Analyzing matching patterns in marriage: Theory and application to Italian data*

Pierre-André Chiappori† Edoardo Ciscato‡ Carla Guerriero§

December 18, 2023

Abstract

Social scientists have long been interested in marital homogamy and its relationship with inequality. However, measuring homogamy is not straightforward, particularly when one is interested in assessing marital sorting based on multiple traits. In this paper, we argue that Separate Extreme Value (SEV) models not only generate a matching function with several desirable theoretical properties, but they are also suited for the study of multidimensional sorting. Specifically, we show (i) how a small number of factors can be identified that capture most of the explained variance in matching patterns, and (ii) how these factors relate to various "outcomes" of the post-matching relationship, such as children's human capital and well-being. We then use rich small-scale survey data to examine sorting among parents of children attending schools in Naples. Our findings show that homogamy is pervasive; not only do men and women sort by age, education, height, and physical characteristics, but they also look for partners that share similar health-related behavioral traits and risk attitude. We also show that marital patterns are well explained by a low number of dimensions, the most important being age and human capital. Moreover, children of parents with a high human capital endowment perform better at school, although they report lower levels of subjective well-being and of perceived quality of relationship with their mothers.

Keywords: homogamy, matching, intergenerational inequality.

*Funding from Programma Star UNINA Linea 1 Junior Principal Investigator Grant 2016 (Compagnia di San Paolo and Istituto Banco di Napoli) is gratefully acknowledged. This paper would have never been possible without the attention and enthusiasm of the children, their parents, and the precious help of the school teachers. A special thank you to the DISES-UNINA administrative staff. A development version of the R package used to produce the quantitative findings of this paper can be found on the following GitHub website: https://github.com/edoardociscato/affinitymatrix.

†Columbia University, Department of Economics

‡KU Leuven, Department of Economics

§University of Naples Federico II, Department of Economics and Statistics

1 Introduction

Since the pioneering work of Becker (1973), a large number of studies have analyzed matching patterns in the marriage market. From a social sciences perspective, one of the major motivations to study partner choice is the relationship between marital patterns and inequality. Assortative matching has a direct impact on inequality *within generations*; if individuals have a large propensity to marry their own likes, initial inequalities in individual endowments tend to be amplified at the household level (Fernández and Rogerson, 2001; Greenwood et al., 2003, 2014). More importantly, recent studies have emphasized the potential impact of assortative matching on social reproduction and *intergenerational* inequality.

While numerous studies in economics, sociology, and demography have investigated marital patterns,[1] they usually concentrate on one specific trait, such as income or human capital.[2] However, the real-life process of marital matching is obviously much more complex, and involves a host of other characteristics: age, race, religion, but also tastes and preferences, cultural background, physical attractiveness, etc. From a methodological viewpoint, whether the empirical strategy adopted can account for this multi-dimensionality and, more importantly, whether it can disentangle the respective impact of multiple traits, especially when the latter appear to be correlated, is an important question.[3]

The present paper has two goals. One is to contribute to the literature on matching and human capital formation by applying and extending a line of research initiated by Dupuy and Galichon (2014) on multidimensional matching. The approach is based on the so-called Separable Extreme Value (SEV) model, which has become dominant in the empirical analysis of matching models. This framework uses a frictionless matching framework with Transferable Utility, in which unobserved heterogeneity is captured through an additive, separable random term that is typically assumed to follow a type 1 extreme value distribution.

Our focus is on the empirical implementation of this framework in a multidimensional context. Specifically, we analyze a rich Italian dataset, a survey of parents and their children aged 6 to 19 attending schools in the Campania region around Naples. This unique dataset of 276 families comprises information on parents, children, and families as a whole. The data collected on parents include sociodemographic variables (age and education), anthropometric characteristics (height and weight), health-related behavior (e.g., healthy eating, smoking and sports activity) and household-level characteristics (e.g., number of children and the time spent by the mother at home). The survey also collects parental psychometric information on risk behavior, with a focus on health and recreational risks. These data provide a perfect sample for analyzing matching patterns when the number of traits is "large" (15 in our case).

We first analyze the interaction of individual characteristics in the matching process using the method introduced by Dupuy and Galichon (2014). Next, we describe how one can estimate more restricted models in which these numerous traits only matter through a small number of unknown "factors". This issue should be

[1] A non-exhaustive list includes Schoen (1981); Qian and Preston (1993); Blackwell and Lichter (2004); Schwartz and Mare (2005); Bouchet-Valat (2014); Schwartz and Han (2014); Gonalons-Pons and Schwartz (2017).

[2] An obvious exception is Dupuy and Galichon (2014), which is discussed later on.

[3] For a general presentation of multidimensional matching, see Chiappori et al. (2010).

considered in the perspective of the applied theory literature on the topic, most of which emphasizes the role of human capital in the matching process. The story, here, is that individuals with a high level of human capital tend to match assortatively; these households tend to invest heavily in their children's human capital, and their investments tend to be particularly productive (Chiappori et al., 2017). Given the complementarities involved in the human capital production function, such trends amplify initial inequalities for the next generation, generating what can be termed an "inequality spiral".[4] An important question, therefore, is whether the emphasis on human capital in general, and on investment into children in particular, can be justified by a direct investigation of existing data (as opposed to simply reflecting economists' unwarranted prejudices about a complex and multidimensional process). In other words, a key issue related to our decomposition is whether a factor reflecting "human capital" naturally emerges as a driving force underlying the matching process when one lets the data "speak by themselves"; and, if so, whether this emergence is specific to one of the partners or common to both genders, as well as whether it is related to children's outcomes.

Indeed, unlike most existing datasets, we also record information on children. First and foremost, we observe class grades in mathematics and Italian for all pupils, as well as standardized test scores (INVALSI) in the same subjects for a subset of them. Other outcomes include anthropometric characteristics, measures of risk preferences, attitude and taste for healthy food, and ability to defer gratification. Moreover, children are surveyed about the quality of the relationship with their parents and the amount of time spent with them, physical and mental health, attitude towards healthy lifestyle and eating habits, the amount of time spent in front of screens, and the children's degree of altruism. This information is particularly interesting, since it relates to parent's investments into their children and to the *outcomes* of these investments.

The second goal of the present paper is precisely to show how these outcomes, although they are not used in the estimation of the matching factors themselves, can be statistically related to these factors. Technically, the main matching factors define a small-dimensional subspace (three dimension in our general case); one can then project the various outcomes onto this subspace, and standard techniques allow us to estimate the statistical significance of the relationship.

In other words, while our data does not allow for an estimation of the production function of children's human capital in the technical sense, it enables us to study the statistical link between parents' matching patterns and children outcomes without relying on any a priori hypothesis about the nature of this relationship. We investigate whether the main factors that drive matching patterns are significantly related to some children outcomes (and if so to which outcomes). In that sense, while our analysis remains essentially descriptive, it provides new and interesting insights into the relationship between parents' matching behavior and children's endowments and welfare.

Finally, our framework allows us to perform some counterfactual experiments.

[4] Becker and Tomes (1979) emphasizes the role of parental investment in the perpetuation of economic inequality. The subsequent literature is summarized in Becker et al. (2018) and Durlauf et al. (2022). A smaller number of papers link mating, parental investment, and economic inequality, including Gayle et al. (2013), Del Boca et al. (2014), Chiappori et al. (2017), Chiappori et al. (2018), and Chiappori et al. (2020).

In particular, we can decompose the observed relationship between matching patterns and children outcomes into the respective contributions of the different factors previously isolated, and simulate a similar model under different matching patterns. We particularly concentrate on two counterfactual scenarios, namely pure random matching and pure assortative matching on human capital. While our results should be taken with some precaution - in particular because the data does not enable us to establish a causality relationship in a rigorous manner - we believe that this approach may help substantiate the vision of an inequality spiral across generations.

Our conclusions can be summarized as follows. First, marital patterns are characterized by a surprisingly large degree of homogamy. In practice, observed preferences for homogamy, as identified from the matching model, are related to age, education and anthropometric measures, but also to *all* behavioral traits under consideration, including those reflecting health and risk attitudes. Formally, we identify the *affinity matrix* of the matching game, which summarizes the respective contributions of individual characteristics to the total surplus generated by any given match (see below for a precise definition). Preferences for homogamy are then reflected by the diagonal of the matrix. In our case, *each* of the 15 diagonal terms is estimated to be positive, a remarkable pattern indeed (whose probability under the null of random matching is significantly less than .1%). An obvious caveat is that, since we observe couples a few years after marriage, some individual traits may have converged. Yet, the existing literature suggests that the traits we consider do not significantly change after marriage (see for instance Dohmen et al. (2012) on risk and trust attitudes).

A second conclusion is that, despite its complexity, the sorting patterns can essentially be reduced to a low-dimensional process. Indeed, a model based on a small number of factors - three for the most general version of our model - constitutes a very good approximation of the marriage market. Quite remarkably, the corresponding factors, which are independently identified for men and women, appear to be surprisingly similar across genders - a finding that was by no means guaranteed ex ante, although it is certainly in line with most theoretical settings used in the literature.

The first sorting factor reflects the underlying segmentation of the dataset in different age cohorts; essentially, people tend to marry spouses within the same age range, a conclusion that is neither too surprising nor particularly interesting. More importantly, the second main dimension can indeed be interpreted as capturing sorting based on human capital. Educated women tend to primarily marry educated men; moreover, sorting also brings together individuals that are health-conscious (e.g., that smoke less and have a preference for healthy food), health being traditionally considered as an important component of human capital. It should be stressed that the importance of human capital is not imposed into the model by a narrow selection of matching traits. In our framework, education is only one of the 15 traits we consider, and attitudes towards health relate to a handful of traits at most. Yet, these aspects are found to play a dominant role in the definition of the second main (and economically most meaningful) factor. Finally, the third factor gathers traits related to lifestyles in general, including aspects such as tastes for a more hedonistic lifestyle, and may be related to parenting styles.

Next, we analyze the relationships between these assortative matching patterns and child outcomes. Positive assortative mating may play an important role in children's direct socialization; as argued in a large literature, parents who want to transmit their characteristics to their children may be more likely to choose a partner with a similar bundle of traits (Bisin and Verdier, 2000; Bisin et al., 2004). Using our projection approach, we find, in particular, that parents with a high level of human capital have children that perform significantly better at school (and have healthier eating habits), as expected from the theory. Interestingly, however, these same children report lower levels of happiness and a worse relationship with their mothers, suggesting that parental investments come at a cost in terms of children's welfare.

To further investigate the relationsip between the factors governing parental matching and their impact on children, we run a series of regressions in which the various child outcomes are explained either by the entire set of parental characteristics, or by the matching factors and their interactions. Our findings are as follows:

- Most child outcomes appear to be largely orthogonal to parents' characteristics. Yet, outcomes that relate to human capital (such as grades) or interactions between parents and children (such as children's happiness) are significantly correlated with the second factor.

- The three factors, which have been exclusively computed from parents' matching patterns, explain basically as much (or as little) of the variance in these children outcomes as the entire set of parental characteristics. This is compatible with the hypothesis of a strong link between the matching process and the technology of human capital production for children

- Regarding children's grades, the only interaction that has a significant impact is between the parents' second factors; and the corresponding coefficient is positive for both the math and the Italian grade (although it is only significant for the Italian grade). In other words, raw data support the complementarity assumption that most theoretical models introduce.

- Interestingly, this complementarity between parental characteristics is only visible on the factor decomposition. In a similar regression using age and education instead of the factors, the interaction coefficient on education is always negative (significantly so in some versions). This strongly suggests that the second factor, while correlated with education, also captures other aspects of human capital (such as health), and that it is this particular combination of characteristics that is relevant for both matching patterns and children outcomes. These findings also provide a further justification for an approach based on a factor decomposition of matching patterns, while raising some concerns about the use of education as the only matching trait.

- However, promoting children's human capital has a cost. When analyzing various measures of children's welfare, we find that the mother's and father's second factor, as well as their interaction, are all negatively correlated with different measures of subjective well-being and with self-reported measures of child-parent relationship quality.

Finally, we perform some counterfactual experiments; specifically, we use the coefficients estimated in our outcome regressions to decompose the impact of matching patterns on the distribution of grades into several channels (here, mathcing factors);

we can then estimate the impact of shutting down some of the channels. In practice, we consider two extreme cases - purely random matching on the one hand, and matching exclusively based on the second, human capital factor on the other hand; both simulations are compared to raw data as well as our estimations. Our estimates fit actual data almost perfectly. Random matching would significantly change the grade distribution among children. Specifically, the distribution of Italian grades in the first counterfactual experiment is significantly shifted to the left. The shift for math grades is less brutal; yet, in both cases the number of children with a grade above (resp. below) the mode of the distribution is lower (higher) under random matching. On the contrary, matching exclusively on the human capital factor would shift the grades to the right, although again the shift is stronger for Italian grades. This suggests that while people do match based on human capital, and while this assortativeness is correlated with better student performances, the real-life process is quite complex. Individuals match on many traits, several (and perhaps most) being unobservable; while human capital stands out as a prominent factor, it is by no means the only one, and some other traits on which spouses match assortatively may actually be substitutes in the children's production function.

The next section presents the main theoretical notions underpinning our work. Section 3 provides an accurate description of the database, while Section 4 details the econometric approach developed for this analysis. Results are presented and discussed in Section 5.

2 Analyzing Marital Patterns: Theory

2.1 The Matching Model

In this section, we review the static, one-to-one, bipartite matching framework with transferable utility of Dupuy and Galichon (2014). We consider two populations, men and women, with equal mass, and each defined by a set of characteristics X and Y, whose distributions are assumed to have densities f and g, respectively. Each woman is described by a vector of characteristics x of length m, each man by a vector y of length n. The matching is described by a *measure* μ on the product space $X \times Y$, and $\mu(x, y)$ denotes the probability that a woman of type x is matched with a man of type y. However, μ is constrained by the fact that its marginals must coincide with $f(x)$ and $g(y)$; in other words, the matching is *feasible* if and only if

$$\mu \in M(f,g) := \left\{ \mu : \begin{array}{ll} \mu(x,y) \geq 0 & \forall x, y, \\ \int_Y \mu(x,y)dy = f(x) & \forall x, \\ \int_X \mu(x,y)dx = g(y) & \forall y \end{array} \right\}. \tag{1}$$

The model aims to explain sorting patterns in the matched population, so the marginal constraints in (1) hold with equality. The model can be extended to include singlehood in the agents' choice sets.[5] However, as explained at the end of this section, the participation margin can be studied separately from the sorting margin under the distributional assumptions made by Dupuy and Galichon (2014).

A match between a woman of type x and a man of type y generates a match surplus $\Phi(x, y)$ on top of additional partner-specific random gains. Each woman of

[5]This is discussed in Appendix D in Dupuy and Galichon (2014).

type x only considers a subset of all men in the population (her *acquaintances*) when choosing a partner. The infinite and countable set of acquaintances $\{(y_k, \varepsilon_k), k \in \mathbb{N}\}$ results from a Poisson process on $Y \times \mathbb{R}$ of intensity $dy \times e^{-\varepsilon} d\varepsilon$. Each potential partner k within this set is characterized by a type y_k and a taste shock ε_k. The woman's utility of choosing k as a husband is $U(x, y_k) + \sigma\varepsilon_k$, where $U(x, y_k)$ is a share of $\Phi(x, y)$ and ε_k is the random component of the wife's gains from marriage.[6] On the other side of the market, men's preferences over women are defined analogously, and a man's utility from matching with a woman l within his set of acquaintances is $V(x_l, y) + \sigma\eta_l$. For any pair (x, y), the shares are constrained by $U(x, y) + V(x, y) \leq \Phi(x, y)$.

A matching μ is *stable* when there is no pair of individuals who would both prefer being matched together rather than with their current partners. When the matching is stable, all agents maximize utility by choosing the best mate for given surplus shares $U(x, y)$ and $V(x, y)$. The latter are equilibrium objects that not only reflect the (exogenous) quality of a match of type (x, y), but also result from the (endogenous) relative bargaining power within a couple of type (x, y). When demand for a type x increases, then $U(x, y)$ will increase relatively to $V(x, y)$.

Demand and supply functions can be derived from the agents' mate choice problems. A woman of type x chooses her husband among her acquaintances in order to maximize her utility from mating

$$\max_k \left\{U(x, y_k) + \sigma\varepsilon_k\right\}. \tag{2}$$

The distributional assumption on ε implies that a woman of type x will choose a man of type y with probability

$$\mu_f(y|x) = \frac{\exp\left(U(x, y)/\sigma\right)}{\int_Y \exp\left(U(x, y)/\sigma\right) dy}. \tag{3}$$

Under the same distributional assumption for η, the analogous of problem (2) for men implies that a man of type y will choose a woman of type x with probability

$$\mu_m(x|y) = \frac{\exp\left(V(x, y)/\sigma\right)}{\int_X \exp\left(V(x, y)/\sigma\right) dx}. \tag{4}$$

The equilibrium matching is both stable and feasible. Combining supply and demand conditions (3) and (4), one can obtain the matching function

$$\mu(x, y) = \exp\left(\frac{\Phi(x, y) - a(x) - b(y)}{2\sigma}\right) \tag{5}$$

and can also characterize the woman's share as

$$U(x, y) = \frac{\Phi(x, y) + a(x) - b(y)}{2}. \tag{6}$$

where $a(x)$ and $b(y)$ are defined as

$$a(x) := \sigma \log \int_Y f(x)^{-1} \exp\left(U(x, y)/\sigma\right) dy \tag{7}$$

$$b(y) := \sigma \log \int_X g(y)^{-1} \exp\left(V(x, y)/\sigma\right) dx. \tag{8}$$

[6] In Choo and Siow (2006), spouses have preferences over *types* rather than over partners. A woman's utility from a match with a man k is given by $U(x, y_k) + \varepsilon(y_k)$, and therefore she is perfectly indifferent between two men of the same type.

Dupuy and Galichon (2014) recasts the matching problem as finding two functions $a(x)$ and $b(y)$ that satisfy the marginal constraints under the matching function (5). In their Theorem 1, they prove that there exist two unique functions $a(x)$ and $b(y)$ (up to a constant) so that the constraints hold. These two functions pin down both the unique equilibrium matching (5) and the equilibrium sharing rule (6).

An important property of this choice model is the Independence from Irrelevant Alternatives (IIA), i.e., the fact that the *relative* probabilities of any set of possible choices do not depend on the presence of other (irrelevant) alternatives. This implies that estimating the probability of choosing a partner of type y instead of type y', *conditional on marriage*, will give the same conclusions as analyzing *unconditional* choices (i.e., also taking singlehood as a possible choice).

2.2 Parameterization and Estimation

The surplus function is conveniently parameterized as follows,

$$\Phi(x, y) = x'Ay = \sum_{k,l} A_{kl} x_k y_l, \tag{9}$$

so that every element of the $m \times n$ *affinity matrix* A tells us whether two traits x_k and y_l are complements ($A_{kl} > 0$) or substitutes ($A_{kl} < 0$). In our context, $m = n$ since the various traits are observed for both spouses. In particular, diagonal terms A_{kk}, with $k = 1, ..., n$, are directly informative about the super- ($A_{kk} > 0$) or sub- ($A_{kk} < 0$) modularity of the surplus with respect to characteristic k.

The affinity matrix can be estimated using a maximum likelihood estimator. With a sample of N couples, the likelihood function corresponds to

$$\mathcal{L}(A) = \frac{1}{N} \sum_{i=1}^{N} \mu^A(x_i, y_i), \tag{10}$$

which means that the equilibrium matching μ^A, whose expression is given in equation (5), must be computed for every choice of A. When the number of matching variables is high relatively to the sample size N, Dupuy et al. (2019) suggest using a penalized likelihood method in order to allow for A to be sparse and increase the precision of the parameter estimates. A way of introducing sparsity is to impose a restriction to the rank of A. In practice, this can be achieved by generalizing (10) with

$$\mathcal{L}(A) = \frac{1}{N} \sum_{i=1}^{N} \mu^A(x_i, y_i) + \tau \left\| A \right\|_*, \tag{11}$$

where $\|A\|_*$ represents the nuclear norm of A, defined as the sum of its singular values, whereas τ is the Lagrangian multiplier associated with the rank constraint.

In the second part of our empirical analysis, we estimate a model with 15 matching variables, and thus with a 15×15 affinity matrix, with a sample of only 201 couples. Hence, penalized MLE helps us considerably reduce the dimensionality of the econometric problem. In practice, we set τ through a four-fold Cross Validation procedure proposed by Dupuy et al. (2019), which restricts the rank of A to be eight at most.

2.3 Dimensionality and Matching Indices

The biquadratic specification offers an important advantage; one can rewrite Φ as the linear combination of independent factors, each capturing a different dimension of assortativeness. This is insightful for multiple reasons. Firstly, it allows one to infer the number of dimensions of assortativeness. For instance, we can test the hypothesis that attractiveness is well summarized by a single index (or a small number of indices) subsuming numerous observable traits. Secondly, when multiple dimensions of assortativeness matter, it quantifies their relative importance. Lastly, it allows one to describe the role played by the observables x and y in each dimension of assortativeness.

In practice, as shown by Dupuy and Galichon (2014), the affinity matrix can be decomposed through a Singular Value Decomposition (SVD) as[7]

$$A = U'\Lambda V,$$

where Λ is a diagonal matrix whose positive nonincreasing elements $(\lambda_1, \dots, \lambda_K)$, with $K = \min\{m, n\}$, capture the relative importance of each sorting dimension, while the columns of U and V are loading vectors that describe the nature of each dimension. In other words, we can define the *factors* (or *indices*) as $\tilde{x} = Ux$ and $\tilde{y} = Vy$, and rewrite the surplus as

$$x'Ay = \tilde{x}'\Lambda\tilde{y} = \sum_{k=1}^{K} \lambda_k \tilde{x}_k \tilde{y}_k, \tag{12}$$

where each k term $\lambda_k \tilde{x}_k \tilde{y}_k$ represents the surplus contribution of an independent dimension of assortativeness.

In our empirical analysis, we can perform SVD on the estimated affinity matrix $\hat{A}$ to obtain estimates of U, V, and Λ. In this way, we can discuss the relative importance and nature of the different dimensions of assortativeness. Confidence intervals for U, V, and Λ can be obtained with bootstrap techniques.

However, what is unknown is how many relevant dimensions of assortativeness do we observe as well as how many elements of Λ are positive and significant.[8] Dupuy and Galichon (2014) outline a method to answer these questions and develop a test of joint significance of the estimated Λ. In summary, the method consists of testing the rank of the estimated affinity matrix $\hat{A}$. The null hypothesis is a restriction on the rank of Z, i.e., $rank(A) = k$; when the number of positive diagonal elements of Λ is higher than k, then the hypothesis is rejected, and we can conclude that the number of relevant dimensions of assortativeness is higher than k.

3 Data

The survey data used in this paper contain information about the preferences, beliefs, and actions of both parents and their children. Compared with previous studies on parents, this rich dataset fleshes out the separate role of each spouse and their

[7] Importantly, we work with demeaned and rescaled data. More precisely, each observable characteristic x_k is demeaned and all characteristics are rescaled so that the diagonal elements of the sample covariance $\hat{\Sigma}$ are one. A similar transformation is applied to the man's traits.

[8] Several approaches have been proposed in the literature to estimate the number of factors in general and/or panel models (see for instance Bai and Ng (2002)). However, the current context - matching models under transferable utility - generate specific issues that will be discussed below.

characteristics in the matching function, making it possible to assess the effects of matching on children's outcomes. This study is part of a large research project, CHILDROLE, that explores the role of children as decision makers within the family. The data were collected in five schools in the Metropolitan City of Naples, Italy's third-largest city, from February to April 2019. The metropolitan city is one of the most densely populated areas in Europe. Naples and the surrounding towns are marked by sharp income and cultural differences, and offer a good setting for collecting representative sample data on Italian households. Campania region also provides the perfect setting to study inequalities in children's academic performance. According to the 2019 INVALSI (National Institute for the Educational Evaluation of Instruction and Training) Report, students' skills in Italian language and Mathematical knowledge were homogeneous across Italy with the exception of Campania and Sardinia, two regions that achieved lower results compared to the national average.

Families were recruited through schools that agreed to participate and classes were randomly selected to take part in the study. Five schools (three elementary, one middle, and one high school) agreed to take part. The schools, all public, are located in different districts of the city and nearby towns, with a good socio-economic mix. The parents were surveyed through a face-to-face interview and a pencil-and-paper questionnaire. To avoid reciprocal influence, fathers and mothers were asked to complete the questionnaire in separate rooms. Children were interviewed individually in class using the same questionnaire format. Younger children were helped to fill in the questionnaire during one-to-one interviews.

Since our sampling design is based on the presence of school-age children, a (stable) match between two individuals should be interpreted as the decision to bring up a child together. Since a vast majority of women eventually have children,[9] our sample is fairly representative of the entire matched population, although we only look at couples after they had at least their first child. Hence, our sample contains a group of individuals who do not necessarily belong to the same age range, but who started a family and had children at roughly the same time.[10] Moreover, it looks at couples who are particularly stable, in that they stayed together to raise their children at least up to school age.

3.1 Matching Variables

The study collected information about parents' demographics, preferences, beliefs, and individual actions. Respondents were asked to report their date and place of birth. Educational attainment is the self-reported highest level achieved: 1) Primary, 2) Middle School, 3) High School, or 4) University. Parents' self-reported height and weight are used to measure Body Mass Index (BMI). Respondents were

[9]In the Labor Force Survey (LFS) run by the Istituto Nazionale di Statistica (ISTAT), 85.6% of women aged between 45 and 49 in the provinces of Naples and Caserta in the period 2017-2019 are living with at least one of their children. Since we do not have fertility history data and some children might already have left their parents' home, this is a lower bound for the total fraction of women (in that cohort) who have had at least one child.

[10]We thank an anonymous referee for pointing out that families with more children are more likely to be sampled. In order to account for this, we can estimate the model with a sample of firstborns only. Moreover, we can reweight our observations by the inverse of the number of children present in the household. We report the findings from these robustness checks in Table 18 in Appendix.

also asked about their health-related actions, measured by three variables: smoking, physical exercise, and propensity for healthy eating. Smoking and physical exercise are measured by multiple-choice questions with three possible answers: "Never" (coded as 0), "Seldom" (1), and "Often" (2). Propensity for healthy dietary choices was assessed through an incentivized question, with each parent choosing a snack to consume after completing the questionnaire.[11] A generic question, "Do you worry about your own health?", also investigated respondents' concern for their own health.

Next, in order to collect information on children's and parents' risk attitude, we drew from Weber et al. (2002), who originally developed a series of questions to measure risk aversion in different conventional domains, including health, safety, and recreational activities. Respondents were asked to agree or disagree with eight statements adapted from the original Weber's questionnaire. These statements explore how risk aversion and risky behavior vary among household members, including children and youths. Three explore risk preferences and actions in the health and safety domain (e.g., "I wear sunscreen to avoid sunburns"), while the other five deal with preferences for recreational risk (e.g., "I would go on a jungle safari"). For further details, see our Appendix B or Guerriero et al. (2018).

3.2 Outcome Variables

As part of the CHILDROLE data project, we also collected information about objective outcomes, such as children's academic performance and anthropometric measures, and subjective indicators, such as children's subjective wellbeing and perceived quality of the relationship with their parents.

The first set of outcome variables relate to children's cognitive competence. From the class register, we observe grades in Italian and math, and whether the child failed the academic year. Additionally, we acquired information on children's nationally standardized OCSE-PISA test results in Italian and math collected by the INVALSI. Hence, we measure children's ability to defer gratification using a simple incentivized experiment assessing whether they prefer to receive one snack today rather than two snacks tomorrow.[12]

The second set of variables relate to children's physical and mental health. Height and weight were directly measured by the interviewers. Using parents' answers, we also have information on children's weight at birth, which has been found to be positively associated with better longer-run outcomes such as IQ, education, and earnings (Black et al., 2007). Cognitive and affective evaluations of one's life were measured using three different questions (Bradshaw, 2015; Nima et al., 2020). The first question investigates how children evaluate their life as a whole with the generic question: "Are you satisfied with your own life?"[13] The affective component of

[11] Parents were shown three different types of snack before the interview and were asked to select one to have immediately after the interview. The choices were: a banana, a Parmesan bar, and a chocolate muffin. According to the Centers for Disease Control, the three correspond to different degrees of healthiness: respectively, very healthy, healthy and unhealthy.

[12] Children were asked whether they wanted to wait one day to receive two (instead of one) of their most preferred snack. Children could choose between three alternatives: a banana, a Parmesan bar, and a chocolate muffin. The snacks were left in the class to show that they would be readily available the day after, to reduce potential distrust about actually receiving the rewards, and because the visibility of the reward itself has been associated with the ability to defer gratification (Mischel et al., 1989; Watts et al., 2018).

[13] During the pilot phase we checked children's comprehension of the question and their capacity to transform their own evaluations to a 5-level Likert scale following González-Carrasco et al.

subjective well-being assesses emotions that people experience in daily life. After extensive piloting, we selected two relevant questions that were easily understood by all age groups in order to measure this affective component: "How often do you feel tired?" and "How often do you feel happy?"

The third set of variables relate to attitudes towards a healthy lifestyle in children. We investigate what they like, what they think to be healthy, and whether they can restrain themselves from eating unhealthy snacks when given the option. For this purpose, we ask them to indicate how much they like three different snacks (a banana, a Parmesan bar, and chocolate) and which one they would like to eat as a reward after the survey. In addition, we ask children whether they worry about their health, how often they eat vegetables and fruit, practice sports, smoke, and consume soft drinks. Finally, we ask them how much time they spend every day using a tablet, smartphone, and watching TV.

The fourth set of variables in our study measures the amount of time children spend with their parents and the type of relationship they have. We asked parents how much time they spend outside the home every day and we asked children how much time they spend with their parents. In order to investigate the relationship between the child and their parent we asked children whether they get along with their mother and father. Finally, we asked children how they perceive the role their parents play within the family.[14]

The fifth set of variables measures children's financial autonomy. The ability to deal with money is an essential skill that people must acquire to successfully function in society (McCormick, 2009; Suiter and Meszaros, 2005). During childhood and adolescence, the availability and use of money is also a powerful measure of parents' non-paternalistic altruism and of their lack of control over children's choices (Barnet-Verzat and Wolff, 2002). Nevertheless, little is known on how much pocket money young children receive from their parents. In our study, we collect information about the frequency and amount of pocket money allowances.

The sixth set of variables considered in this study measure children's risk preferences and actions in the health and safety domain, using the same questions posed to parents. Parents can mold their children's preferences in order to align them with their own. A study conducted by Dohmen et al. (2012) on a large sample of German citizens shows that willingness to take risks and to trust others are transmitted from parents to children.

Finally, we measured children's degree of altruism. We asked children whether they would try to help a classmate in trouble to assess their generic altruism, and whether they would give money to a classmate that needs to buy a snack to assess their non-paternalistic altruism (Guerriero et al., 2018).

3.3 Summary Statistics

In the context of this study, 632 children were surveyed in schools. For 332 of them, at least one parent participated in the survey, and for 276 of them, both parents participated in the survey.[15] The latter constitute the core of our sample, since

(2015).

[14]Children were asked who works and who takes care of the home in order to identify if tasks are equally shared or parents specialize (Amato and Booth, 1995; Kaufman, 2000).

[15]Table 12 in Appendix describes parents' participation to the survey. About 50% of all mothers participated, while only 44% of all fathers did. When asked about their family composition, 97% of all children reported living with *both* parents, as described in Table 13. According to the ISTAT

information on both parents is necessary for our empirical analysis.

Table 1 reports summary statistics on parents' education and age. We have complete information on age, educational attainment, height, and BMI for 254 of 276 participant couples. Henceforth, this group of 254 couples will be termed Sample 1. Table 2 reports summary statistics on parents' health and recreational risk behavior. 201 of 276 participant couples had both parents complete the behavioral questionnaire. This group of 201 couples will be termed Sample 2. In Table 1, we see that men are on average 43 years old and women 40 years old. Women are also slightly better educated.[16] The average man in our sample is overweight according to WHO standards (BMI greater than 25.0), while women are on average in the healthy range (BMI between 18.5 and 24.9). Table 2 shows summary statistics on risk attitude. Women are slightly less likely to smoke but also less active. Preferences for healthy snacks and concerns about own health are on average similar across genders. Gender differences are more pronounced for specific health and recreational risk behaviors. Women are more likely to use sunscreen at the beach, more scared of riding a fast moped, and less interested in doing extreme sports.

Table 1: Summary statistics - parents

	Mean	St Dev	10th P	90th P
Mother's education	2.9	0.8	2.0	4.0
Mother's age	39.6	6.7	31.0	49.0
Mother's height (cm)	163.6	5.4	158.0	170.0
Mother's BMI	24.5	3.9	20.5	29.4
Father's education	2.7	0.8	2.0	4.0
Father's age	43.3	7.3	35.0	53.0
Father's height (cm)	175.5	6.7	168.0	184.0
Father's BMI	26.8	3.4	23.0	31.1

Notes. 254 observations (Sample 1). The table reports mean, standard deviation, 10th and 90th percentile of each variable. Education is coded as a four-category variable. Individuals that reported implausibly low anthropometric measures were excluded (height below 130cm or weight below 40kg).

Labor Force Survey, the fraction of children aged between 6 and 18 who live with both parents in the provinces of Naples and Caserta is lower (84%). This figure is in line with the corresponding country-level statistic for Italy, but is slightly higher than the OECD average, and much higher than the same statistic for the UK, France, and the U.S.

[16]The age and education patterns in our estimation sample are close to those documented by the ISTAT Labor Force Survey for mothers and fathers of school-age children residing in the Naples and Caserta provinces, with the average education of parents in our sample being only slightly higher than what found in the ISTAT representative sample. See Table 15 for summary statistics of the ISTAT Labor Force Survey sample.

Table 2: Summary statistics - parents

	Mother		Father	
	Mean	St Dev	Mean	St Dev
Smokes	0.6	0.8	0.8	0.9
Does sports	0.8	0.7	0.9	0.7
Likes healthy snacks	0.8	0.8	0.8	0.8
Puts sunscreen	1.4	0.7	1.1	0.8
Washes hands	1.8	0.4	1.8	0.5
Worries about own health	1.5	0.6	1.5	0.6
Would do a safari	0.8	0.8	0.8	0.8
Hates speed	1.4	0.7	0.7	0.7
Likes usual vacation	1.1	0.7	1.1	0.7
Likes extreme sports	0.2	0.5	0.4	0.6
Crosses carefully	1.9	0.3	1.8	0.5

Notes. 201 observations (Sample 2). The table reports mean and standard deviation of each variable. Possible answers are (0) never, (1) sometimes, (2) often. The only exception is the question about healthy snacks, for which possible answers are (0) chocolate muffin, (1) Parmesan bar, (2) banana, ordered from the least healthy to the healthiest. See Table 10 for more details.

4 Results: Matching Patterns

4.1 Reduced Affinity Matrix

We start with a small-size version of our model, in which we only consider matching patterns on four socio-demographic characteristics: age, education, height, and BMI. The corresponding matrix is displayed in Table 3. All diagonal coefficients are positive and significant, implying that, for all characteristics, homogamy increases the surplus generated by the match. Not surprisingly, the largest and most significant association relates to age, a feature that essentially reflects the presence of several cohorts among parents. Strongly significant is homogamy on education, which was expected, BMI, and height.[17] Additional patterns emerge; for instance, more educated men tend to have older and thinner wives (while the opposite is not significant); and more educated wives tend to have taller husbands.

The factor decomposition is given in Table 4. The last line indicates that the first factor (Index 1) weighs three times more than the three other factors combined, and mostly reflects age differences. In other words, parents in our sample belong to different "cohorts" (defined by year of birth), and people tend to marry a spouse from a cohort that is close to their own. It should be noted that, in our sample, age is positively correlated with education, reflecting the well-known fact that more edu-

[17]Positive estimates for the diagonal coefficients could also capture the fact that individuals only meet potential partners with similar traits on their local marriage market. In order to shed more light on this issue, we estimate an alternative model where we divide couples into clusters based on the school where they have been interviewed. We report the findings in Table 18 in Appendix. Comparing columns (5) and (7), we do find that, after accounting for the presence of clusters, the estimated educational and age complementarities are slightly weaker than what initially found. For other matching variables though, we do not find significant differences with our benchmark results.

Table 3: Estimated affinity matrix (Sample 1)

Husband \ Wife	Education	Age	Height	BMI
Education	**0.77**	**0.62**	0.13	**−0.26**
	(0.11)	(0.18)	(0.09)	(0.10)
Age	0.23	**3.28**	0.25	0.07
	(0.16)	(0.33)	(0.13)	(0.13)
Height	**0.24**	**0.38**	**0.17**	−0.07
	(0.09)	(0.14)	(0.07)	(0.08)
BMI	0.10	0.20	0.00	**0.28**
	(0.08)	(0.14)	(0.07)	(0.07)

Notes. 254 couples. Standard errors in parentheses. They are calculated as in Dupuy and Galichon (2014). Boldfaced estimates are significant at the 5% level. All matching variables are scaled by their standard deviation.

Table 4: Saliency analysis (Sample 1)

	Men		Women	
	Index 1	Index 2	Index 1	Index 2
Education	**0.21**	**0.93**	**0.12**	**0.92**
	(0.02)	(0.02)	(0.02)	(0.02)
Age	**0.97**	**−0.23**	**0.99**	**−0.12**
	(0.01)	(0.02)	(0.00)	(0.02)
Height	**0.12**	**0.28**	**0.08**	0.14
	(0.02)	(0.06)	(0.03)	(0.08)
BMI	**0.06**	−0.04	0.01	**−0.36**
	(0.02)	(0.05)	(0.02)	(0.05)
Index share	**0.74**	**0.17**	**0.74**	**0.17**
	(0.07)	(0.02)	(0.07)	(0.02)

Notes. The table reports men's and women's singular vectors, V and U respectively, and singular values, $diag(\Lambda)$, from the singular value decomposition of $\hat{A} = U'\Lambda V$. We report standard errors in parentheses; they are obtained with 2,000 bootstrap replications (Milan and Whittaker, 1995). Boldfaced estimates are significant at the 5% level. In the last line, each value of $diag(\Lambda)$ can be interpreted as the relative importance of each sorting dimension.

cated people tend to both marry and have children later. This explains the positive and significant impact of education on the first factor. The second factor (Index 2) is particularly interesting. It singles out individuals who are more educated, as well as taller and, at least for women, thinner. In other words, matching patterns, while primarily driven by age, also capture a mix of education and physical appearance, possibly reflecting various dimensions of social status.

4.2 Global Affinity Matrix

Here, we look at the global (15×15) affinity matrix. A first and very striking feature is the high level of homogamy that prevails within the population. *Each* of the 15 diagonal coefficients in Table 5 is positive, indicating positive assortativeness along that specific dimension. The probability of getting such a pattern under random matching would be less than .01%. Moreover, all but one are statistically significant at 5%, and most are actually significant at 1%. This is particularly remarkable given the relatively small sample size of only 201 couples.

Table 5: Estimated affinity matrix (Sample 2)

Husband \ Wife	Education	Age	Height	BMI	Smokes	Does sports	Chooses healthy snacks	Wears sun-screen	Washes hands	Worries about health	Would go on safari	Fears speed	Likes holidays in known places	Would do extreme sports	Careful when crossing
Education	**0.29**	**0.31**	**0.09**	**−0.12**	−0.03	0.05	**0.13**	0.04	0.01	−0.03	0.05	−0.02	**−0.11**	0.05	−0.06
	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
Age	**0.16**	**0.91**	**0.09**	0.05	0.05	0.04	0.03	**0.16**	0.02	−0.01	−0.07	0.02	**−0.12**	0.05	−0.05
	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.03)	(0.03)
Height	**0.10**	0.04	**0.04**	**−0.06**	0.00	−0.01	0.01	−0.01	0.02	−0.04	−0.00	−0.02	−0.02	0.01	−0.01
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	(0.03)
BMI	−0.02	0.04	−0.00	**0.11**	−0.05	−0.03	−0.01	0.05	0.02	0.02	−0.03	0.02	−0.02	−0.01	−0.06
	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	(0.03)
Smokes	**−0.09**	−0.02	−0.03	−0.01	**0.13**	−0.02	−0.08	−0.04	−0.01	−0.03	0.01	−0.01	−0.03	−0.01	**0.08**
	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Does sports	−0.07	0.01	0.00	−0.04	−0.00	**0.05**	−0.04	−0.04	0.00	0.05	0.04	0.00	0.02	0.00	0.07
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)
Chooses healthy snacks	**0.12**	0.01	−0.01	−0.03	−0.02	0.01	**0.26**	0.04	0.02	−0.03	0.02	0.06	0.06	0.02	−0.03
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
Wears sunscreen	0.07	−0.03	0.01	0.00	**−0.08**	0.01	**0.12**	**0.06**	−0.01	0.02	−0.05	−0.00	0.03	−0.00	−0.06
	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)
Washes hands	**−0.08**	0.04	0.03	0.02	−0.01	−0.01	0.00	0.03	**0.10**	0.03	−0.03	0.01	0.03	−0.03	0.07
	(0.04)	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	(0.03)
Worries about health	−0.06	0.01	0.02	0.01	−0.05	0.03	−0.03	0.02	0.01	**0.07**	−0.02	−0.02	−0.02	−0.01	0.02
	(0.04)	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)
Would go on safari	0.03	−0.06	0.00	−0.07	−0.01	0.03	0.02	−0.07	0.03	0.01	**0.15**	0.03	−0.03	0.02	0.05
	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.03)	(0.04)
Fears speed	−0.05	0.04	−0.04	0.05	0.04	−0.01	0.04	0.01	−0.01	−0.01	0.00	**0.05**	0.04	0.00	0.01
	(0.04)	(0.04)	(0.02)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Likes holidays in known places	0.00	0.02	−0.05	0.06	−0.06	0.01	0.05	−0.01	−0.05	−0.03	0.01	**0.07**	**0.15**	0.03	**−0.08**
	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.04)	(0.04)
Would do extreme sports	0.04	0.06	0.01	−0.02	0.01	0.01	0.06	0.01	0.01	−0.02	0.01	0.01	0.01	0.01	0.00
	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
Careful when crossing	**−0.15**	−0.04	0.00	−0.03	0.06	0.02	−0.01	−0.01	**0.06**	0.05	−0.00	−0.01	0.01	−0.03	**0.15**
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)

Notes. 201 couples. Standard errors in parentheses; they are obtained with 2,000 bootstrap replications. Boldfaced estimates are significant at the 5% level. The estimates are obtained with the penalized Maximum Likelihood Estimation technique described by Dupuy et al. (2019). A four-fold Cross Validation procedure results in the upper bound $rank(A) \leq 8$. All matching variables are scaled by their standard deviation.

Several aspects revealed in the reduced matrix are still visible here - for instance, a positive and significant interaction between his education and her age, or between her education and his height, as well as a negative interaction between his education and her BMI. Others are less expected. Wives of more educated men are more likely to choose healthy snacks and to favor vacations at an unknown place; husbands of more educated women are more likely to eat healthy food and less likely to smoke, but are also less careful when crossing a street[18] and wash their hands less often.

It should be noticed that spouses might become more similar in terms of health-related behavior and risk attitude *after* the marriage. In this case, our estimates of the diagonal coefficients would be upward biased.[19] However, there is evidence in the existing literature that some of these traits do not significantly alter after marriage. Dohmen et al. (2012) find that correlation patterns between spouses' risk attitude and trust do not change over the relationship. Using retrospective information for couples aged 50 and older, Jackson et al. (2015) show that health-related behaviors are very persistent through the life-cycle, and only a small fraction of individuals successfully quit smoking or lose weight while getting older. They argue that convergence between spouses' healthy habits is less likely if one partner consistently maintains a healthy lifestyle. On the other hand, partners are more inclined to make positive changes at the same time, i.e., married individuals make healthy changes in their behavior if their partners also do so.

The factor decomposition enriches the conclusions drawn from the reduced matrix.[20] Factor loadings are reported in Table 6. The first factor, which accounts for about a third of the total systematic surplus by itself, essentially recaptures the cohort pattern observed on the reduced matrix. As can be seen in the first and fourth column of Table 6, age plays a dominant role in the first sorting dimension, while the role of education follows from the observed positive correlation between education and both age at marriage and age at first birth. Height, healthy habits, and attitude towards recreational risk play a much smaller role.

The second factor is of a different nature. It singles out individuals who are younger, more educated, taller (men) or thinner (women) and more health-conscious. They are more likely to eat healthy food and wear sunscreen, and less likely to smoke or experience health problems. All in all, Index 2 appears to capture various dimensions of parents' *human capital*, which includes not only education but also health. In turn, this dimension is likely to be highly correlated with social status. While our data does not allow us to analyze this aspect in detail, one can note that parents with a high Index 2 tend to live in more residential neighborhoods (hence, the reduced requirement to pay attention when crossing the street) and to wash their hands less often (possibly because they are less likely to have a manual

[18]Street crossing behavior also reflects the safety of the neighborhood where the family lives. Individuals who live in safer places, with more pedestrian areas and less dangerous crossing points, are less careful when crossing the street. Only 71% of parents from students at the centrally located Umberto I high school report being very careful when crossing the street, compared to 87% in other schools. Even after controlling for education, parents of students enrolled in Umberto I high school are the least likely to be careful when crossing the street.

[19]Unfortunately, we do not have information on relationship duration, and thus we cannot check how spousal correlation patterns change with duration.

[20]In Appendix C, we plot the marginal distributions of Index 1 and 2, as well as their joint distributions. This helps visualize the strength of sorting on different dimensions. By definition, sorting is stronger on the first dimension.

Table 6: Saliency analysis (Sample 2)

	Men			Women		
	Index 1	Index 2	Index 3	Index 1	Index 2	Index 3
Education	**0.40**	**0.45**	**0.44**	**0.28**	**0.62**	**0.23**
	(0.02)	(0.05)	(0.06)	(0.02)	(0.04)	(0.06)
Age	**0.90**	**−0.30**	**−0.11**	**0.91**	**−0.27**	**−0.06**
	(0.01)	(0.02)	(0.03)	(0.01)	(0.02)	(0.03)
Height	**0.07**	**0.15**	**0.20**	**0.12**	0.03	**0.19**
	(0.03)	(0.06)	(0.08)	(0.03)	(0.06)	(0.08)
BMI	0.04	−0.06	**−0.34**	0.00	**−0.17**	**−0.55**
	(0.03)	(0.05)	(0.07)	(0.03)	(0.05)	(0.06)
Smokes	**−0.06**	**−0.32**	**0.21**	0.02	**−0.25**	**0.23**
	(0.03)	(0.05)	(0.06)	(0.02)	(0.05)	(0.06)
Does sports	−0.03	**−0.15**	**0.18**	0.05	0.03	0.13
	(0.03)	(0.06)	(0.07)	(0.03)	(0.06)	(0.07)
Chooses healthy snacks	**0.07**	**0.51**	**−0.13**	**0.10**	**0.55**	**−0.14**
	(0.02)	(0.04)	(0.06)	(0.02)	(0.04)	(0.05)
Wears sunscreen	0.01	**0.30**	**−0.24**	**0.16**	0.02	**−0.24**
	(0.03)	(0.06)	(0.08)	(0.03)	(0.06)	(0.08)
Washes hands	0.01	**−0.18**	−0.07	0.02	−0.06	0.12
	(0.03)	(0.05)	(0.06)	(0.03)	(0.05)	(0.06)
Worries about health	−0.01	**−0.13**	−0.03	−0.03	**−0.13**	0.00
	(0.03)	(0.05)	(0.07)	(0.03)	(0.05)	(0.07)
Would go on safari	**−0.06**	**0.13**	**0.42**	−0.05	**0.13**	**0.34**
	(0.03)	(0.05)	(0.06)	(0.03)	(0.05)	(0.06)
Fears speed	0.02	−0.07	**−0.21**	0.01	0.07	**−0.18**
	(0.03)	(0.06)	(0.08)	(0.03)	(0.06)	(0.07)
Likes holidays in known places	0.00	**0.16**	**−0.47**	**−0.14**	0.09	**−0.41**
	(0.02)	(0.04)	(0.06)	(0.03)	(0.05)	(0.05)
Would do extreme sports	**0.07**	0.09	0.03	**0.07**	0.09	0.02
	(0.03)	(0.07)	(0.08)	(0.03)	(0.06)	(0.08)
Careful when crossing	**−0.09**	**−0.32**	**0.19**	**−0.09**	**−0.29**	**0.36**
	(0.03)	(0.05)	(0.07)	(0.02)	(0.04)	(0.06)
Index share	**0.38**	**0.18**	**0.13**	**0.38**	**0.18**	**0.13**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)

Notes. The table reports men's and women's singular vectors, V and U respectively, and singular values, $diag(\Lambda)$, from the singular value decomposition of $\hat{A} = U'\Lambda V$. We report standard errors in parentheses; they are obtained with 2,000 bootstrap replications (Milan and Whittaker, 1995). Boldfaced estimates are significant at the 5% level. In the last line, each value of $diag(\Lambda)$ can be interpreted as the relative importance of each sorting dimension.

occupation). In addition, some traits are more idiosyncratic, such as individuals with a high Index 2 being fond of holidays in known destinations, possibly because they own a secondary residence.

Finally, the third factor emphasizes yet other traits, several of which are related to lifestyle in general and physical shape in particular. Individuals with a high Index 3 are younger, more educated, but also taller, thinner, and more likely to do sports. Yet, they are more likely to adopt risky behavior in the health and recreational domains, since they are also less likely to choose healthy snacks and wear sunscreen, but more likely to smoke. Moreover, they dislike usual holidays, would like to go on safari, and are not scared of speed. All in all, these traits may suggest a more hedonistic approach to life in general - and to relationship with children in particular.

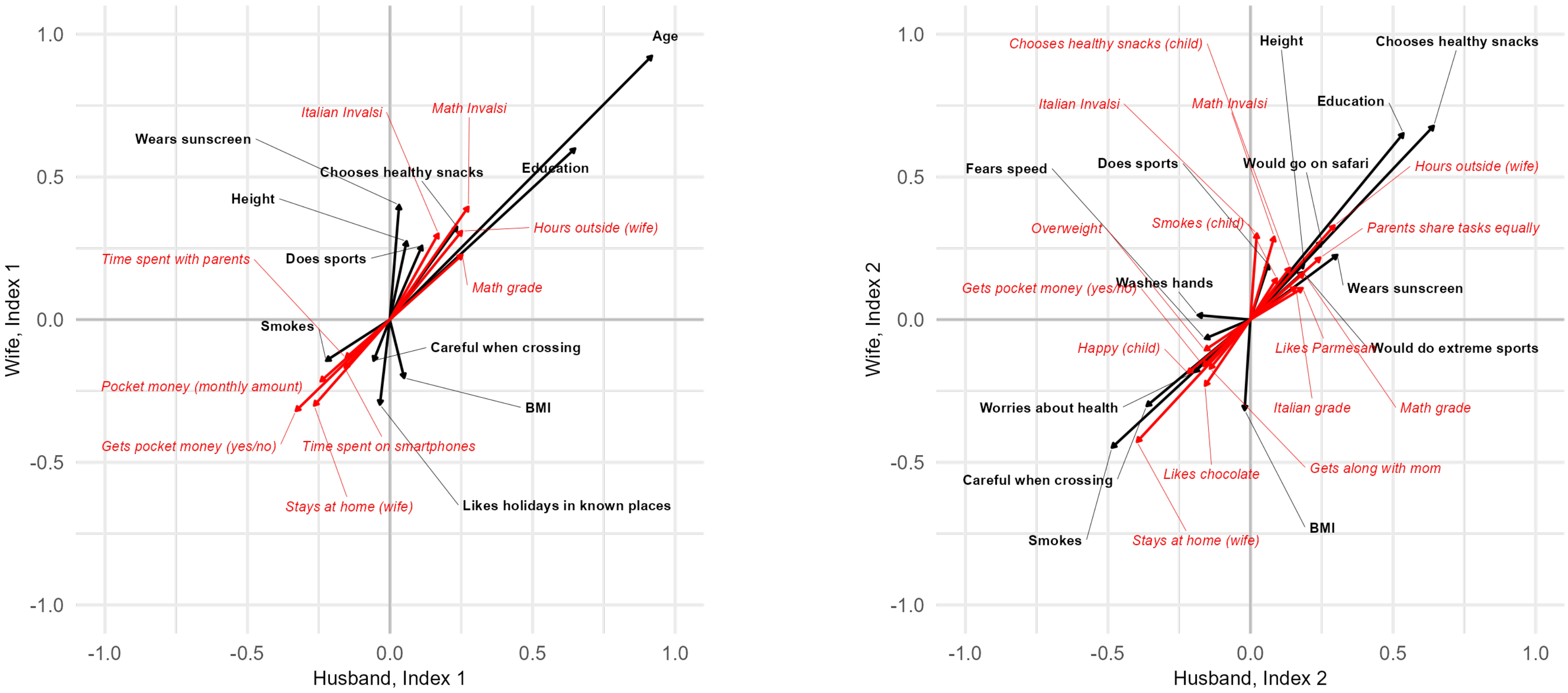
Together, these three factors explain more than the remaining 12 factors combined. A formal test, presented in Table 17, suggests that these factors are in fact sufficient to fully summarize the matching affinity matrix; the null hypothesis that its rank is less than or equal to 3 is not rejected.

5 Results: Child Outcomes

An interesting aspect of our data is that they include what could be considered as "outcome" variables, i.e., indicators reflecting choices made within the household and their consequences. A standard example is labor supply behavior. While most married men are active on the labor market, women may or may not participate, and these decisions appear to be related to matching patterns. Even more topical is the impact of matching on children. In our data, we observe both objective outcomes, such as grades and academic performance, but also general behavior; and more subjective indicators, such as the child's subjective well-being and the perceived quality of the relationship with their parents. A list of all outcome variables is provided in Appendix A.

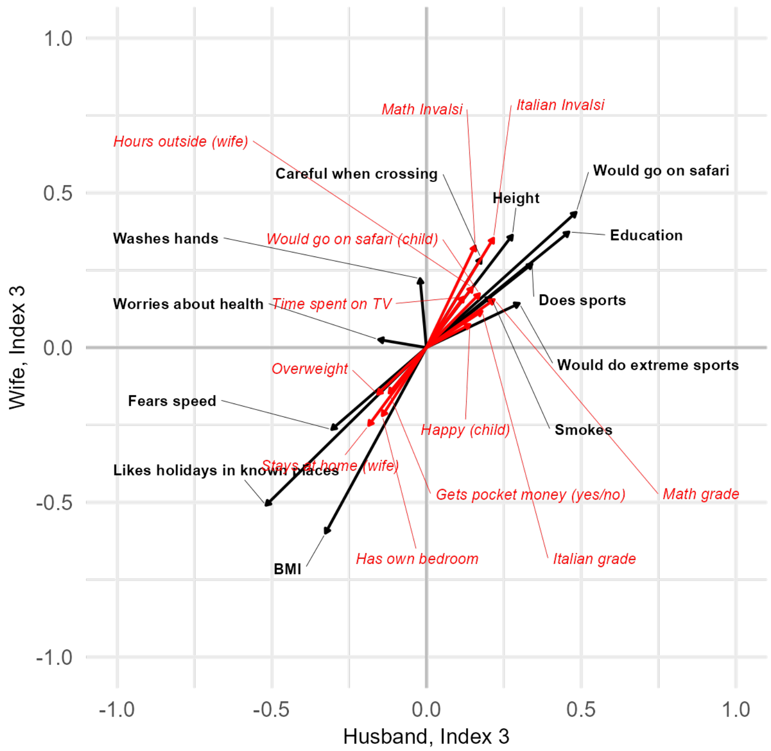
We first provide a graphical visualization of the correlation patterns between parents' matching factors and family outcomes using simple Pearson correlation rates. Then, we run a series of regressions relating these outcomes to parents' characteristics in general, and to our three factors in particular. Finally, we select certain outcomes based on our preliminary analysis, namely child class grades and different measures of children's well-being, and assess the relative importance of each factor (and of their interactions) in explaining the outcome variation.

Figure 1: Correlation between matching indices and matching variables (black) and between matching indices and outcomes (red) (Sample 2)



(a) Index 1

(b) Index 2



(c) Index 3

Notes. We plot correlation rates of both matching variables and outcome variables with the husband's Index k (x-axis) and the wife's Index k (y-axis). In the North-East region, we find the defining characteristics of households where both spouses rank high along the k-th matching dimension, whereas in the South-West region, we find the characteristics of households where both spouses rank low along the same dimension. For instance, in Figure 1b, we see that parents with a high Index 2 are more educated (education being positively correlated with Index 2 for both husbands and wives) and the wife is less likely to stay at home (the wife staying at home being negatively correlated with both the husband's and the wife's Index 2). The matching variables (in black with boldfaced labels) include all the parents' background variables used in the main estimation. The outcomes (in red with labels in italics) include additional variables that are excluded from the main estimation, e.g., the number of children, the child's grades, and the wife's labor supply. In order to improve the readability of the graph, we only plot those variables whose correlation rate is significantly different from zero at the 5% level in a two-tailed test. All correlation rates and the respective p-values are reported in Table 19 in Appendix B.

5.1 A Graphical Illustration

The relationship between factors, matching traits and outcomes can first be illustrated in a series of graphs (Figures 1a, 1b, and 1c). Their interpretation goes as follows. The factors driving matching patterns define a low dimensional subspace. We plot the projection, on the corresponding subspaces, of both the variables describing marital traits, which were used to construct the factors, *and* our outcome variables, which were absent from the factor estimation but may nevertheless be significantly correlated to them. In other words, families where both parents rank high on a certain matching dimension, thereby having a high value for the corresponding factor, will be characterized by high (low) values for the variables that we find in the North-East (South-West) region of the graph. Not all correlation patterns are perfectly gender symmetric; variables that are only correlated with the husband's (wife's) index are located along the horizontal (vertical) axis. In principle, certain variables could be correlated positively with the husband's index but negatively with the wife's (and vice versa), but in practice the second and fourth quadrants are almost empty. In order to improve the readability, in the Figures, we only keep variables whose correlation with the corresponding factor is significant at the 5% level, whereas Table 19 in Appendix B reports all correlation rates between matching factors and outcome variables, as well as their respective p-values.

We find that parents with a high Index 1 are older. This characteristic appears to be correlated not only with wives being more likely to work outside the home, but also with children reporting lower levels of subjective well-being and worse relationships with their parents. These correlations, however, are spurious, since they mostly reflect the child's age. After accounting for the child's age, they weaken or disappear. In the plots, we only report the correlations between the indices and the residualized child's outcome variables obtained after regressing them on the child's age.[21] All in all, the cohort component of the matching patterns does not seem to be strongly correlated with any output. Yet, since older parents are also often more educated, a higher Index 1 is associated with higher grades, particularly in math. Children of parents with a high Index 1 are also less likely to receive pocket money and to spend time on smartphones, but also to spend time with their parents; this might be explained by generational differences in parenting style, a point that we will discuss in greater detail later.

Things are quite different with the second factor, which we interpret as being mostly driven by parents' human capital. In Figure 1b, we see that women with a high Index 2 are more likely to work outside the home; interestingly, these couples are also more likely to share domestic tasks equally. More importantly, high levels of Index 2 are strongly correlated with better educational outcomes, as measured by children's grades, including both nationally standardized INVALSI test results (math and Italian) and class grades (always math and Italian, measured as the deviation from the class median). These children are less likely to like chocolate, more likely to prefer Parmesan cheese, and less likely to be overweight. When asked to choose a snack, they opt for healthier options. Interestingly, Index 2 is also strongly correlated with children reporting to be less happy, less likely to get along with their

[21]In other words, we only use the variation within the child's age cohort. We apply this correction to all variables that are expected to change as children grow older. Some variables are already standardized and thus do not need this correction (e.g., BMI, standardized Italian and math tests). The full list is provided in Table 11.

mothers, and also slightly more likely to smoke. A possible interpretation is that educated parents' investment in their children's human capital, while fruitful in terms of health-related habits and academic performance, comes at a price, since children appear to resent the corresponding pressure. One explanation is that children of working parents might resent their absence from home (Heinrich, 2014). Another explanation is that the parenting style of educated parents is more intrusive, and possibly more authoritative (Doepke and Zilibotti, 2017). This means that educated parents pressure children to perform well at school, which might result in a welfare loss for the child.

This is in stark contrast with couples exhibiting a high level of Index 3, which is shown in Figure 1c. Their children also have better grades and less pocket money, but watch TV more often, are more likely to share their bedroom with their siblings, are more excited about the idea of going on safari, and are less likely to be overweight. Also in this case, mothers with a high Index 3 are more likely to work, but the correlation is weaker than on the first two dimensions. Last but not least, children of parents with a high Index 3 appear to be overall happier than their (nerdier) peers. In Section 4.2, we highlighted how parents with a high Index 3 are in better physical shape and more likely to adopt risky behavior in the health and recreational domains. These traits might have a direct positive impact on children (e.g., more outdoor activities) or might reflect into a more permissive parenting style.

What is remarkable here is that the factors are recovered *exclusively from matching patterns*; neither parental investment nor any output variable is used for their estimation. Yet, parents' human capital, the most important factor driving assortativeness (after parental age/cohort), appears to be strongly correlated with both children's human capital achievements (positively) and well-being (negatively). This strongly suggests not only that future investments in children's human capital are an explicit part of individuals' marital strategies, but also that these aspects are crucially important, since the corresponding factor dominates all other patterns.

5.2 All Characteristics & All Outcomes

We first run a series of regressions of all observable outcomes over all parental characteristics, as well as over the matching indices (and their interactions). This first step should be considered as purely exploratory; that is, we want to see which outcomes appear to have any relationship with parental characteristics.

Table 7 gives, for each outcome, the (adjusted) R^2 for these regressions. A first remark is that, for many outcomes, the null of no linear correlation is not rejected; even when the R^2 significantly differs from 0, it is often inferior to 5%. While this conclusion is obviously linked to our small sample size, it still indicates that even our rich data set fails to capture the full complexity of many parent-children interactions. Yet, a few outcomes appear to be significantly correlated with parental characteristics. Besides the mother's labor supply behavior, this is the case for outcomes linked with children's human capital, such as grades, as well as certain measures of children's happiness.

A second finding is that, in the vast majority of cases, the adjusted R^2 for the second set of regression (over matching factors), when significant, are equal to or even larger than those of the first set (over all parental characteristics). A notable exception is the INVALSI math grade, for which our sample size is unfortunately even smaller. This is all the more remarkable that children outcomes were not used

Table 7: Adjusted R^2 from regressing family and child outcomes on parental traits

.	Observed traits (1)	Matching indices *(w/o interactions)* (2)	Matching indices *(w/ interactions)* (3)
Num. children	−0.03	0.02	0.01
Hours outside (wife)	0.15	0.15	0.18
Stays at home (wife)	0.23	0.23	0.23
Parents share tasks equally	0.05	0.05	0.06
Height (child)	0.12	−0.01	0.03
BMI (child)	0.06	0.02	0.02
Overweight	0.00	0.03	0.02
Year failed	−0.00	−0.00	−0.01
Italian grade	0.02	0.03	0.08
Math grade	0.05	0.09	0.10
Italian Invalsi	0.23	0.22	0.28
Math Invalsi	0.39	0.23	0.23
Patience (child)	0.01	−0.00	−0.01
Sub. well-being (child)	−0.07	−0.02	−0.01
Happy (child)	0.03	0.06	0.05
Gets along with mom	0.02	0.00	0.01
Gets along with dad	−0.03	−0.01	−0.02
Likes chocolate	0.03	0.03	0.03
Likes Parmesan	0.04	0.01	0.02
Likes bananas	0.07	−0.02	−0.02
Chooses healthy snacks (child)	0.07	0.01	0.01
Gets pocket money (yes/no)	0.13	0.10	0.10
Pocket money (monthly amount)	0.08	0.04	0.06
Number of daily snacks	0.08	−0.02	−0.00
Drinks sugary drinks	−0.01	0.03	0.02
Time spent on tablets	−0.01	−0.00	0.02
Time spent on smartphones	0.02	0.01	0.04
Time spent on TV	0.01	0.03	0.03
Time spent on screens (total)	−0.02	−0.01	−0.02
Time spent with parents	0.05	0.03	0.03
Eats vegetables	0.02	0.02	0.03
Often tired	−0.05	−0.03	−0.01
Has own bedroom	0.02	0.04	0.07
Smokes (child)	−0.01	0.01	0.02
Does sports (child)	−0.02	−0.01	−0.02
Worries about own health	0.06	−0.02	−0.03
Wears sunscreen (child)	0.04	−0.02	−0.01
Washes hands (child)	−0.02	−0.02	−0.02
Would go on safari (child)	0.13	0.01	0.01
Likes holidays in known places (child)	0.00	−0.01	0.01
Would do extreme sports (child)	0.05	−0.01	−0.03
Careful when crossing (child)	0.00	−0.00	−0.01
Altruism I	0.05	−0.02	−0.01
Altruism II	0.02	−0.01	−0.01

Notes. In column (1), we report the adjusted R^2 from separately regressing each child or family outcome on the parents' observed characteristics (15 variable for the father, 15 for the mother). In column (2), we report the adjusted R^2 from regressing each outcome on the parents' main matching indices (three for the father and three for the mother). In column (3), we report the adjusted R^2 obtained from regressing each outcome on both the parents' indices and their pairwise interactions. Since the number of regressors is different across columns, and the number of observations is sometimes different across rows, we report the adjusted R^2 to facilitate the comparison. Certain outcomes were residualized with respect to child's age: see Appendix A.

in the construction of the factors (which are exclusively based on parents' matching patterns).

Finally, a quick look at column (2) and (3) clearly suggests that, for several outcomes including child grades, introducing interactions between the parents' factors significantly increases the adjusted R^2. In other words, what matters is not only how large each parent's matching factor is, but also how the two sets of factors interact. Obviously, this remark has a special weight in a matching context.

5.3 Unobserved Matching Characteristics

Before we further explore the statistical relationship between matching indices and child outcomes, it is worth discussing the role played by unobserved heterogeneity. Parents match on several traits, many (if not most) of which are not observable by the econometrician. In the SEV approach, these are captured by random terms (that are moreover assumed to follow a type I extreme value distribution). However, a simple OLS regression of children outcomes on parents' characteristics (either as observed or as summarized by the matching factors) would implicitly require that these unobservable traits have no impact on children outcomes; otherwise, the fact that we only consider households that actually formed introduces a selection bias.

Given our econometric structure, a natural way of correcting for this bias is to introduce a Heckman-type control - which, in our setting, can equivalently be interpreted as the expected value of the "quality of the match" (as generated by the random shocks that summarize unobserved heterogeneity). Specifically, the expected match surplus for any pair (x, y) randomly drawn from the entire population is

$$\mathbb{E}[U(x_k, y_l) + \sigma\varepsilon_l + V(x_k, y_l) + \sigma\eta_k | x_k, y_l] = \Phi(x_k, y_l). \tag{13}$$

However, if two individuals k and l are observed as a couple, then k has optimally chosen l, and vice versa (this is the selection bias). Hence, their expected match surplus conditional on being matched is

$$\begin{aligned}\mathcal{E}(x_k, y_l) :=& \mathbb{E}[U(x_k, y_l) + \sigma\varepsilon_l + V(x_k, y_l) + \sigma\eta_l | x_k, y_l, k \text{ and } l \text{ matched}] \\ =& \Phi(x_k, y_l) - \sigma \log \mu_f(y_l|x_k) - \sigma \log \mu_m(x_k|y_l) \\ =& \Phi(x_k, y_l) - 2\sigma \log \mu(x_k, y_l) + \sigma \log f(x_k) + \sigma \log g(y_l).\end{aligned} \tag{14}$$

The expected (unobserved) match quality, normalized by σ, is thus

$$\frac{\mathcal{E}(x_k, y_l) - \Phi(x_k, y_l)}{\sigma} = -2 \log \mu(x_k, y_l) + \log f(x_k) + \log g(y_l), \tag{15}$$

and is inversely related to the probability of matching. This is the control term we introduce in our regressions.

5.4 Child Grades

Next, we concentrate on the relationship between matching factors and a few specific children outcomes. We start with class grades in math and Italian.[22] These outcomes

[22] Class grades are available for all children in the classes selected for the survey, and were taken directly from class registers. They range from 4 to 10, with 6 being the minimum passing mark. We normalize them at the class level by taking the distance from the class median. The resulting normalized grades range from -4 to 3. The INVALSI exam results, while already nationally standardized, are only available for a subsample of pupils, raising serious sample size issues.

are interesting not only because we aim to shed light on how children's human capital relate to matching patterns, but also because Table 7 suggests that the relationship between child grades and matching factors is strong relatively to other outcomes.

Table 8 gives the estimated coefficients for several ordered probit specifications.[23] Regarding Italian grades, the first two columns present a regression over parental factors and their interaction, respectively without and with control for selection. The husband's Index 2, interpreted as human capital, has a significant positive impact; more interesting, the interaction between the parents' second factor is also positive and significant. In terms of matching theory, this result suggests a complementarity between spouses' human capital, which would partly explain the high degree of assortative matching observed in the data. Grades in math exhibit similar patterns, although the corresponding coefficients are not significant at the 5% threshold - a result that is in line with the existing literature on child development.[24] Lastly, math grades are correlated positively with the father's third factor, while both Italian and math grades are negatively correlated with the interaction of parents' third factor, although the coefficient is not significant when selection is controlled for.

The next two columns, (3) and (4), present similar regressions when the matching factors are replaced by the spouses' age and education, as well as their interactions. Our goal, here, is to check whether the previous conclusions could be directly reached from an investigation of the key variables related to the parents' human capital, without the factor decomposition of the matching process. The answer is particularly interesting. We still find that the father's education is positively correlated with child grades. However, the interaction between parents' education has now a *negative* impact on grades, suggesting a relationship of substitution. Notably, the negative sign obtains for both math and Italian grades in all specifications of the model.

Finally, we run a last series of regressions in which right hand-side variables includes matching factors *and* parents' age and education. While several coefficients are no longer significant, reflecting the correlation between education and the second matching factor, the qualitative patterns remain unchanged. In particular, the interaction of parents' second factor has a positive impact on both Italian and math grades, and the impact is significant for Italian grades, while the coefficients of education interaction is negative, in all specifications.

5.5 Children's Well-being

Another set of interesting outcomes relates to various measures of children's welfare. In the survey, we use four indicators: children's self-reported happiness and well-being, as well as the quality of the relationship with both parents, always as reported by the child. Again, we regress these variables over the parents' matching factors and

[23]Given the discreteness of child grades, ordered probit models are a natural choice. Previously, in Table 7, we estimated a linear model of child grades on matching indices in order to compare the adjusted R^2 with other family outcomes. If we stuck to a linear model also throughout this section, the qualitative findings would be very similar.

[24]While the link between parents' education and their children's literacy skills is firmly established, a higher level of parental human capital does not consistently translate into better math skills for children (Siegler et al., 2019). A possible explanation for this discrepancy lies in the preference that parents often exhibit for participating in literacy-related activities with their children, such as reading bedtime stories, as opposed to engaging in math-related activities like counting and number naming (Hart et al., 2016). In Italy, pediatricians recommend literacy-related activities for newborns from the earliest months of life while numerical activities are not emphasized or recommended (Pizzi et al., 2022).

Table 8: Regressing grades on matching traits

	Italian grade					Math grade				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Index 1 (husb)	0.06	0.07			0.31	0.23	0.23			−0.01
	(0.69)	(0.67)			(0.60)	(0.13)	(0.13)			(0.98)
Index 2 (husb)	0.24	0.22			0.13	0.19	0.18			0.14
	(0.01)	(0.03)			(0.36)	(0.05)	(0.07)			(0.30)
Index 3 (husb)	0.13	0.12			0.07	0.17	0.16			0.10
	(0.14)	(0.16)			(0.45)	(0.06)	(0.06)			(0.30)
Index 1 (wife)	−0.09	−0.10			−0.31	−0.02	−0.02			0.21
	(0.55)	(0.52)			(0.33)	(0.89)	(0.87)			(0.52)
Index 2 (wife)	−0.07	−0.05			−0.11	−0.06	−0.05			−0.20
	(0.50)	(0.63)			(0.43)	(0.57)	(0.63)			(0.14)
Index 3 (wife)	0.07	0.06			0.03	0.04	0.03			−0.05
	(0.40)	(0.50)			(0.73)	(0.68)	(0.74)			(0.62)
Index 1 (husb)*Index 1 (wife)	−0.12	−0.07			−0.01	−0.06	−0.04			0.06
	(0.02)	(0.28)			(0.96)	(0.21)	(0.53)			(0.61)
Index 2 (husb)*Index 2 (wife)	0.12	0.15			0.16	0.06	0.07			0.07
	(0.07)	(0.04)			(0.03)	(0.38)	(0.32)			(0.34)
Index 3 (husb)*Index 3 (wife)	−0.14	−0.09			−0.09	−0.12	−0.10			−0.10
	(0.01)	(0.19)			(0.23)	(0.03)	(0.17)			(0.19)
Educ (husb)			0.22	0.26	0.08			0.39	0.39	0.29
			(0.02)	(0.01)	(0.73)			(0.00)	(0.00)	(0.24)
Educ (wife)			0.06	0.09	0.19			0.07	0.07	0.11
			(0.52)	(0.34)	(0.13)			(0.45)	(0.47)	(0.38)
Educ (husb)*Educ (wife)			−0.17	−0.14	−0.13			−0.15	−0.15	−0.17
			(0.04)	(0.09)	(0.19)			(0.06)	(0.07)	(0.10)
Age (husb)				−0.09	−0.31				0.13	0.18
				(0.56)	(0.59)				(0.36)	(0.75)
Age (wife)				−0.07	0.19				−0.10	−0.31
				(0.66)	(0.56)				(0.52)	(0.34)
Age (husb)*Age (wife)				−0.12	−0.09				−0.06	−0.12
				(0.08)	(0.57)				(0.41)	(0.45)
Exp. match quality		0.07			0.06		0.04			0.02
		(0.34)			(0.41)		(0.65)			(0.80)
N	201	201	201	201	201	201	201	201	201	201

Notes. Coefficients obtained from ordered probit models. P-values in parentheses. Expected match quality is calculated from equation (15). Grades are normalized at the class level by taking the distance from the class median. The resulting normalized grades range from -4 to 3.

their interactions, plus the child's age and age squared, and a control for selection.

Table 9: Regressing children's subjective well-being measures on matching traits

	Sub. well-being		Happiness		Gets along with mom		Gets along with dad	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Index 1 (husb)	−0.04	−0.04	0.13	0.13	0.06	0.05	0.08	0.06
	(0.81)	(0.81)	(0.51)	(0.51)	(0.75)	(0.78)	(0.67)	(0.73)
Index 2 (husb)	−0.04	−0.04	−0.27	−0.27	−0.11	−0.08	−0.03	0.00
	(0.69)	(0.70)	(0.03)	(0.04)	(0.38)	(0.54)	(0.83)	(0.97)
Index 3 (husb)	0.13	0.13	0.26	0.26	−0.03	−0.02	−0.01	−0.00
	(0.22)	(0.23)	(0.04)	(0.04)	(0.79)	(0.87)	(0.95)	(0.99)
Index 1 (wife)	0.07	0.07	−0.23	−0.23	−0.13	−0.11	−0.03	−0.01
	(0.71)	(0.71)	(0.27)	(0.27)	(0.51)	(0.56)	(0.87)	(0.94)
Index 2 (wife)	−0.03	−0.03	−0.10	−0.10	−0.13	−0.16	−0.14	−0.17
	(0.79)	(0.79)	(0.48)	(0.47)	(0.33)	(0.23)	(0.29)	(0.21)
Index 3 (wife)	−0.16	−0.16	−0.00	−0.00	−0.02	0.01	0.05	0.07
	(0.14)	(0.15)	(0.98)	(1.00)	(0.88)	(0.96)	(0.67)	(0.52)
Index 1 (husb)*Index 1 (wife)	−0.09	−0.09	−0.08	−0.08	−0.03	−0.10	0.04	−0.02
	(0.15)	(0.28)	(0.29)	(0.37)	(0.70)	(0.28)	(0.62)	(0.78)
Index 2 (husb)*Index 2 (wife)	−0.12	−0.12	−0.05	−0.06	−0.16	−0.20	−0.07	−0.11
	(0.11)	(0.14)	(0.51)	(0.51)	(0.05)	(0.02)	(0.35)	(0.19)
Index 3 (husb)*Index 3 (wife)	0.07	0.07	−0.03	−0.04	0.02	−0.05	−0.05	−0.12
	(0.39)	(0.49)	(0.65)	(0.67)	(0.76)	(0.57)	(0.39)	(0.16)
Child's age	−0.13	−0.13	−0.08	−0.08	0.04	0.04	−0.00	−0.00
	(0.00)	(0.00)	(0.10)	(0.10)	(0.35)	(0.35)	(0.94)	(0.92)
Child's age (sq)	0.01	0.01	0.02	0.02	0.00	0.00	0.01	0.01
	(0.16)	(0.16)	(0.03)	(0.03)	(0.88)	(0.85)	(0.56)	(0.56)
Exp. match quality		−0.00		−0.01		−0.12		−0.11
		(1.00)		(0.90)		(0.21)		(0.26)
N	193	193	199	199	198	198	200	200

Notes. Coefficients obtained from ordered probit models. P-values in parentheses. Expected match quality is calculated from equation (15). Children's subjective well-being is measured on a 5-level Likert scale. Self-reported happiness is measured through the question "Are you happy?", with possible answers being "Never", "Sometimes", or "Often". The relationship quality between parents and children is measured through the question "Do you get along with your mum/dad?", with possible answers being "Never", "Sometimes", or "Often". See Appendix A for futher clarifications.

Two conclusions emerge from these regressions. First and primarily, *all* coefficients involving the mother's and father's second matching factor, and also their interaction, are negative (sometimes significantly so) in all specifications. In others words, the emphasis put by high human capital parents on their children's academic performance, while productive in terms of scholarly achievement, comes at a cost in terms of child happiness and quality of parent-child relationship. Secondly, a high Index 3 for the father appears to be positively and significantly correlated with a child's happiness - possibly suggesting the benefits of a more "hedonistic" approach to parenthood (at least when compared to "nerdier" families).

6 Counterfactual Experiments

Finally, we present some counterfactual experiments based on our estimates. We use the coefficients estimated in the outcome regressions, controlling for selection bias, to assess what the distribution of grades would be in some counterfactual contexts. It is important to stress that these counterfactual experiments should be taken for what they are - namely, as a decomposition exercise. Our data do not allow us to assess true causal impacts; our goal is rather to illustrate what our estimates would imply in terms of the respective correlations between matching patterns and children outcomes. Yet, we believe that they provide an interesting perspective on the matching process.

While many experiments of this type could be performed, we concentrate on two benchmark cases. One is the "pure" random matching scenario, in which spouses are allocated to each other independently of any (observable or unobservable) characteristic. In practice, we start with the actual sample of men and women, and re-create virtual couples by matching them randomly; we then use our estimated coefficients to predict the grade for each (virtual) couple's child, and plot the corresponding distribution.

The alternative, polar case is one in which spouses exclusively match on the second factor. The idea, here, is to capture a scenario where human capital, as proxied by the second factor, is the *exclusive* determinant of matching patterns. In practice, we calculate the equilibrium assignment resulting from the model presented in Section 2 when the expected match surplus $\Phi(x, y)$ is only given by the interaction between the husband's and wife's Index 2, i.e., $\Phi(x, y) = \lambda_2 \tilde{x}_2 \tilde{y}_2$.

In Figure 2, we plot the distribution of Italian and math grades under (i) raw data, (ii) our estimated framework, (iii) pure random matching and (iv) exclusive matching on Index 2. Our ordered probit estimations fit actual data almost perfectly. We find that random matching would significantly change the grade distribution among children. Specifically, the distribution of Italian grades in the first counterfactual experiment, plotted in gray, is significantly shifted to the left. The shift for math grades is less brutal; yet, in both cases the number of children with a grade above (resp. below) the mode of the distribution is lower (higher) under random matching. On the contrary, matching exclusively on Index 2 would shift the grades to the right, with the counterfactual distribution plotted in red, although again the shift is stronger for Italian grades.[25]

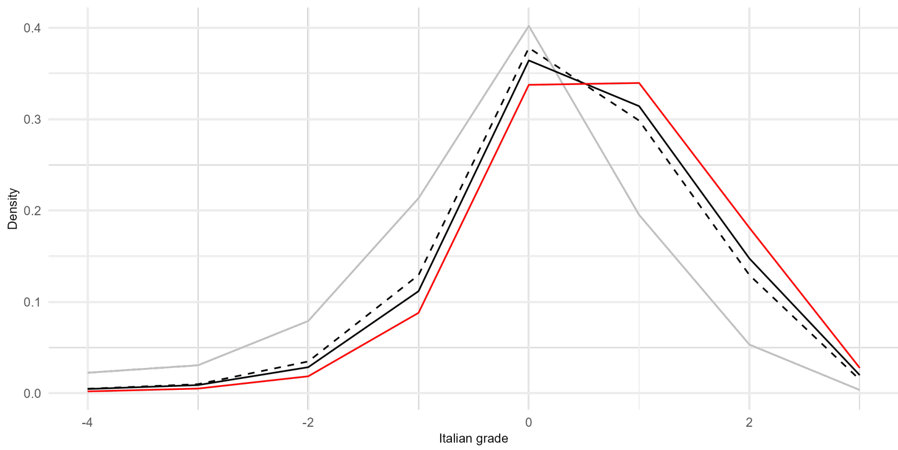
Our interpretation is that while people do match based on human capital, and while this assortativeness is correlated with better student performances, the real-life process is quite complex. Individuals match on many traits, several (and perhaps most) being unobservable. While human capital stands out as a prominent factor, it is by no means the only one, and some other traits on which spouses match assortatively may actually be substitutes in the children's production function.

7 Conclusion

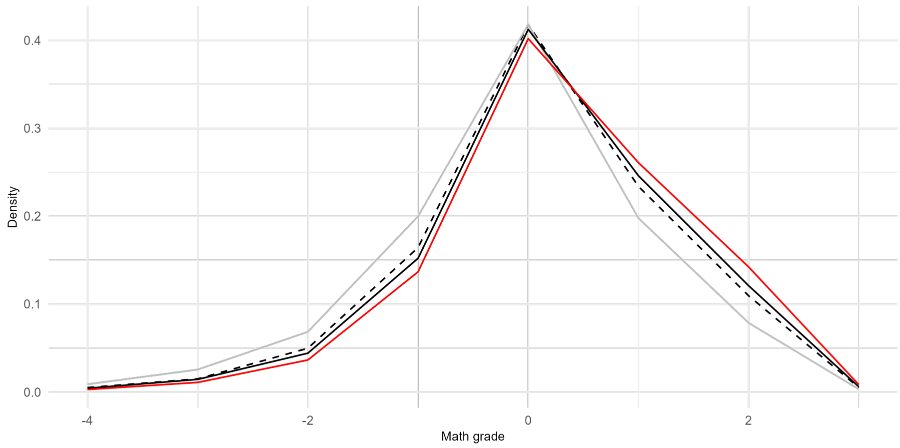
A large body of literature exists across economics, sociology, and demography that has studied homogamy and measured it using data on marital patterns. In this paper, we argue that Separable Extreme Value (SEV) models can be used to study multidimensional sorting and can easily handle numerous discrete classes and continuous variables. We show that the SEV approach can generate rich empirical findings by estimating a multidimensional and parametric model, borrowed from Dupuy and Galichon (2014), with data from a survey of parents of children attending schools in Campania, a region of Southern Italy. We show that marital patterns are char-

[25]Note that the same counterfactual experiments, if based on the regressions of grades over parents' age and education only (instead of matching factors), would give totally different results. In particular, because the interaction terms on parental education have a negative coefficient, the random matching counterfactual would actually shift the grade distribution *to the right*.

Figure 2: Class grade distribution: observed, fitted, and counterfactuals (Sample 2)



(a) Italian



(b) Math

Notes. Grades are normalized at the class level by taking the distance from the class median, calculated exactly from class registers. The resulting normalized grades range from -4 to 3. Dashed black lines represent the probability distribution of observed grades. Solid black lines represent the grade distribution as predicted by the fitted ordered probit model. Gray lines represent the counterfactual grade distribution under random matching between parents. Red lines represent the counterfactual grade distribution if parents sorted on human capital only (i.e., on the second sorting dimension).

acterized by a high level of homogamy; not only do men and women sort based on demographic and socioeconomic traits such as age, BMI, height, and education, but they also look for partners that share similar health-related behavioral traits and risk attitudes. Our estimates are also insightful about the number and nature of the sorting dimensions that can rationalize the marital patterns observed in the data. We find that a relatively low number of sorting dimensions, three in our sample of couples, are sufficient to summarize an individual's attractiveness on marriage markets. While the first dimension of sorting mainly captures market segmentation across age cohorts, the second dimension describes sorting on human capital, so that educated and health-conscious women tend to marry men with similar traits. When we look at family outcomes, we find that children of parents with a high level of human capital perform better at school and are more likely to exhibit healthy habits, but this comes at a cost, as they are also more likely to report lower levels of subjective well-being and a worse relationship with their parents. In particular, we show the existence of a positive association of human capital complementarities with child grades, particularly in Italian, but a negative association with the quality of mother-child relationship. Finally, through a counterfactual experiment, we show that, if parents were matched randomly, child grades would be lower overall than what observed in the data, precisely due to the presence of the aforementioned complementarities. Conversely, if parents matched exclusively on the human capital factor, and thus disregarded the other matching dimensions, child grades would be higher. While our data does not enable us to establish a causal relationship between parental characteristics and child outcomes, these results are suggestive of a tight relationship between matching on marriage markets and parenting, through complementarities in the production of children's human capital.

A Variable Description

Table 10 contains the survey questions that were used to measure different behavioral traits. In Table 11, a complete list of all variables that were used in section 5 as "outcome" variables are reported.

Table 10: Matching variables

Matching variables		
Variable	Question	Possible answers
Chooses healthy snacks	"At the end of the experiment, we will give you a snack. Which one do your prefer?"	1=chocolate, 2=Parmesan bar, 3=banana
Smokes	"I smoke"	0=Never, 1=Sometimes, 2=Often
Does sports	"I do sports"	0=Never, 1=Sometimes, 2=Often
Wears sunscreen	"I wear sunscreen to avoid sunburns"	0=Never, 1=Sometimes, 2=Often
Washes hands	"I wash my hands before eating"	0=Never, 1=Sometimes, 2=Often
Worries about health	"I worry about my health"	0=Never, 1=Sometimes, 2=Often
Would go on safari	"I would go on a jungle safari"	0=Never, 1=Sometimes, 2=Often
Fears speed	"I am scared of mopeds riding fast"	0=Never, 1=Sometimes, 2=Often
Likes holidays in known places	"I like holidays in places I know because it is safer"	0=Never, 1=Sometimes, 2=Often
Would do extreme sports	"I would do extreme sports"	0=Never, 1=Sometimes, 2=Often
Careful when crossing	"I am very careful when crossing the street"	0=Never, 1=Sometimes, 2=Often

Notes. The table reports the translated text of the questions that were asked in the written questionnaire.

Table 11: Outcome variables

Outcome variables	
Variable	Description
Number of children*	Children report family composition, including number and sex of siblings.
Stays at home (wife)*	Dummy variable. Answers "housewife" when asked about profession.

Outcome variables	
Variable	Description
Hours outside (wife)*	Mothers are asked: "If you work, how many hours do you usually spend outside the home?" Possible answers: Does not work, $3-5$ hours, $6-8$ hours, >8 hours.
Share equal tasks*	Dummy variable. Answers "Both parents work and share household chores" when asked about gender roles in their family.
Height*	Measured by interviewers.
BMI	Measured by interviewers.
Overweight	BMI greater than 25.
Birth weight	Reported by mothers.
Failed the school year	From class register.
Italian grade	From class register. It ranges from 4 to 10. We use deviation from class median.
Math grade	From class register. It ranges from 4 to 10. We use deviation from class median.
Italian Invalsi grade	Standardized national test. It ranges from 0 to 300.
Math Invalsi grade	Standardized national test. It ranges from 0 to 300.
Patience*	Dummy variable. Chooses to wait one day to have two snacks instead of one. See Section 5.
Subjective well-being*	Children are asked: "How happy are you about your life?" Possible answers are: 1=Very sad, ... 5=Very happy.
Happy*	Children are asked: "Are you happy?" Possible answers are: 0=Never, 1=Sometimes, 2=Often.
Often tired*	Self-reported. Possible answers: 0=Never, 1=Sometimes, 2=Often.
Has own bedroom*	Dummy variable. Child does not share his/her bedroom with siblings/parents.
Gets along with mum*	Children are asked: "Do you get along with your mum?" Possible answers are: 0=Never, 1=Sometimes, 2=Often.
Gets along with dad*	Children are asked: "Do you get along with your dad?" Possible answers are: 0=Never, 1=Sometimes, 2=Often.
Likes chocolate*	Children are asked if they like chocolate. Possible answers range from 1=Not at all, to 5=Very much.
Likes Parmesan*	Children are asked if they like Parmesan. Possible answers range from 1=Not at all, to 5=Very much.
Likes bananas*	Children are asked if they like bananas. Possible answers range from 1=Not at all, to 5=Very much.

Outcome variables	
Variable	Description
Chooses healthy snacks*	Children are asked to choose a snack to eat after the interview. Possible choices are: 1=chocolate, 2=Parmesan bar, 3=banana.
Gets pocket money*	1=Children report receiving at least 5€per month, 0=Otherwise.
Pocket money*	Children's self-reported amount of pocket money per month. Trimmed at 450€.
Number of daily snacks*	Self-reported.
Drinks sugary drinks*	1=Reports drinking sodas or other sugary drinks when snacking, 0=Otherwise.
Time spent on tablets*	Self-reported. Possible answers: < 1 hour per day, $1-3$ hours per day, > 3 hours per day.
Time spent on smartphones*	Self-reported. Possible answers: same as "time spent on tablets".
Time spent on TV*	Self-reported. Possible answers: same as "time spent on tablets".
Time spent on screens*	Sum of time spent on tablets, smartphones, and TV.
Time spent with parents*	Self-reported. Possible answers: 0=Seldom, 1=Often, 2=Very often.
Eats vegetables*	Measures vegetables *and* fruit consumption. Self-reported. Possible answers: 0=Never, 1=Sometimes, 2=Often.
Does sports*	Children are asked the same questions about health behavior and risk attitudes as their parents. See Table 10 for details.
Smokes*	See Table 10.
Wears sunscreen*	See Table 10.
Washes hands*	See Table 10.
Worries about health*	See Table 10.
Would go on safari*	See Table 10.
Likes holidays in known places*	See Table 10.
Would do extreme sports*	See Table 10.
Careful when crossing*	See Table 10.
Altruism I*	Children are asked: "If your classmate is in trouble, do you try to help him/her?" Possible answers are: 0=Never, 1=Sometimes, 2=Often.
Altruism II*	Children are asked: "Would you give money to your classmate if he/she has no money to buy a snack?" Possible answers are: 0=Never, 1=Sometimes, 2=Often.

Notes. The table clarifies how outcome variables were measured. Some variables were built straight from the class register or direct measurement, others from the answers to the survey. To construct Figures 1a, 1b, and 1c, as well as Table 7, all variables marked with * were normalized by taking the residuals after regressing the raw variable on the child's age.

B Additional Tables

Table 12: Frequency of children by parents' survey participation

	Father participates		
Mother participates	No	Yes	**Total**
No	326	8	334
Yes	48	289	337
Total	374	297	671

Notes. Survey participation is coded as the parent reporting at least some basic information.

Table 13: Frequency of children by parents' presence at home

	Father is present		
Mother is present	No	Yes	**Total**
No	0	3	3
Yes	16	625	641
Total	16	628	644

Notes. A parent is present if he/she participates to the survey and/or is reported as living at home by the child.

Table 14: Frequency of children by school type

Type of school	No.
Elementary school	183
Middle school	34
High school	37
Total	254

Notes. Elementary school is for children aged between 6 and 10, middle school for children aged between 11 and 13, and high school for children aged between 14 and 18.

Table 15: Age and education by gender for parents residing in the provinces of Naples and Caserta

	Mean	St Dev	10th P	90th P
Mother's education	2.6	0.9	2.0	4.0
Mother's age	39.8	5.8	32.0	48.0
Father's education	2.5	0.8	2.0	4.0
Father's age	42.9	6.1	35.0	51.0

Notes. ISTAT Labor Force Survey 2017-2019 data. These data are maintained and distributed as part of the IPUMS International series by the Center (2020). The sample contains 906 individuals with a child in school age and resident in the provinces of Naples and Caserta. The table reports mean, standard deviation, 10th and 90th percentile of each variable. Education is coded as a four-category variable.

Table 16: Rank test for $\hat{A}$ (Sample 1)

H_0: $rk(A) = k$	$k = 1$	$k = 2$	$k = 3$
χ^2	70.69	17.36	2.24
df	9	4	1
P-value	0.00	0.00	0.13

Notes: Each column reports the statistic resulting from testing the null hypothesis that the rank of $\hat{A}$ is equal to k. We report the corresponding p-values. These tests lead us to conclude that sorting occurs on at least 3 orthogonal dimensions.

Table 17: Rank test for $\hat{A}$ (Sample 2)

H_0: $rk(A) = k$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
χ^2	551.04	226.04	126.56	60.54	24.40	11.02	4.25
df	147	120	95	72	51	32	15
P-value	0.00	0.00	0.02	0.83	1.00	1.00	1.00

Notes: Each column reports the statistic resulting from testing the null hypothesis that the rank of $\hat{A}$ is equal to k. We report the corresponding p-values. The tests show that we cannot reject the hypothesis that sorting occurs on three orthogonal dimensions at the 1% level.

Table 18: Estimates of the affinity matrix diagonal: comparison across samples

	Benchmark (1)	Firstborns (2)	Reweighted (3)	Segmented (4)	Benchmark (5)	Reweighted (6)	Segmented (7)
Education	**0.77**	**1.17**	**0.69**	**0.88**	**0.29**	**0.35**	**0.18**
	(0.11)	(0.26)	(0.12)	(0.12)	(0.04)	(0.04)	(0.04)
Age	**3.27**	**2.76**	**3.51**	**3.12**	**0.91**	**0.92**	**0.78**
	(0.33)	(0.49)	(0.37)	(0.31)	(0.06)	(0.06)	(0.06)
Height	**0.17**	**0.29**	**0.17**	**0.19**	**0.04**	**0.05**	**0.04**
	(0.07)	(0.14)	(0.07)	(0.07)	(0.03)	(0.03)	(0.03)
BMI	**0.28**	0.20	**0.28**	**0.27**	**0.11**	**0.10**	**0.09**
	(0.07)	(0.12)	(0.07)	(0.07)	(0.04)	(0.04)	(0.04)
Smokes					**0.13**	**0.13**	**0.14**
					(0.04)	(0.04)	(0.04)
Does sports					**0.05**	**0.05**	**0.05**
					(0.04)	(0.04)	(0.03)
Chooses healthy snacks					**0.26**	**0.25**	**0.24**
					(0.05)	(0.05)	(0.05)
Wears sunscreen					**0.06**	**0.08**	**0.06**
					(0.04)	(0.04)	(0.03)
Washes hands					**0.10**	**0.08**	**0.08**
					(0.04)	(0.04)	(0.04)
Worries about health					**0.07**	**0.06**	**0.07**
					(0.04)	(0.04)	(0.03)
Would go on safari					**0.15**	**0.14**	**0.14**
					(0.05)	(0.05)	(0.05)
Fears speed					**0.05**	**0.07**	**0.04**
					(0.03)	(0.03)	(0.03)
Likes holidays in known places					**0.15**	**0.12**	**0.14**
					(0.05)	(0.05)	(0.05)
Would do extreme sports					0.01	0.01	0.01
					(0.02)	(0.02)	(0.02)
Careful when crossing					**0.15**	**0.15**	**0.15**
					(0.04)	(0.04)	(0.04)
N	254	88	254	254	201	201	201

Notes. Standard errors in parentheses; they are obtained with 2,000 bootstrap replications. Boldfaced estimates are significant at the 5% level. Columns (1) and (5) correspond to the diagonal of our benchmark estimates found in Tables 3 and 5 respectively. In Column (2), we report the estimates obtained with a sample of firstborns. In Columns (3) and (6), we report the estimates obtained after reweighting each observation in our main samples by the inverse of the number of children in the household. In Columns (4) and (7), we report the estimates obtained with an alternative model where individuals can only mate locally (i.e. with parents of children attending the same school as theirs).

Table 19: Correlation of matching indices with household outcomes (Sample 2)

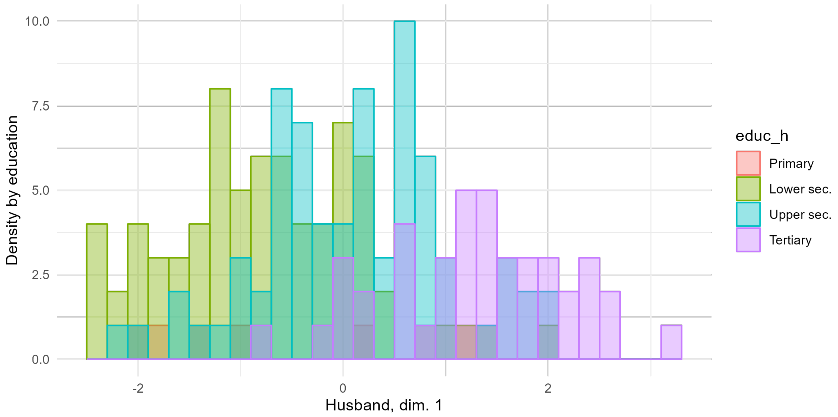
	Men			Women		
	Index 1	Index 2	Index 3	Index 1	Index 2	Index 3
Num. children	−0.01	0.02	−0.09	0.07	0.09	−0.01
	(0.91)	(0.79)	(0.20)	(0.30)	(0.20)	(0.88)
Hours outside (wife)	**0.25**	**0.30**	**0.15**	**0.31**	**0.33**	**0.19**
	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)	(0.01)
Stays at home (wife)	**−0.27**	**−0.40**	**−0.19**	**−0.30**	**−0.43**	**−0.25**
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Parents share tasks equally	0.03	**0.25**	0.03	0.03	**0.22**	0.11
	(0.63)	(0.00)	(0.66)	(0.71)	(0.00)	(0.11)
Height (child)	0.12	0.05	0.02	0.09	0.03	0.03
	(0.10)	(0.46)	(0.78)	(0.22)	(0.70)	(0.71)
BMI (child)	0.13	−0.10	−0.11	0.10	−0.07	−0.13
	(0.08)	(0.17)	(0.14)	(0.18)	(0.31)	(0.07)
Overweight	0.02	**−0.16**	**−0.16**	0.01	−0.11	**−0.15**
	(0.83)	(0.03)	(0.03)	(0.92)	(0.13)	(0.04)
Year failed	−0.03	0.03	−0.01	0.04	0.03	0.06
	(0.68)	(0.69)	(0.87)	(0.59)	(0.67)	(0.36)
Italian grade	0.03	**0.16**	**0.18**	0.02	0.11	0.12
	(0.67)	(0.02)	(0.01)	(0.77)	(0.11)	(0.10)
Math grade	**0.25**	**0.19**	**0.22**	**0.23**	**0.16**	**0.16**
	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)	(0.03)
Italian Invalsi	0.17	0.02	0.22	**0.30**	**0.30**	**0.35**
	(0.17)	(0.85)	(0.08)	(0.01)	(0.01)	(0.00)
Math Invalsi	**0.28**	0.08	0.16	**0.40**	**0.29**	**0.33**
	(0.02)	(0.50)	(0.20)	(0.00)	(0.02)	(0.01)
Patience (child)	−0.10	−0.02	0.09	−0.09	−0.09	0.04
	(0.14)	(0.78)	(0.22)	(0.19)	(0.23)	(0.55)
Sub. well-being (child)	−0.03	−0.03	0.04	−0.03	−0.04	−0.05
	(0.63)	(0.69)	(0.61)	(0.67)	(0.55)	(0.48)
Happy (child)	−0.11	**−0.22**	**0.14**	−0.14	**−0.19**	0.07
	(0.13)	(0.00)	(0.05)	(0.05)	(0.01)	(0.29)
Gets along with mom	−0.05	**−0.14**	−0.01	−0.07	**−0.17**	−0.02
	(0.48)	(0.05)	(0.87)	(0.30)	(0.02)	(0.76)
Gets along with dad	0.05	−0.07	0.04	0.03	−0.09	0.06
	(0.51)	(0.33)	(0.59)	(0.67)	(0.23)	(0.38)
Likes chocolate	−0.09	**−0.16**	0.05	−0.11	**−0.23**	0.01
	(0.21)	(0.03)	(0.49)	(0.14)	(0.00)	(0.86)
Likes Parmesan	0.02	**0.18**	0.03	0.05	0.11	0.06
	(0.73)	(0.01)	(0.64)	(0.48)	(0.13)	(0.37)
Likes bananas	−0.04	0.06	−0.03	−0.03	0.09	−0.00
	(0.61)	(0.38)	(0.67)	(0.67)	(0.23)	(0.98)
Chooses healthy snacks (child)	0.02	0.14	0.00	0.07	**0.18**	−0.02
	(0.76)	(0.05)	(0.95)	(0.31)	(0.01)	(0.75)
Gets pocket money (yes/no)	**−0.33**	**−0.16**	−0.12	**−0.32**	**−0.16**	**−0.15**
	(0.00)	(0.02)	(0.09)	(0.00)	(0.02)	(0.04)
Pocket money (monthly amount)	**−0.24**	−0.08	−0.02	**−0.22**	−0.02	−0.01
	(0.00)	(0.32)	(0.75)	(0.00)	(0.82)	(0.92)
Number of daily snacks	−0.08	−0.06	0.00	−0.07	−0.05	−0.04
	(0.25)	(0.38)	(1.00)	(0.33)	(0.45)	(0.61)

	Men			Women		
	Index 1	Index 2	Index 3	Index 1	Index 2	Index 3
Drinks sugary drinks	−0.03	0.01	0.02	−0.06	0.13	−0.06
	(0.70)	(0.87)	(0.76)	(0.42)	(0.06)	(0.42)
Time spent on tablets	0.03	0.09	0.02	0.01	0.09	−0.09
	(0.71)	(0.18)	(0.81)	(0.84)	(0.21)	(0.19)
Time spent on smartphones	**−0.16**	−0.08	−0.06	**−0.17**	−0.09	−0.12
	(0.03)	(0.23)	(0.39)	(0.02)	(0.23)	(0.10)
Time spent on TV	−0.09	−0.10	0.12	−0.09	−0.03	**0.16**
	(0.22)	(0.16)	(0.09)	(0.22)	(0.72)	(0.02)
Time spent on screens (total)	−0.10	−0.01	0.02	−0.12	0.01	−0.04
	(0.16)	(0.88)	(0.83)	(0.09)	(0.91)	(0.54)
Time spent with parents	**−0.15**	−0.10	0.13	−0.13	−0.09	0.03
	(0.03)	(0.15)	(0.06)	(0.06)	(0.23)	(0.68)
Eats vegetables	0.11	−0.02	0.11	0.10	0.04	−0.04
	(0.14)	(0.79)	(0.13)	(0.16)	(0.61)	(0.56)
Often tired	−0.03	−0.00	−0.08	−0.03	0.00	−0.04
	(0.71)	(0.99)	(0.29)	(0.71)	(0.96)	(0.61)
Has own bedroom	−0.04	−0.11	**−0.14**	−0.07	−0.02	**−0.22**
	(0.61)	(0.13)	(0.05)	(0.36)	(0.74)	(0.00)
Smokes (child)	0.03	0.09	−0.11	0.04	**0.14**	0.01
	(0.64)	(0.20)	(0.12)	(0.56)	(0.04)	(0.94)
Does sports (child)	−0.05	−0.00	0.11	−0.04	−0.04	0.08
	(0.49)	(0.99)	(0.14)	(0.57)	(0.60)	(0.23)
Worries about own health	−0.04	0.00	0.09	−0.04	0.02	0.02
	(0.60)	(0.97)	(0.22)	(0.62)	(0.79)	(0.79)
Wears sunscreen (child)	0.05	0.05	−0.01	0.01	−0.00	−0.06
	(0.48)	(0.53)	(0.86)	(0.88)	(0.96)	(0.43)
Washes hands (child)	0.03	−0.01	−0.01	0.02	0.01	0.08
	(0.62)	(0.88)	(0.92)	(0.74)	(0.94)	(0.29)
Would go on safari (child)	0.04	0.08	**0.17**	0.05	0.10	**0.17**
	(0.53)	(0.26)	(0.02)	(0.52)	(0.16)	(0.01)
Likes holidays in known places (child)	−0.07	−0.05	0.06	−0.07	−0.06	−0.06
	(0.34)	(0.48)	(0.42)	(0.36)	(0.36)	(0.36)
Would do extreme sports (child)	0.08	0.04	0.00	0.09	0.06	0.09
	(0.24)	(0.62)	(0.99)	(0.20)	(0.40)	(0.23)
Careful when crossing (child)	0.12	−0.00	−0.01	0.10	−0.06	0.02
	(0.09)	(0.97)	(0.93)	(0.18)	(0.43)	(0.76)
Altruism I	0.06	0.07	0.04	0.02	0.05	0.05
	(0.42)	(0.33)	(0.59)	(0.73)	(0.47)	(0.45)
Altruism II	0.07	0.09	0.11	0.06	0.03	0.04
	(0.33)	(0.21)	(0.14)	(0.42)	(0.62)	(0.56)
Age (child)	**0.56**	**0.19**	**0.22**	**0.62**	**0.23**	**0.21**
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)

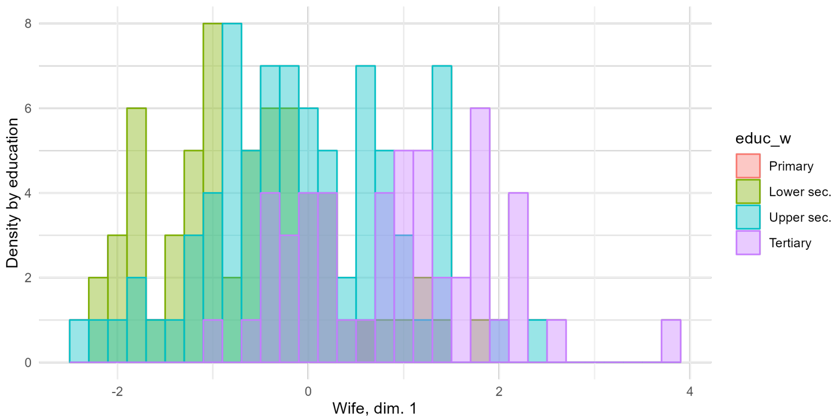
Notes. The table presents pairwise Pearson's correlation rate between matching indices and household outcomes. P-values of the two-tailed significance test in parentheses.

C Additional Figures

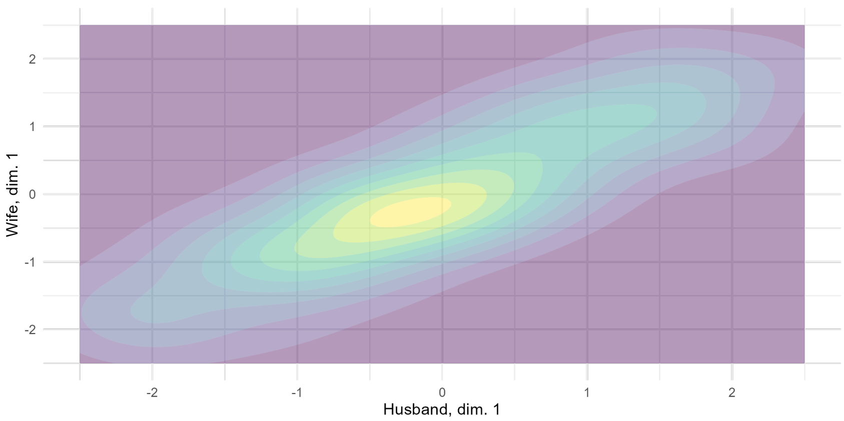
Figure 3: Distribution of Index 1 by education and gender (Sample 2)



(a) Men



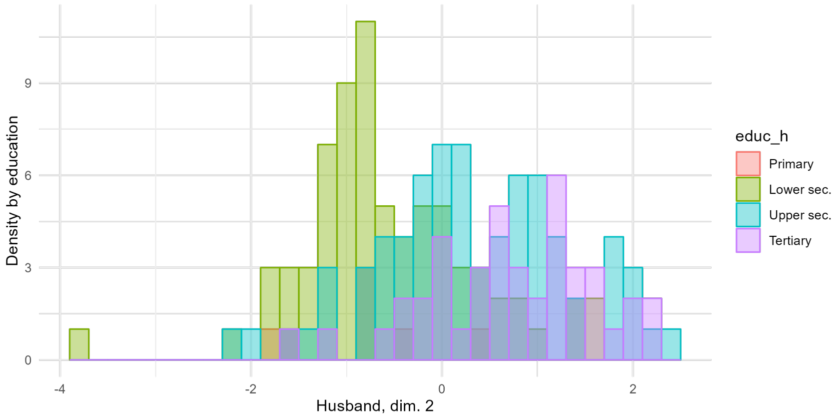
(b) Women



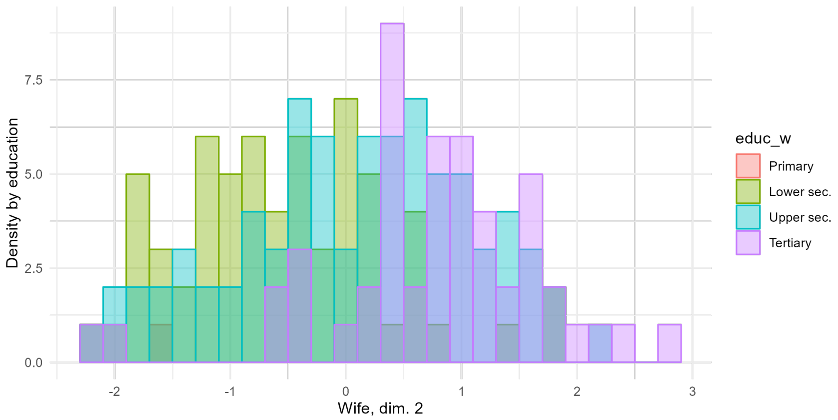
(c) Joint

Notes. We plot the distribution of the men's Index 1 by their education (a), the distribution of women's Index 1 by their education (b), and the joint distribution of men's Index 1 and women's Index 1 (c).

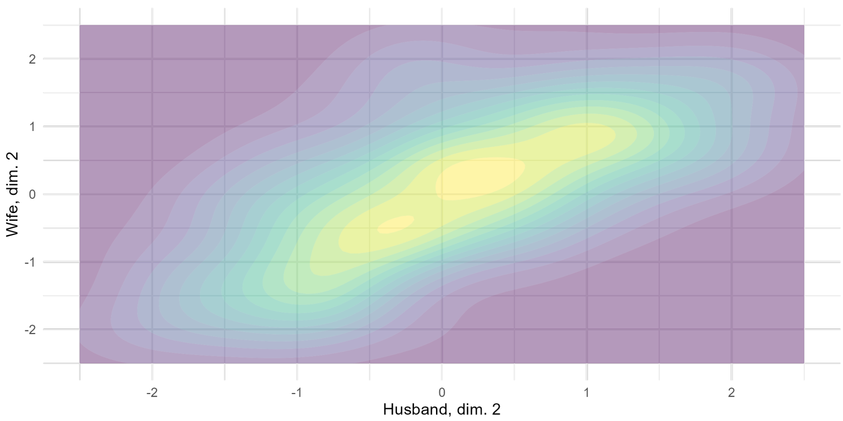
Figure 4: Distribution of Index 2 by education and gender (Sample 2)



(a) Men



(b) Women



(c) Joint

Notes. We plot the distribution of the men's Index 2 by their education (a), the distribution of women's Index 2 by their education (b), and the joint distribution of men's Index 2 and women's Index 2 (c).

References

Paul R. Amato and Alan Booth. Changes in gender role attitudes and perceived marital quality. *American Sociological Review*, 60(1):58–66, 1995.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

Christine Barnet-Verzat and François-Charles Wolff. Motives for pocket money allowance and family incentives. *Journal of Economic Psychology*, 23(3):339–366, 2002.

Gary S Becker. A theory of marriage: Part i. *Journal of Political Economy*, 81(4): 813–846, 1973.

Gary S Becker and Nigel Tomes. An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of political Economy*, 87(6):1153–1189, 1979.

Gary S Becker, Scott Duke Kominers, Kevin M Murphy, and Jörg L Spenkuch. A theory of intergenerational mobility. *Journal of Political Economy*, 126(S1): S7–S25, 2018.

Alberto Bisin and Thierry Verdier. "beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics*, 115(3):955–988, 2000.

Alberto Bisin, Giorgio Topa, and Thierry Verdier. Religious intermarriage and socialization in the united states. *Journal of political Economy*, 112(3):615–664, 2004.

Sandra E Black, Paul J Devereux, and Kjell G Salvanes. From the cradle to the labor market? the effect of birth weight on adult outcomes. *The Quarterly Journal of Economics*, 122(1):409–439, 2007.

Debra L Blackwell and Daniel T Lichter. Homogamy among dating, cohabiting, and married couples. *The Sociological Quarterly*, 45(4):719–737, 2004.

Milan Bouchet-Valat. Changes in educational, social class and social class of origin homogamy in france (1969-2011). *Revue Française de Sociologie*, 55(3):459–505, 2014.

Jonathan Bradshaw. Subjective well-being and social policy: Can nations make their children happier? *Child Indicators Research*, 8(1):227–241, 2015.

Minnesota Population Center. Integrated public use microdata series, international: Version 7.3 [dataset]. Technical report, 2020.

Pierre-André Chiappori, Robert J McCann, and Lars P Nesheim. Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Economic Theory*, 42(2):317–354, 2010.

Pierre-André Chiappori, Bernard Salanié, and Yoram Weiss. Partner choice, investment in children, and the marital college premium. *American Economic Review*, 107(8):2109–67, 2017.

Pierre-André Chiappori, Monica Costa Dias, and Costas Meghir. The marriage market, labor supply, and education choice. *Journal of Political Economy*, 126 (S1):S26–S72, 2018.

Pierre-André Chiappori, Monica Costa Dias, and Costas Meghir. Changes in assortative matching: Theory and evidence for the us. Technical report, National Bureau of Economic Research, 2020.

Eugene Choo and Aloysius Siow. Who marries whom and why. *Journal of Political Economy*, 114(1):175–201, 2006.

Daniela Del Boca, Christopher Flinn, and Matthew Wiswall. Household choices and child development. *Review of Economic Studies*, 81(1):137–185, 2014.

Matthias Doepke and Fabrizio Zilibotti. Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica*, 85(5):1331–1371, 2017.

Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. The intergenerational transmission of risk and trust attitudes. *The Review of Economic Studies*, 79(2):645–677, 2012.

Arnaud Dupuy and Alfred Galichon. Personality traits and the marriage market. *Journal of Political Economy*, 122(6):1271–1319, 2014.

Arnaud Dupuy, Alfred Galichon, and Yifei Sun. Estimating matching affinity matrices under low-rank constraints. *Information and Inference: A Journal of the IMA*, 8(4):677–689, 2019.

Steven N Durlauf, Andros Kourtellos, and Chih Ming Tan. The great gatsby curve. *Annual Review of Economics*, 14:571–605, 2022.

Raquel Fernández and Richard Rogerson. Sorting and long-run inequality. *Quarterly Journal of Economics*, 116(4):1305–1341, 2001.

George-Levi Gayle, Limor Golan, and Mehmet A Soytas. What accounts for the racial gap in time allocation and intergenerational transmission of human capital? 2013.

Pilar Gonalons-Pons and Christine R Schwartz. Trends in economic homogamy: changes in assortative mating or the division of labor in marriage? *Demography*, 54(3):985–1005, 2017.

Mònica González-Carrasco, Sara Malo, Ferran Casas, Gemma Crous, Mireia Baena, and Dolors Navarro. The assessment of subjective well-being in young children: Strengths and challenges. *Journal of Social Research & Policy*, 6(2), 2015.

Jeremy Greenwood, Nezih Guner, and John A Knowles. More on marriage, fertility, and the distribution of income. *International Economic Review*, 44(3):827–862, 2003.

Jeremy Greenwood, Nezih Guner, Georgi Kocharkov, and Cezar Santos. Marry your like: Assortative mating and income inequality. *American Economic Review*, 104 (5):348–53, 2014.

Carla Guerriero, John Cairns, Fabrizio Bianchi, and Liliana Cori. Are children rational decision makers when they are asked to value their own health? a contingent valuation study conducted with children and their parents. *Health Economics*, 27 (2):e55–e68, 2018.

Sara A Hart, Colleen M Ganley, and David J Purpura. Understanding the home math environment and its role in predicting parent report of children's math skills. *PloS one*, 11(12):e0168227, 2016.

Carolyn J Heinrich. Parents' employment and children's wellbeing. *The Future of Children*, 24(1):121–146, 2014.

Sarah E Jackson, Andrew Steptoe, and Jane Wardle. The influence of partner's behavior on health behavior change: the english longitudinal study of ageing. *JAMA internal medicine*, 175(3):385–392, 2015.

Gayle Kaufman. Do gender role attitudes matter? family formation and dissolution among traditional and egalitarian men and women. *Journal of Family Issues*, 21 (1):128–144, 2000.

Martha Henn McCormick. The effectiveness of youth financial education: A review of the literature. *Journal of Financial Counseling and Planning Volume*, 20(1): 71, 2009.

Luis Milan and Joe Whittaker. Application of the parametric bootstrap to models that incorporate a singular value decomposition. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44(1):31–49, 1995.

Walter Mischel, Yuichi Shoda, and Monica I Rodriguez. Delay of gratification in children. *Science*, 244(4907):933–938, 1989.

AA Nima, KM Cloninger, BN Persson, S Sikström, and D Garcia. Validation of subjective well-being measures using item response theory. *Frontiers in psychology*, 10:3036, 2020.

Enrica Pizzi, Michele Antonio Salvatore, Serena Donati, Silvia Andreozzi, Serena Battilomo, and Privitera Maria Grazia. Sorveglianza bambini 0-2 anni, sintesi dei risultati, edizione 2022. Technical report, Istituto Superiore di Sanità, 2022.

Zhenchao Qian and Samuel H Preston. Changes in american marriage, 1972 to 1987: Availability and forces of attraction by age and education. *American Sociological Review*, 58(4):482–495, 1993.

Robert Schoen. The harmonic mean as the basis of a realistic two-sex marriage model. *Demography*, 18(2):201–216, 1981.

Christine R Schwartz and Hongyun Han. The reversal of the gender gap in education and trends in marital dissolution. *American Sociological Review*, 79(4):605–629, 2014.

Christine R Schwartz and Robert D Mare. Trends in educational assortative marriage from 1940 to 2003. *Demography*, 42(4):621–646, 2005.

Robert S Siegler, Jenny Saffran, Nancy Eisenberg, and Elizabeth" Gershoff. *How Children Develop*. Worth Publishers; 6th ed. edition, 2019.

Mary Suiter and B Meszaros. Teaching about saving and investing in the elementary and middle school grades. *Social Education*, 69(2):92–95, 2005.

Tyler W Watts, Greg J Duncan, and Haonan Quan. Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological science*, 29(7):1159–1177, 2018.

Elke U Weber, Ann-Renee Blais, and Nancy E Betz. A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15(4):263–290, 2002.