0514	-0.129
	(0.0862)	(0.0790)	(0.0918)	(0.0884)	(0.0823)
White	0.390***	0.388***	0.0931	0.401***	0.498***
	(0.119)	(0.109)	(0.126)	(0.122)	(0.113)
Black	0.116	0.0541	-0.148	0.0986	0.198*
	(0.110)	(0.101)	(0.118)	(0.113)	(0.105)
Constant	4.438***	4.795***	4.224***	4.683***	3.676***
	(0.127)	(0.116)	(0.135)	(0.130)	(0.121)
Observations	3331	3332	3332	3330	3328

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the number of days per week a mother does a given activity with her child when the child is 5 years old. High education means the mother at least attended college.

Table 17: Child investments, period 4

	(1) Play outside	(2) Watch TV	(3) Video games	(4) Read stories	(5) Play inside	(6) Current events
Married	-0.0663	-0.374***	-0.135	-0.167	0.0149	-0.121
	(0.0919)	(0.119)	(0.0947)	(0.122)	(0.0843)	(0.131)
Cohabiting	0.0118	0.119	-0.0386	-0.305*	-0.249**	-0.147
	(0.134)	(0.175)	(0.139)	(0.179)	(0.123)	(0.192)
Employment status	-0.255***	-0.178*	-0.0832	-0.0643	-0.147**	-0.0943
	(0.0779)	(0.101)	(0.0803)	(0.104)	(0.0715)	(0.111)
Bad relationship	0.0526	0.0428	0.0837	-0.0681	0.101	-0.0737
	(0.0779)	(0.101)	(0.0803)	(0.103)	(0.0715)	(0.111)
High education (mom)	0.0349	-0.587***	-0.470***	0.207*	-0.150**	0.225*
	(0.0808)	(0.105)	(0.0833)	(0.107)	(0.0742)	(0.116)
White	0.435***	-0.0999	0.0618	0.272*	0.180*	0.114
	(0.114)	(0.149)	(0.118)	(0.152)	(0.105)	(0.164)
Black	-0.0100	0.376***	0.400***	0.392***	0.179*	0.775***
	(0.107)	(0.139)	(0.111)	(0.142)	(0.0984)	(0.153)
Constant	1.835***	4.349***	1.393***	3.418***	1.554***	3.147***
	(0.124)	(0.162)	(0.128)	(0.165)	(0.114)	(0.178)
Observations	2688	2688	2688	2688	2688	2688

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the number of days per week a mother does a given activity with her child when the child is 9 years old. High education means the most at least attended college.

Appendix B: Reduced form evidence- health outcomes

Table 18: Child's health

	Dependent variable = child health (1-5 scale)				
	(1)	(2)	(3)	(4)	(5)
	Period 1	Period 2	Period 3	Period 4	Period 5
Married	0.0826**	0.0640**	0.115***	0.0969**	0.120***
	(0.0335)	(0.0319)	(0.0314)	(0.0385)	(0.0440)
Cohabiting	0.0484	0.0953***	-0.0127	0.0157	-0.111
	(0.0312)	(0.0339)	(0.0398)	(0.0563)	(0.0783)
Employment status	0.0182	0.0533**	0.0694***	0.0363	0.106***
	(0.0259)	(0.0258)	(0.0264)	(0.0326)	(0.0385)
Bad relationship	-0.0593**	-0.0748***	-0.0493*	-0.0657**	-0.0960***
	(0.0260)	(0.0259)	(0.0264)	(0.0326)	(0.0364)
High education (mom)	0.131***	0.125***	0.0879***	0.0676**	0.0882**
	(0.0285)	(0.0279)	(0.0280)	(0.0339)	(0.0372)
White	0.171***	0.148***	0.0833**	0.155***	0.169***
	(0.0374)	(0.0374)	(0.0385)	(0.0479)	(0.0528)
Black	0.155***	0.133***	0.0501	0.0121	0.0763
	(0.0350)	(0.0353)	(0.0358)	(0.0449)	(0.0493)
Constant	4.294***	4.283***	4.361***	4.286***	4.153***
	(0.0396)	(0.0406)	(0.0412)	(0.0521)	(0.0589)
Observations	4028	3666	3332	2688	2301

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. High education means the mother at least attended college.

Appendix C: Reduced form evidence- academic outcomes

Table 19: Academic outcomes, period 2

	Dependent variable = test scores (percentiles)	
	(1)	(2)
	TVIP	PPVT
Married	-4.636	3.105***
	(3.347)	(0.883)
Cohabiting	-7.283*	-0.981
	(3.743)	(0.887)
Employment status	-2.005	2.268***
	(2.611)	(0.690)
Bad relationship	-3.249	-0.498
	(2.434)	(0.688)
High education (mom)	-0.737	6.870***
	(3.689)	(0.761)
White	1.676	7.574***
	(2.528)	(1.094)
Black	-14.62	-0.148
	(13.16)	(1.014)
Constant	103.4***	79.54***
	(3.681)	(1.142)
Observations	118	2156

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table shows determinants of test scores at age 3. High education means the mom at least attended college.

Table 20: Academic outcomes, period 3

	Dependent variable = test scores (percentiles)	
	(1)	(2)
	PPVT	WJ22
Married	3.546***	6.098***
	(0.798)	(1.517)
Cohabiting	-0.104	2.117
	(0.966)	(1.829)
Employment status	2.868***	6.731***
	(0.653)	(1.239)
Bad relationship	-0.648	-2.461**
	(0.651)	(1.236)
High education (mom)	8.198***	12.50***
	(0.706)	(1.338)
White	7.887***	0.948
	(1.006)	(1.911)
Black	1.057	7.555***
	(0.915)	(1.739)
Constant	85.48***	35.93***
	(1.034)	(1.967)
Observations	2037	2050

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table shows the determinants of test scores at age 5. High education means the mom at least attended college.

Table 21: Academic outcomes, period 4

Dependent variable = test scores (percentiles)			
	(1)	(2)	(3)
	PPVT	WJ9	WJ10
Married	4.021***	6.276***	6.630***
	(0.663)	(1.162)	(1.327)
Cohabiting	-1.541	-0.275	-2.326
	(0.952)	(1.666)	(1.897)
Employment status	1.576***	1.379	3.336***
	(0.561)	(0.981)	(1.120)
Bad relationship	-0.768	-1.872*	-2.165*
	(0.561)	(0.981)	(1.120)
High education (mom)	8.834***	11.59***	12.62***
	(0.581)	(1.017)	(1.161)
White	5.326***	6.661***	5.883***
	(0.824)	(1.443)	(1.645)
Black	-2.576***	0.610	-5.716***
	(0.767)	(1.344)	(1.535)
Constant	88.11***	28.58***	41.91***
	(0.894)	(1.565)	(1.789)
Observations	2516	2506	2512

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table shows determinants of test scores at age 9. High education means the mom at least attended college.

Table 22: Child investments, period 5

	Dependent variable = grades (1-4 scale)			
	(1)	(2)	(3)	(4)
	English	Math	History	Science
Married	0.233***	0.207***	0.215***	0.154***
	(0.0470)	(0.0519)	(0.0512)	(0.0507)
Cohabiting	0.197**	0.0935	0.170*	0.228**
	(0.0845)	(0.0937)	(0.0920)	(0.0909)
Employment status	0.0208	0.0564	0.0867*	0.0697
	(0.0419)	(0.0464)	(0.0462)	(0.0452)
Bad relationship	-0.0559	-0.133***	-0.0662	-0.128***
	(0.0390)	(0.0431)	(0.0426)	(0.0420)
High education (mom)	0.175***	0.154***	0.253***	0.183***
	(0.0398)	(0.0440)	(0.0436)	(0.0429)
White	0.110*	0.0718	0.0439	0.148**
	(0.0563)	(0.0624)	(0.0627)	(0.0607)
Black	-0.0862*	-0.0624	-0.196***	-0.117**
	(0.0523)	(0.0580)	(0.0581)	(0.0563)
Constant	2.824***	2.691***	2.837***	2.778***
	(0.0629)	(0.0700)	(0.0695)	(0.0677)
Observations	2120	2124	1975	2086

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table shows determinants of grades at age 15. High education means the mom at least attended college.

Appendix D: Two period model

In this section, we solve for the closed form solution to the two period model to show how we identify the utility and child quality transition parameters. We ignore demographics for simplicity. At the start of period 1, we have relationship status s_0 and child quality q_0 which are exogenously given. We ignore the role of investment given that it is not included in our current results. For simplicity, we ignore labor market decisions, assuming that the choice set $S = \{1, 2, 3\}$ indicating married, cohabiting, and single.

First consider the period 2 decision. Assume a person chooses relationship status s_1 in period 1. The value function can be written as follows:

$$V_2\left(s_1, q_1, \{\varepsilon_{s2}\}\right) = \max_{s_2} k_{s_1,s_2} + q_1 + \varepsilon_{s_2}$$

The choice probabilities take a logit form because we assumed that the payoff shocks follow the extreme value distribution.

$$Pr_2\left(s_2|q_1, s_1\right) = \frac{\exp\left(k_{s_1,s_2} + q_1\right)}{\sum_{s\in S}\exp\left(k_{s_1,s} + q_1\right)}$$

We can then write the log odds ratio, considering the ratio of the choice probability for choice s^A versus choice s^B:

$$\begin{aligned}
\log\left(\frac{Pr_2\left(s^A|q_1, s_1\right)}{Pr_2\left(s^B|q_1, s_1\right)}\right) &= \log\left(\frac{\exp\left(k_{s_1,s^A} + q_1\right)}{\exp\left(k_{s_1,s^B} + q_1\right)}\right) \\
&= \log\left(\exp\left(k_{s_1,s^A} - k_{s_1,s^B}\right)\right) \\
&= k_{s_1,s^A} - k_{s_1,s^B} \qquad (16)
\end{aligned}$$

We can use equation (16) to identify the difference between the net utilities of each choice pair. To do this, we first need to normalize utility. We do this by setting $k_{11} = 0$. This is necessary because the k terms are only identified to scale. Since there are 3 choices and therefore 9 possible pairs, there will be 8 utility terms we need to identify. To see how to do this, consider the log odds ratio for choosing $s_2 = 1$ versus $s_2 = 2$, conditional on original relationship status $s_1 = 1$.

$$\log\left(\frac{Pr_2(s_2 = 1|q_1, s_1 = 1)}{Pr_2\left(s_2 = 2|q_1, s_1 = 1\right)}\right) = k_{11} - k_{12} \qquad (17)$$

$$\log\left(\frac{Pr_2(s_2 = 1|q_1, s_1 = 1)}{Pr_2\left(s_2 = 3|q_1, s_1 = 1\right)}\right) = k_{11} - k_{13} \qquad (18)$$

We cannot consider the third log odds ratio conditioning on $s_1 = 1$ because it is redundant. Because we set $k_{11} = 0$, we can use equations (17) and (18) identify k_{12} and k_{13}. We can also write down the log odds ratios conditioning on $s_1 = 2$ and $s_1 = 3$.

$$\log\left(\frac{Pr_2(s_2 = 1|q_1, s_1 = 2)}{Pr_2(s_2 = 2|q_1, s_1 = 2)}\right) = k_{21} - k_{22} \tag{19}$$

$$\log\left(\frac{Pr_2(s_2 = 1|q_1, s_1 = 2)}{Pr_2(s_2 = 3|q_1, s_1 = 2)}\right) = k_{21} - k_{23} \tag{20}$$

$$\log\left(\frac{Pr_2(s_2 = 1|q_1, s_1 = 3)}{Pr_2(s_2 = 2|q_1, s_1 = 3)}\right) = k_{31} - k_{32} \tag{21}$$

$$\log\left(\frac{Pr_2(s_2 = 1|q_1, s_1 = 3)}{Pr_2(s_2 = 3|q_1, s_1 = 3)}\right) = k_{31} - k_{33} \tag{22}$$

We are left with 6 utility terms to identify, and only 4 equations. Therefore to identify the remaining parameters we need to look at the period 1 decision.
Using the properties of the extreme value distribution to solve for the expected continuation values, we can write the period 1 value function as follows:[13]

$$\begin{aligned}
V_1(s_0, q_0, \{\varepsilon_{s1}\}) &= \max_{s_1 \in S} k_{s_0,s_1} + q_0 + \varepsilon_{s1} + \beta \log\left(\sum_{s_2 \in S} \exp\left(k_{s_1,s_2} + q_0 + \delta_{1s_1} + \delta_{2s_1} \log(1 + \exp(q_0))\right)\right) \\
&= \max_{s_1 \in S} \Bigg\{ k_{s_0,s_1} + q_0 + \varepsilon_{s1} + \beta\left(q_0 + \delta_{1s_1} + \delta_{2s_1} \log(1 + \exp(q_0))\right) \\
&+ \beta \log\left(\sum_{s_2 \in S} \exp\left(k_{s_1,s_2}\right)\right)\Bigg\}
\end{aligned}$$

We can solve for the choice probabilities over relationship status in period 1:

$$Pr_1(s_1|q_0, s_0) = \frac{\exp\left(k_{s_0,s_1} + q_0 + \beta\left(q_0 + \delta_{1s_1} + \delta_{2s_1}\log(1 + \exp(q_0))\right) + \beta \log\left(\sum_{s_2} \exp(k_{s_1,s_2})\right)\right)}{\sum_s \exp\left(k_{s_0,s} + q_0 + \beta\left(q_0 + \delta_{1s} + \delta_{2s}\log(1 + \exp(q_0))\right) + \beta \log\left(\sum_{s_2} \exp(k_{s,s_2})\right)\right)}.$$

We can now calculate the log odds ratio next. After doing some algebra, we get

$$\begin{aligned}
\log \frac{Pr_1(s^A|q_0, s_0)}{Pr_1(s^B|q_0, s_0)} &= \left(k_{s_0,s^A} - k_{s_0,s^B}\right) + \beta\left(\delta_{1s^A} - \delta_{1s^B}\right) + \beta \log(1 + \exp(q_0))\left(\delta_{2s^A} - \delta_{2s^B}\right) \\
&+ \beta\left(\log\left(\sum_{s_2} \exp\left(k_{s^A,s_2}\right)\right) - \log\left(\sum_{s_2} \exp\left(k_{s^B,s_2}\right)\right)\right)
\end{aligned}$$

[13] We assume that each shock is equal to an extreme value type I shock minus Euler's constant, which causes Euler's constant to drop out of the expected continuation value.

Now consider the log ratio of the odds ratios in periods 1 and 2. To do this, consider a person starting either period 1 or period 2 in a given state (call this $\tilde{s}$). Then we consider the odds ratio of her picking state s^A and s^B. We do this for periods 1 and 2, and take the ratio of the ratios for each period. Explicitly, we define $Q(\tilde{s}, s^A, s^B)$ as follows

$$Q\left(\tilde{s}, s^A, s^B\right) \quad = \log\left(\frac{P_2\left(s^A|\tilde{s},q_1\right)/P_2\left(s^B|\tilde{s},q_1\right)}{P_1(s^A|\tilde{s},q_0)/P_1(s^B|\tilde{s},q_0)}\right)$$

Lets consider a person starting in state 1 ($\tilde{s} = 1$), and considering the odds ratio for choices 1 and 2 ($s^A = 1$ and $s^B = 2$):

$$\begin{aligned} Q(1,1,2) &= \beta\left(\delta_{12} - \delta_{11}\right) + \beta \log\left(1 + \exp\left(q_0\right)\right)\left(\delta_{22} - \delta_{21}\right) \\ &- \beta \log\left(e^{k11} + e^{k12} + e^{k13}\right) + \beta \log\left(e^{k21} + e^{k22} + e^{k23}\right) \end{aligned}$$

Assuming we know q_0, we can now derive more equations that will help identify the k and δ terms. Just as above, we need to normalize the δ terms, with the normalization $\delta_{11} = \delta_{21} = 0$. First, consider a regression with the ratio of the log odds ratio on the left hand side, and the right hand side variable is q_0. The constant term is this regression is a function of the k and δ_1 terms, and the slope is a function of δ_2 terms. First let's consider the constant term. From the period 2 decision, we can write k_{22} and k_{23} as a function of k_{21}. In addition, the period 2 decision already identified k_{11}, k_{12}, and k_{13}. Then we are left with δ_{12} and k_{12}, which are not separately identified without further information. In fact, δ_{12} needs to be identified off of changes in q, as a function of marital status. Therefore we need separate information on q transitions to identify δ_{12}. Once we have that, we can then identify k_{12}, which gives us k_{22} and k_{23} also. We can repeat this exercise with the remaining log odds ratios to identify the rest of the parameters.

We normalized $\delta_{21} = 0$. Then the slope in this regression identifies δ_{22}. We can repeat this with another log odds ratio to identify δ_{23}.