Propensity Score Estimates of the Effect of Fertility on Marital Dissolution

Daniela Vuri*†

European University Institute

July 24, 2001

Abstract

In recent years many studies have reported significant empirical associations between fertility and marital dissolution. Whether this is a causal effect or only a correlation is not clear. This issue is explored by using matching methods. First the effect of "having children" (binary treatment) on marital disruption is investigated. Then, the method is extended to the case of "number of children in the household" (multi-valued treatment). The main findings indicate that parents do not divorce less in the presence of children but they only postpone the decision to divorce until children get older.

Keywords: Fertility; Marital dissolution; Propensity score methods; Counterfactual

Jel Classification: C12, C2, J12, J13

1 Introduction

In recent years many studies have reported significant empirical associations between fertility and marital dissolution. Whether this is a causal effect or only a correlation is not clear. The goal of this empirical analysis is to find out whether there is a true causal effect of fertility on marital dissolution. If fertility were randomly assigned to the population of married couples,

*Department of Economics, European University Institute, Via dei Roccettini 9, I-50016 San Domenico di Fiesole (FI), Italy. E-mail: vuri@iue.it. Fax: +39-055-4685202. Tel: +39-328-8524310

†I would like to thank Andrea Ichino for comments and guidance. Many thanks to participants at the Spring Meeting of Young Economists 2001 in Copenhagen, participants at the ESPE 2001 Conference in Athens, participants at the BHPS 2001 Conference in Colchester, and to seminar participants at the EUI, especially Anna Sanz de Galdeano, with whom I programmed the propensity scores methods, and Sascha Becker. A special thank to Erich Battistin who gave me the initial hints to program the propensity score methods and very useful comments. All errors are mine.

then the observed marital dissolution differential by fertility could be interpreted as a causal effect. However, as shown in Vuri (2001), the marriage continuation probability increases as the number of children increases, but at the same time the *potential* stability of the marriage may affect the arrival of children. Therefore, fertility may not be exogenous to the decision of marital dissolution.

There is a substantial body of literature studying the effect of fertility on marital dissolution and it can be divided into two categories based on the methodological approach used: studies considering fertility as an exogenous variable, and studies addressing the problem of endogeneity of fertility.

Studies in the first category show that children increase the stability of their parents' marriage throughout their preschool years, while children born before marriage increase significantly the chances that the couple will dissolve (Becker, Landes and Michael 1977, White and Lillard 1991, Peter 1986, Ono 1998). The positive effect of having children on marital stability does not seem to hold in the case of high number of children (Thornton, 1977).

However, this literature is not satisfactory because it neglects the potential problem of endogeneity of fertility. This implies that if fertility is not an exogenous variable in the divorce equation, all these studies provide biased estimates of the effect of fertility on marital dissolution. There are two potential sources of bias. First, the couples might differ systematically in their observable characteristics by fertility, i.e. if characteristics like religion, age and earnings differ between couples with children and childless couples, this might explain the observed marital dissolution differential by fertility. The second source of bias might be due to unobservable factors that affect both fertility and marital instability, in which case at least part of the observed relationship between them is spurious. The existence of any of the two biases would imply that households with children would behave differently from households with no children, independently of any true causal effect of fertility on divorce (selection bias problem).

Studies in the second category acknowledge the problem of endogeneity of fertility but their analysis is not always convincing. For instance, Becker et al. (1977) initially suggest the use of a simultaneous equations model to identify the causation between children and dissolution, but then they decide against this strategy by constructing a situation (they select women aged 40-55 whose fertility is already completed) that excludes causation running from marital (in)stability to fertility. Koo and Janowitz (1983) formulate, for married couples, a simultaneous model of fertility and marital dissolution, but then they estimate the two equations individually by single equation logit method, ignoring the issue of simultaneity previously addressed. Lillard and White (1993) use instrumental variables techniques to identify the simultaneous model of marital separation and fertility. However, the instrument chosen to identify the separation equation - the legal environment for divorce in the state of current residence - is weak because it need not affect separation, which is the outcome of interest- but only legal separation and divorce. Finally, Brien, Lillard and Stern (1999) propose a way to model endogenous investment in children in a model of cohabitation, marriage and divorce; unfortunately, they revert to exogenous investment in children as an element of the cost of divorce in their estimation because of computational costs.

This paper takes into account the problem of endogeneity of fertility but it uses the framework known as the *potential outcome approach* to identify and estimate the effect of interest. In particular, the relationship between fertility and marital dissolution is formulated in a treatment-outcome framework similar to an experiment where the treatment is randomly assigned. The treatment of interest (fertility in this case) is defined in terms of potential marital outcomes for the couples with children (treated); in particular the following question is explored: what would have been the marital outcome of a couple with children had they not had children? In this paper, I draw on matching methods developed in the statistcs literature (Rubin 1977, 1979, Rosenbaum and Rubin 1983, Heckman et al. 1989, 1997, 1998) that exploit full information contained in observable covariates. The matching method provides a way to estimate treatment effects when controlled randomization is not possible and there are no convincing natural experiments which could substitute randomization. The main purpose of the method is to identify a systematic way to construct a correct sample counterpart for the missing information on the treated outcomes had they not been treated and to pair units in the two groups.

The main findings indicate that parents do not divorce less in the presence of children but they only postpone the decision to divorce until children get older. Furthermore, not only the presence of children has an effect on marital dissolution but also the number of children in different age groups matters.

The remainder of the paper is organized as follows: Section 2 briefly summarizes the decision model of fertility and divorce presented in Vuri (2001). Section 3 introduces the potential outcome approach and it identifies the treatment effect under the causal effect model. Sections 4 describes the matching approach for the binary case (having children or not). Section 5 extends the methodology to the multi-valued treatment case (number of children). Section 6 describes the data sets used (the German Socio-Economic Panel-GSOEP, the British Household Panel Survey-BHPS, and the Panel Study of Income Dynamics-PSID), and the process of sample selection. In section 7, the results are presented. Finally, section 8 presents some concluding remarks and direction for further research.

2 The theoretical framework

A useful tool for examining the relationship between fertility and divorce is Becker's analysis of marriage (1974), according to which marriages and cohabitations are seen as voluntary arrangements between two adults, formed to coordinate consumption and production activities, including the conception of children. When couples marry, they begin to acquire various "things" together, including a dwelling and its furnishings, shared interest to friends and so on. These are defined as "general" investments because they retain their value regardless of the couple's marital status. However, there are also other types of investments made by the couples called "marital-specific" because they belong to the couple rather than to either one of the partners separately (e.g. information on the partner's preferences, a division of labor inside and outside the household, sexual affinity and children). One immediate implication of this distinction between marital investments is the way they affect a couple's divorce probability because the accumulation of marital-specific capital raises the expected gain from remaining married and consequently discourages dissolution (Becker et al. 1977).

This is particularly true for children because they represent the most important marital-specific investment of a couple during their marriage. Therefore, parenthood provides an important basis for marital stability and children greatly lower the risk of marital disruption (see Becker et al. 1977, Cherlin 1977, Becker 1991, Morgan and Rindfuss 1985). The presence of children may not only make the marriage more stable but it may also delay divorce[1] by increasing the gains from marriage and making it more costly than continuation in the marriage for two reasons: i) because of the anticipated complications attending a divorce action, such as problems with child custody, visitation plans, coparenting and single-parent problems; ii) because of the increasing awareness of the financial and psychological costs of divorce for children. Consequently, children appear to constitute financial, legal, and emotional[2] barriers to divorce.

However, causation also runs in the other direction, i.e. the arrival of children may be affected by the potential stability of the parents' marriage. In fact, a couple's "divorce inclination" may influence their decision to begin a family and their willingness to add children to an existing family. Therefore, couples who face a relatively high likelihood that they will not stay together may delay the decision to have children, because of the higher costs of ending a marriage with children with respect to one without (Weiss and Willis, 1985).

In Vuri (2001) a simultaneous equations model of marriage status and fertility decisions which

[1] This paper focuses only on the couple's first marriage . Dissolution, if any, is measured as of the date husband and wife started to live separately, regardless of whether the legal formality of a divorce decree took place subsequently.

[2] I am referring in this case to a sort of "stigma" which is sometimes attached to persons who divorce when they have children, especially very young, which might discourage couples from divorcing.

considers both directions of causality is presented. In particular, in the context of a model of marital-specific investment, it is shown that the marriage continuation probability increases as the number of children increases, and that the number of children is increasing in an unobservable measure of the quality of the marriage, which in turn influences the perceived marriage duration. This framework leads (with some simplifications) to a simple estimable model described by the following two equations (for the complete derivation of the model see Vuri (2001, sections 4 and 5):

$$D_i = \beta C_i + \gamma \mathbf{X}_i + \epsilon_i \tag{1}$$

$$C_i = \delta \mathbf{X}_i + \nu_i \tag{2}$$

where D_i is the binary variable identifying whether the couple is observed to divorce (D_i =1) or to stay married (D_i =0); C_i represents an indicator for having children or not;[3] $\mathbf{X}_i$ represents couple's demographic and social characteristics.

Equation 1 says that the decision to divorce is influenced by children in the household, by some observable characteristics $\mathbf{X}_i$ and by some unobservable factors ϵ_i. Equation 2 models the decision of a couple to have children, which depends on some observed characteristics $\mathbf{X}_i$ and some unobserved factors ν_i.

If fertility is exogenous to the divorce decision, then ordinary least square regression of the effect of fertility on marital dissolution yields an unbiased estimate of the treatment effect β in equation 1. However, fertility might be endogenous to the divorce decision if there is dependence between fertility C_i and the error term of the "divorce" relationship ϵ_i. The correlation between C_i and ϵ_i can arise for one of two not necessarily mutually exclusive reasons: (a) dependence between ϵ_i and ν_i, or (b) dependence between $\mathbf{X}_i$ and ϵ_i. The first case is referred as *selection on unobservables* (Heckman and Robb, 1985) and the second case as *selection on observables* (Rosenbaum and Rubin, 1983).

The methodology followed in this paper pursues the selection on observables approach and does not extend to selection on unobservables. In what follows, the framework of the potential outcome approach to causality is described.

3 The potential-outcome approach

Using the terminology of the evaluation literature, let C_i denote a binary variable indicating treatment status "having children or not" ($C_i \in \{0,1\}$); furthermore, let $D_i\,(1)$ denote the

[3] In section 5, the analysis is extended to the multivalued treatment "number of children in the household".

potential marital outcome of a couple i under the treatment state "having children" ($C_i = 1$), and $D_i(0)$ the potential marital outcome if the same couple i receives no treatment "having no children" ($C_i = 0$). Thus, $D_i = C_i D_i(1) + (1 - C_i) D_i(0)$ is the observed marital outcome for a couple i. The individual treatment effect is $\beta_i = D_i(1) - D_i(0)$, which, however, is not observable since either $D_i(1)$ or $D_i(0)$ is missing. Alternatively, one might focus on the average effect of treatment on the treated couples (ATT henceforth):

$$\widehat{\beta}_{|C_i=1} = E(\beta_i \mid C_i = 1) = E[D_i(1) | C_i = 1] - E[D_i(0) | C_i = 1] \tag{3}$$

which implies comparing the marital outcome of a couple with children to the counterfactual case with no children, i.e. what would have been the marital outcome of a couple with children had not they had children. It is thus necessary that each couple is potentially exposable to any of the two treatments.[4]

While the first expectation $E[D_i(1) | C_i = 1]$ can be identified in the subsample of the treatment group, the counterfactual expectation $E[D_i(0) | C_i = 1]$ is not identifiable without invoking further assumptions. To overcome this problem, one has to rely on the untreated couples $(D_i(0))$ of the comparison group to obtain information on the counterfactual outcome of the treated in the no-treatment status. The replacement of $E[D_i(0) | C_i = 1]$ with $E[D_i(0) | C_i = 0]$ does not seem the right strategy since treated and untreated couples tend to differ in their characteristics that determine the outcome if they themselves select into treatment.

An ideal randomized experiment would solve this problem because random assignment of couple to the treatment ensures that potential outcomes are independent of treatment status.[5] Hence, the treatment effect could consistently be estimated by the difference between the observed means of the outcome variable in the treatment group and in the no-treatment group. However, in this non-experimental setting, the choice of fertility is not likely to be random: fertility decision of a couple may depend on some (un)observed characteristics which could also influence its marital outcome. For example, being catholic could affect both the decision to have children and at the same time discourage marital separation; likewise, parents who are less committed to their families may be more likely to divorce and less likely to childbearing.

In this case, when randomized experiments are not available, other estimators have to be devised, relying on appropriate identifying assumptions.

[4] Note that already at this stage the stable unit-treatment value assumption (SUTVA) has to be made. In our case, it requires that the marital outcome of a couple depends only on its own treatment status, not on the treatment status of other couples in the population, and that whether couples have children or not does not depend on the fertility decisions of others (no peers effect).

[5] Randomization implies that: $C_i \perp (D_i(0), D_i(1))$ and therefore: $E[D_i(0) | C_i = 1] = E[D_i(0) | C_i = 0] = E[D_i | C_i = 0]$

In what follows, the approach used to construct a suitable comparison group, namely the matching method, and the identifying assumptions on which it is based, namely CIA, are described for the binary treatment case. In section 5 the approach is extended to the multivalued treatment case.

4 The matching approach

4.1 The Conditional Independence Assumption (CIA)

One approach to construct a correct sample counterpart for the missing information on the treated couples had they not been treated is based on statistical matching. Matching estimators try to re-establish the condition of an experiment when no such data is available by stratifying the sample of treated and untreated couples with respect to covariates X_i that rule both the selection into treatment and the outcome under study. Such a stratification eliminates selection bias provided all variables X_i are observed and balanced between treated and control group. In this case, each stratum (or cell) would represent a separate small randomized experiment and simple differences between treated and controls outcomes would provide an unbiased estimates of the treatment effect.

The matching method relies on the assumption that the relevant differences between any two couples, in terms of potential outcomes, are captured in their observable attributes. This underlying identifying assumption, called "conditional independence assumption"[6] (CIA henceforth) requires that, conditional on observed attributes X_i, the distribution of the counterfactual outcome $D_i(0)$ in the treated group is the same as the (observed) distribution of $D_i(0)$ in the non-treated group.[7] In other words, the outcomes of the non-treated are independent on the participation into treatment C_i, once one controls for the observable variables X_i. In symbols:

$$D_i(0) \perp C_i | X_i \tag{4}$$

It implies that, given X_i, the non-treated outcomes are what the treated outcomes would have been had they not been treated. This rules out the possibility that variables other than

[6] Also "unconfoundedness", or "ignorable treatment assignment".

[7] This is actually the weaker version of CIA. The strong version (Rosenbaum and Rubin, 1983) asserts that the assignment to treatment C_i is unrelated to the pair of potential outcomes $(D_i(1), D_i(0))$, within subpopulations homogeneous in X_i. Formally:

$$C_i \perp (D_i(0), D_i(1)) \mid X_i$$

However, since our objective is only the construction of the counterfactual $E(D_i(0) \mid X_i, C_i = 1)$ in equation 3, the weaker version of the CIA sufficies to identify the ATT.

X_i, on which the analyst cannot condition, affect both $D_i(0)$ and C_i, i.e. there is no *selection on unobservables.* Moreover, assume that $Pr(C_i = 0 \mid X_i = x) > 0$ for all x which guarantees that, with positive probability, there are untreated couples for each x.[8] From the previous two assumptions, it follows that $E(D_i(0) \mid X_i, C_i = 1) = E(D_i(0) \mid X_i, C_i = 0)$ (see Rosenbaum and Rubin 1983). The conditional mean response of the *treated* under no treatment for a given X can thus be estimated by the conditional mean response of the untreated under no treatment (the technique is simply to replace the unobserved outcomes of the treated had they not been treated with the outcome of non-participants with the same X_i characteristics, since they are statistically equivalent). In other words, the matched non-treated couples are used to measure how treated would have behaved, on average, had they not been treated.[9]

However, the CIA is controversial because it is based on the assumption that the conditioning variables available to the econometricians are sufficiently rich to justify application of matching. In particular, the CIA requires that the set of the X_i's should contain all the variables that jointly influence the outcome with no-treatment as well as the selection into the treatment.[10] To justify the assumption, econometricians implicitly make conjectures about what variables enter in the decision set of couples, and how unobserved (by the analysts) relevant variables are related to observables.

4.2 The average treatment effect for the treated

Under the CIA, the average effect of treatment on the treated can be computed as follows:

$$\begin{aligned}
\widehat{\beta}_{|C_i=1} &\equiv E\left[D_i(1)\,|C_i=1\right] - E\left[D_i(0)\,|C_i=1\right] = \\
& E_X\left\{E\left(D_i(1) \mid X_i, C_i=1\right) - E\left(D_i(0) \mid X_i, C_i=1\right) \mid C_i=1\right\} = \\
& \overset{CIA}{=} E_X\left\{E\left(D_i(1) \mid X_i, C_i=1\right) - E\left(D_i(0) \mid X_i, C_i=0\right) \mid C_i=1\right\} = \\
& E_X\left\{E\left(D_i \mid X_i, C_i=1\right) - E\left(D_i \mid X_i, C_i=0\right) \mid C_i=1\right\}
\end{aligned} \tag{5}$$

The ATT is estimated by taking the difference of the outcomes in the two groups conditional on covariates and then averaging over the distribution of observable variables in the treated population $X_i \mid C_i = 1$.[11] Practically, equation 5 is equivalent to stratifying the data into

[8] This implies to match couples only over the common support region of X_i where the treated and non-treated group overlap. Consequently, the ATT will be computed only for those treated couples falling within the common support. The drawback of this selection is that if the treatment effect is heterogeneous across couples, restricting the sample of treated to the common support can change the parameter estimated.

[9] Note that under the "conditional assumption", it is not necessary to make assumptions about specific functional forms of outcome equations, decision process or distribution of unobservables.

[10] In fact, it is called a "data hungry" identification strategy (Heckman *et al.*, 1998).

[11] The regression equivalent of this procedure requires the inclusion of all the possible interactions between the

cells defined by each particular value of X_i; then, within each cell (i.e. conditioning on X_i) the difference between the average outcomes of the treated and control couples is computed. Finally, these differences are averaged with respect to the distribution of X_i for the treated couples.

However, in a finite sample balancing X is problematic if the vector of observables is of high dimension. As the number of variables increases, the number of matching cells increases exponentially, and very often there will be cells containing either treated couples or control couples but not both, making the comparisons impossible.

Rubin (1977) and Rosenbaum and Rubin (1983) suggest to alternatively use the conditional probability to participate into the treatment $p(X_i) \equiv \Pr(C_i = 1 \mid X_i = x) = E(C_i \mid X_i)$, the *propensity score*, for purposes of stratifying the sample. They show that by definition treated and non-treated couples with the same value of the propensity score have the same distribution of the full vector of observables X_i. This is the so-called *balancing property* of the propensity score: $X_i \perp C_i \mid p(X_i)$. Furthermore, they demonstrate that if $D_i(0)$ is independent of C_i given X_i, $D_i(0)$ and C_i are also independent given $p(X_i)$. This implies that matching can be performed on $p(X_i)$ alone, thus reducing a potentially high dimensional matching problem to a one dimensional problem.

Matching treated and untreated couples with the same propensity scores and placing them into one cell means that the decision whether to participate or not is random in such a cell and the probability of participation in this cell equals the propensity score. Consequently the difference between the treatment and the non treatment average outcomes at any value of $p(X_i)$ is an unbiased estimate of the average treatment effect for the treated at that value of $p(X_i)$. Formally:

$$\widehat{\beta}_{|C_i=1} = E_{p(X)} \left\{ \left[E(D_i \mid C_i = 1, p(X_i)) - E(D_i \mid C_i = 0, p(X_i)) \right] \mid C_i = 1 \right\} \tag{6}$$

Therefore, an unbiased estimate of the ATT can be obtained conditioning on $p(X_i)$, which is equal to exact matching on the $p(X_i)$.

However, some drawbacks accompany this strategy. First, the propensity score itself has to be estimated (see the Appendix for a description of the algorithm used to estimate the propensity score). Second, since it is a continuous variable exact matches will rarely be achieved and a certain distance between treated and untreated couples has to be accepted. Several alternative and feasible procedures based on stratifying and matching (nearest and radius methods of matching) on the basis of the estimated propensity score have been proposed in the literature to solve this

observables X_i. The essential difference between regression and matching is the weighting scheme used to average estimates at different values of X.

problem(see the Appendix for a detailed description of the methods). In the next section, the extension of the matching methodology to the multivalued treatment case is described.

5 Estimation of the average causal effect with multi-valued treatment

Imbens (2000) and Lechner (1999) have proposed an extension of the propensity score methodology that allows for estimation of average causal effects with multi-valued treatments. The key insight of this method is that "for estimation of average causal effects it is not necessary to divide the population into subpopulations where causal comparisons are valid, as the propensity score does; it is sufficient to divide the population into subpopulations where average potential outcomes can be estimated" (Imbens, 2000: 706).

Here it is assumed that the treatment C_i can take values between 0 and K, i.e. $k = 0, 1, ..., K$. I am interested in average outcomes, $E\{D_i(k)\}$, for all values of k, and in particular in differences of the form $E\{D_i(k) - D_i(s)\}$, i.e. the average causal effect of exposing all units to treatment k rather than to treatment s. The key assumption, like in the binary case, is that adjusting for covariates solves the problem of drawing causal inferences. This is formalized by using a weak version of CIA in the multivalued treatment case. Let $T_i(k)$ be the indicator of receiving the treatment k:

$$\begin{aligned} \mathrm{T}_i(k) &= 1 \quad \text{if } \mathrm{C}_i = k \\ &= 0 \quad \text{otherwise} \end{aligned}$$

The weak version of CIA (also *weak unconfoundedness*) states that the assignment to treatment C_i is weakly unconfounded, given the covariates X_i if:

$$T_i(k) \perp D_i(k) \mid X_i \text{ for all } k = 0, ...K$$

Weak unconfoudedness requires only pairwise independence of the treatment with each of the potential outcomes. Furthermore, the independence of the potential outcome $D_i(k)$ with the treatment has to be only "local" at the treatment level of interest, i.e. with the indicator $T_i(k)$ rather than with the treatment level C_i. The important result is that under weak unconfoundedness the expected value of $D_i(k)$ can be estimated, by adjusting for X_i :

$$E\{D_i(k) \mid X_i\} = E\{D_i(k) \mid T_i(k) = 1, X_i\} = E\{D_i \mid C_i = k, X_i\}$$

Then, average outcomes can be estimated by averaging the conditional means: $E\{D_i(k)\} = E[E\{D_i(k) \mid X_i\}]$.

For the same motivations seen for the binary case, it can be difficult to estimate $E\{D_i(k)\}$ when the dimension of X_i is large. To solve the problem, Imbens proposes the multivalued version of the propensity score methodology. Firstly, he defines the "generalized" propensity score (GPS), which is the conditional probability of receiving a particular level of the treatment given the covariates. In symbols:

$$r(k, x_i) \equiv \Pr(C_i = k \mid X_i = x_i) = E[T_i(k) \mid X_i = x_i]$$

Note that as in the binary case, GPS satisfies by definition the balancing property, i.e. $T_i(k) \perp X_i \mid r(k, X_i)$ for all $k = 0, ...K$.

Then, by using the same argument as in the binary treatment case, he proves that the CIA given the generalized propensity score $r(k, X_i)$ holds:

$$T_i(k) \perp D_i(k) \mid r(k, X_i) \quad \text{for all } k = 0, ...K$$

Consequently, the average outcomes can be estimated by conditioning only on the generalized propensity score and the difference $E\{D_i(k) - D_i(s)\}$ can be easily computed for any k and s.

6 Data and measures

The empirical analysis of this paper is based on data from Germany, the UK and the USA.[12] The German data come from the *German Socio-Economic Panel* (GSOEP) in its 95% public-use version (see Haisken-De New and Frick, 1998). The British data come from the *British Household Panel Survey* (BHPS) (see Rose et al., 1991). The USA data come from the *Panel Study of Income Dynamics* (PSID) (see Martha S.Hill, 1992).

In order to build the final sample of analysis, I follow a simple procedure of sample selection: firstly, I select only the couples married for the first time in 1990 or before but still married in 1990 (1992 for the UK), then I follow these couples in the next five years and I identify whether

[12] For the German sample, I only make use of the West German and Foreigners subsamples for the waves 7-12 (1990-1995). This choice is due to the fact that the sample of immigrants has been collected only since 1994, while the East German subsample is excluded because the income variables are not comparable with those of the first two subsamples, at least for the two years after the German reunification in 1989. For the USA sample, I use the five waves 1990-1995. For the British sample, I make use of the five waves (1992-1997), in order to extract comparable datasets (with the German and the US samples) in terms of the number of years of analysis. However, for Germany and the USA, I access a simplified version of their panels, the CNEF 1980-1997, which contains equivalently defined variables for the PSID and for the GSOEP. Since the CNEF 1980-1997 can be merged with the original surveys, I incorporate these constructed variables into my current analyses.

they divorced or separated in one of these five years or are still married at the end of the period analyzed.[13] Couples in which one of the two spouses dies during the period and couples who marry after 1990 (1992 for the UK) are excluded from the sample. The dependent variable is the indicator for marital status (equal to one if the couple divorces or separates between 1990 and 1995 (1992 to 1997 for the UK), 0 if the couple is still married at the end of 1995 (1997 for the UK)). The covariates of interest are the indicator of the *Number of children between 0 and 18 years old, Number of children between 0 and 6 years old, Number of children between 7 and 18 years old,* and dummy variables for *Having children between 0 and 18 years old or not, Having children between 0 and 6, Having children between 6 and 18.*[14] Each record describes family characteristics, like yearly total household income and duration of marriage, and personal characteristics, like age, education, labor earnings, religious affiliation for both partners (being both Catholic, Protestant or atheist), and 3 country dummy variables, all recorded in 1990 (1992 for the UK). Furthermore, I select only couples where the wife is less than 45 years old and whose oldest child was less than 18 years of age at the time of the interview.[15]

After restricting the sample to households with complete records in the critical variables, **3351** records remain (**848** for Germany, **923** for the UK, and **1580** for the USA). This constitutes the pooled restricted sample of household observations in the three countries on which the estimation results are based (see table 1 for more details on the procedure of sample selection). Tables 2 shows descriptive statistics (mean and standard deviation) of the controls and the treatments of interest included in the regressions for the pooled sample. In addition, tables 2-4 show summary statistics of the covariates for the three groups of treated (couples with children aged 0-18, aged 0-6 and aged 6-18 respectively) and controls (couples with no children in the age group 0-18, 0-6 and 6-18), and t-statistics from the test of equality of means between them are reported. Overall, there are substantial differences between the two groups in the three cases under study in terms of most of the explanatory variables. The existence of such differences highlights the need for the careful statistical adjustment procedures described in the previous sections.

The variables selected in this paper are the same considered in the literature on this topic. In particular, previous studies have usually included husband-wife characteristics at the time of marriage (or at the time of interview) like education, age, marriage duration, earnings, previous

[13] The choice of a period of this length comes from the fact that the decision of divorce or separating usually takes a long time, particularly because of the length of legal procedures.

[14] The covariates of interest are recorded in 1990 ('92 for the UK); I exclude the couples who have an additional child(ren) during the five years of analysis in order to compute the effect of the children present in 1990 ('92) on the probability of being still married five years later.

[15] This sample selection allows me to exclude from the sample old couples with children who likely moved out from the household .

cohabitation, pre-marital births, and traits that for most individuals do not vary over time like religion, race etc. (see Lecher 2988, Becker et al. 1977, Brien, Lillard and Stern 1999, Lillard and Waite 1993, Ermish and Francesconi 1996).

For this reason, it can be reasonably assumed that these attributes contain relevant observable information influencing both the marital outcome and fertility decision. Therefore, the conditional independence assumption can be considered as valid for the remainder of this paper.[16]

7 Estimation results

In section 7.1, estimates of the effect of fertility on marital dissolution obtained through parametric methods are presented, namely OLS and probit, the latter taking into account the binary nature of the outcome variable. These estimates represent the benchmark for the comparison with the propensity score matching estimates presented in section 7.2. The analysis is performed both for the binary treatment and for the multivalued treatment.

7.1 OLS and probit estimates

Firstly, three different binary measures of fertility are considered, i.e."*having children between 0-18 years old*", "*having children between 0-6 years old*","*having children between 6-18 years old*" (rows 1-3). In column (1) of table 5, the OLS-estimates of the effects of having at least one child in each of the three groups on the probability of marital dissolution are reported, controlling for the vector of observed variables, indicated by X_i in equation (7) and listed in section 6.[17] The estimated coefficient on the fertility binary variable "*having children between 0-18 years old*" is positive and equal to 0.036, which implies that having children has a positive effect on the dissolution rate. This result seems to be at odds with previous empirical findings and theoretical considerations (see sections 1 and 2). To investigate this result further, I disentangle the treatment "*having children aged 0-18*" into two measures of fertility, i.e."*having young children aged 0-6*" and "*having older children aged 6-18*". The results show that the positive effect previously

[16] Of course, there may be substantial arguments claiming that this is not true. For example, if one believes that there are additional unobserved factors correlated with outcomes and selection into treatment, not captured when we condition on the observables, then, of course, this invalidates the CIA and the following analysis. Moreover, the presence of infertile couples could also invalidate the CIA, because any couple has to be potentially exposable to both the treatments. However, it can be reasonably assumed that the proportion of infertile people in the sample is negligible.

[17] The list of covariates X_i included in the regression is slightly different from the one listed in section 6 to avoid problems of collinearity; it includes the couple's marital duration, the couple's average age, the couple's average age to the square, the couples's average education, household total income (in log), wife labor earnings (in log), three country dummies and three dummy for being Catholic, Protestant and atheist.

estimated is mainly due to the effect of older children (aged 6-18) on dissolution (0.08 percentage points), while having young children (aged 0-6) reduces significantly the dissolution probability by 0.021 percentage points. The same result is obtained from probit estimation.[18]

In rows 4-6, the OLS and probit estimates for the multivalued case (when the treatments are "*number of children aged 0-18*", "*number of children aged 0-6*","*number of children aged 6-18*") are reported. These results are very similar to the ones of the binary case, but smaller in size. In particular, an additional child in each of the three groups of interest has an effect respectively of 1.9%, of -3.0% and of 3.3% on marital dissolution.

However, as already pointed out, OLS (or probit) estimates can be biased because of the self-selection problem and the potential correlation of fertility with some observable characteristics that make these estimates biased. Therefore, we turn to the propensity score matching estimators which provide unbiased estimates of the causal effect of fertility on divorce.

7.2 Results using the propensity scores

This section is organized in the following way: the first part focuses on the analysis of the binary treatment case, in which the estimation of the propensity score, and the results from the stratification and matching procedures are presented; the second part is devoted to the multi-valued treatment results.

7.2.1 The dichotomous treatment case

Estimating the propensity score The first step in the implementation of this methodology is to estimate the propensity score for the three treatments under study. In general, any standard probability model can be used to estimate the propensity score. For example, $\Pr\{C_i = 1 \mid X_i\} = F(h(X_i))$, where F(.) is the normal or the logistic cumulative distribution and $h(X_i)$ is a function of covariates with linear and higher order terms. The choice of which higher order terms to include is determined by the need of obtaining an estimate of the propensity score that satisfies the balancing property. In this paper,the propensity score for the three treatments of interest is estimated using a probit model and following a simple algorithm proposed by Dehejia and Wahba (1998), which is described in more details in the Statistical Appendix. Essentially, observations are grouped into blocks defined on the estimated propensity score and it is checked whether the score and the covariates are balanced across the treated and the controls within each stratum. Interaction and higher order terms are added and blocks are divided into finer blocks until this balance is achieved. In my case, I have started with five blocks based on the quintiles of the estimated propensity score for the treated, and then I have tested whether the

[18] Marginal effects of probit estimation are reported in all the tables in order to be comparable with the OLS estimates.

means of the score for the two groups are statistically different. I have built finer blocks until the test is satisfied for all of them and 9 blocks are identified for the treatments "*having children between 0-18 years old*" and "*having children between aged 6-18*", 10 blocks are identified for the treatment "*having children between 0-6 years old*". There has been no need to add interaction and higher order terms. Once the balance is achieved for the score, also the distributions of covariates X_i between the two groups should be identical for the balancing property. I provide an example of it by testing for equality of means between the treated and the control groups for each of the nine variables in X_i, within each block and for each treatment. In almost all cases I find equality of means of the X_i at the 5% confidence level, and none of the covariates does systematically fail the test in all the blocks. Remember that when the same test was performed on the whole sets of control and treated units, rather then within each stratum, I rejected equal means for twelve out of sixteen variables used in the regression for the treatment "children 0-18", eight out of sixteen variables for the treatment "children 0-6", and fourteen out of sixteen variables for the treatment "children 6-18" (see Tables 2-4). Figure 1 plots the histograms of the estimated propensity scores for the three treatments. Note that they do not include the controls whose estimated propensity score is less than the minimum estimated propensity score for the treated units. There is no need to discard control units at the top of the distribution because they are below the maximum value of the estimated propensity score for the treated in all the three cases. This selection is necessary in order to assure that the treated and the control units lie on a common support. The figure reveals that there is a discrete overlap in terms of the propensity score in each block, while in the extreme bins there is only a limited overlap, as expected, because the number of treated units increases and the number of control units decreases at high values of the propensity score. However, this does not generate bias in my estimates as long as the balancing property is satisfied; this ensures that the treated couples in each block are observationally identical to the controls in the same block and only by chance does the treatment status differ in the two groups.

The next step consists in estimating the ATT using equation 6, where the ATT is computed as the difference between the treatment and the control average outcomes at any value of $p(X_i)$. However, as mentioned in section 4, the exact matching on $p(X_i)$, implicit in this strategy, is unfeasible in practice because the probability of observing two couples with exactly the same value of the propensity score is in principle zero since $\mathrm{p}(X_i)$ is a continuous variable. There are, however, several methods proposed in the literature based on stratifying and matching (nearest and radius methods of matching) on the basis of the estimated propensity score (the two methods are described in details in the Statistical Appendix).Estimating the treatment effect

Blocking (stratification) estimator The stratification estimator relies on the same division into strata defined in section 7.2.1, where the covariates are balanced across treated and control couples by construction. Then, within each block, the difference between the average outcomes of the treated and the controls is computed. The ATT of interest is finally obtained as an average of the ATT of each block with weights given by the distribution of treated units across blocks.

The stratification estimates show a positive effect of fertility on divorce for the cases "*having children aged 0-18*" and "*having children aged 6-18*" (column 1 named unadjusted), the second effect being larger than the first one (see tables 6-8). In fact, having children aged 0-18 increases the probability of divorce by 2.0 percentage points, while having older children between 6 and 18 years old increases this probability by 3.0 percentage points. Instead, the effect of having young children aged 0-6 discourages marital dissolution by 0.3 percentage points (but it is not statistically significant).

An alternative is the linear and probit regression adjustments (column 2, named adjusted), which eliminate the remaining within-block differences in the covariates. The results are similar to the unadjusted estimate in column 1 providing further evidence that the covariates are well balanced.

Matching estimator In the matching method, each treated couple is matched with replacement to a control couple such that their propensity scores are close enough to be considered approximately the same. Matching with replacement means that more than one control unit can be matched to the same treated unit while the unmatched controls are discarded. The simplest method is to match each treated unit to the single control unit with the closest propensity score (*nearest-match method*). However, it is obvious that some of these matches are fairly poor. The *radius method* of matching offers a solution to this problem because it consists in matching each treated to the control couple(s) whose propensity score is within a $\delta-$radius chosen by the researcher.[19] In this way only higher quality matches are selected, even if it has the disadvantage of reducing the sample size.

In tables 6-8, the results of the matching estimators computed using both the *nearest matching method* and the *radius method* are reported for the three treatments of interest. For the last method, three different measures of the δ-radius are chosen in order to check the robustness of the estimates to this choice. Note that, in the δ-radius method, not only the more distant controls are discarded, like in the nearest matching method, but also the treated units for which a match within the δ chosen could not be found. Consequently the lower is the radius chosen,

[19] The choice of δ depends on the willingness of the researcher to select more accurate matches because lower δ implies selecting higher quality matches.

the smaller is the number of remaining units (column 3).

Like for stratification estimator, the treatment effect can be either estimated as a difference in means in marital outcomes across these pairs of treated and matched control units (unadjusted), or a linear regression of marital outcomes on covariates on the balanced sample can be performed (adjusted). The results show that controlling for covariates does not alter the estimates significantly.

In particular, the matching estimates provide evidence of a positive effect of *having children aged 0-18* and of *having children aged 6-18* on marital dissolution. The first effect ranges from 0.019 for the nearest method to 0.004 for radius method with $\delta = 0.0005$, the second effect ranges from 0.050 for the nearest method to 0.018 for radius method with $\delta = 0.0001$ (column 1). The effect of having children aged 0-6 on the contrary is negative and ranges from -0.016 for the nearest method to -0.001 for radius method with $\delta = 0.0005$ (column 1) Furthermore, they are also consistent with the stratification results.

In conclusion, matching methods and stratification method yield similar results to the parametric methods (OLS and Probit) reported in section 7.1, and they support the idea that having young children only delay parental's decision of divorce until children get older.

7.2.2 The multi-valued treatment case

Table 9 shows the estimates in the multi-valued treatment cases, respectively "*number of children aged 0-18*", "*number of children aged 0-6*" and "*number of children aged 6-18*". It is important to note that in this case not only the average effect of having an additional child on the probability of divorce can be computed (as for the OLS and Probit estimates, see table 3), but the effect of each additional child on marital disruption can be analyzed, i.e. the effect of going from 0 to 1 child, from 1 to 2 children, from 2 to 3, and finally from 3 to more than 3 children.[20] In order to compute the generalized propensity score an ordered probit of each of the three treatments of interest on the covariates X_i is computed.

The estimates show that there is a positive effect of the treatment "*number of children aged 0-18*" on divorce especially for the first two children (respectively 0.042 for going from zero to one child and 0.048 for going from one to two children), while the other effects seem to be of lower magnitude (0.026 for going from two to three children and 0.003 for going from three to four or more children). By computing a weighted average of these four effects, the evidence shows that having an additional child in the range 0-18 years old on average increases the probability of marital dissolution by 3.1 percentage point. The same pattern is observed for the treatment "*number of children aged 6-18*". Only the effect of going from one to two

[20] The other cases (couples with more than 4 children) were too rare to be considered separately; consequently, we have decided to group all the couples with more than three children.

children is of opposite sign, but it is not statistically significant. The average effect in this case is 7.0 percentage points. Finally, the estimates of the treatment "*number of children aged 0-6*" on marital dissolution are significantly negative and the larger effect is obtained by going from 0 to one child (-6.6 percentage points). The average effect shows that having an additional young child reduces the probability of divorce by 6.0 percentage points. Like for the binary treatment case, the results support the evidence of the OLS (and probit) analysis of Section 7.1 but they are larger in size. Furthermore, the estimates for the multivalued case are higher than the estimates for the binary case showing that not only the presence of children has an effect on marital dissolution but also the total number of children in each group matters.

8 Conclusion

In this paper, I shown how to estimate the treatment effect of fertility on marital dissolution in presence of non random assignment using propensity score methods. In particular, I have analyzed the effects of three binary treatments "*having children aged 0-18*", "*having children aged 0-6*", and "*having children aged 6-18*" on marital dissolution, by using stratification and matching techniques, and the effects of the multivalued treatments "*number of children aged 0-18*", "*number of children aged 0-6*", and "*number of children aged 6-18*" on marital dissolution. The empirical analysis strengthens the evidence that parents do not divorce less in the presence of children but they only postpone the decision to divorce until children get older; in addition, the results for the multivalued case support the evidence found for the binary case but they are larger showing that not only the presence of children has an effect on marital dissolution but also the total number of children in each group matters.

Two directions of researches are currently under study. First, the temporal dimension of the data is used and the propensity score methodology is applied to the panel data in the hope of controlling for some fixed unobservable factors that could invalidate the CIA. Second, the three countries are studied separately. Preliminary estimates show that Germany behaves differently from the UK and the USA in that children aged 0-18 have a negative effect on marital dissolution, whilst the opposite happens in the other two countries. My goal for the next future is to investigate in further details on these two points and in particular to identifies the institutional and cultural differences that could explain the different behaviors in the three countries.

References

Angrist J.D. and Evans W., (1998), "*Children and their parents' labor supply: evidence from exogenous variation in family size*", The American Economic Review 88(3): 450-477.

Angrist J.D., Imbens G.W and Rubin D.W., (1996), "*Identification of causal effects using instrumental variables*", Journal of the American Statistical Association, 91 (434), 444-472

Angrist J.D. and Krueger A.B., (1998), "*Empirical Strategies in Labor Economics*", Handbook of Labor Economics (2000).

Becker G.S., Landes E.M. and Michael R.T, (1977), "*An economic analysis of marital instability*", Journal of Political Economy 85(1): 1141-1187.

Becker G.S., (1991); A Treatise on the Family; Harvard University Press.

Becker G.S., (1974), "*A theory of marriage: Part II*", Journal of Political Economy 82(2): S11-S26.

Blundell R. and Costa Dias M. (2000), "*Evaluation Methods for Non-Experimental Data*", mimeo UCL

Brien M.J., Lillard L.A. and Stern S., (1999), "*Cohabitation, Marriage, and Divorce in a Model of Match Quality*", mimeo, University of Virginia.

Cherlin A., (1977), "*The effect of children on marital dissolution*", Demography 14: 265-72.

Dehejia R and Wahba S. (1998), "*Propensity Score Matching Methods for Non-Experimental Causal Studies*", NBER Working Paper n° 6829

Dehejia R and Wahba S. (1999), "*Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs*", JASA: 1053-1062

DIW (ed.), (1998), The German Socio-Economic Panel (GSOEP) Desktop Companion, Berlin.

Ermish J. and Francesconi M. (1996), "Pertnership Formation and Dissolution in Great Britain", WP 96-10 of the ESRC, Colchester, University of Essex.

Green W.H., (1987), Econometric Analysis, 3rd. edition, Prentice Hall.

Heckman J. and Hotz V.J. (1989), "*Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: the Case of Manpower Training*", JASA December: 862-880.

Heckman J., Ichimura H, and Todd P. (1998), "*Matching as an Econometric Evaluation Estimator*", Review of Economic Studies 65: 261-294

Heckman J., Ichimura H, and Todd P. (1997), "*Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme*", Review of Economic Studies 64: 605-654

Heckman J., Ichimura H, Smith J.A., and Todd P. (1998), "*Characterising Selection Bias Using Experimental Data*", Econometrica 66.

Hill Martha S., (1992), The panel study of income dynamics: a user's guide, Harvard University

Holland P.W., (1986), "*Statistics and causal inference*", Journal of the American Statistical Association, 81 945-960.

Koo H and Janowitz B., (1983), "*Interrelationship between fertility and marital dissolution: results of a simultaneous logit model*", Demography 20 (2): 129-145.

Imbens G., (2000), "*The Role of Propensity Score in Estimating Dose Response Functions*", Biometrika 87.

Imbens G. and Rubin D.B., (1997), "*Estimating Outcome Distributions for Compliers in Instrumental Variables Models*", Review of Economic Studies 64: 555-574.

Imbens G., Rubin D.B., and Sacerdote B, (1999), "*Estimating the Effect of Unearned Income on Labor Supply, Earnings, Savings and Consumption: Evidence from a Survey of Lottery Players*", NBER Technical Working Paper Series n°7001.

Lehrer E.L. (1988), "*Determinants of Marital Instability*", Applied Economics 20: 195-210.

Lechner M. (2001), "*Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies*", University of St Gallen, mimeo.

Lillard L.A. and Waite L.J. (1993), "*A Joint Model of Marital Childbearing and Marital Disruption*", Demography 30: 653-681.

Manski C.F., Sandefur G.D., McLanahan S. and Powers D., (1992), "*Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation*", JASA 87: 25-37.

Morgan S. and Rindfuss R., (1985), "*Marital Disruption: Structural and Temporal Dimensions*", American Journal of Sociology 90: 1055-1077.

Ono H., (1998), "*Husbands' and Wives' Resources and Marital Dissolution*", Journal of Marriage and the Family 60: 674-689.

Peters E., (1986), "*Marriage and Divorce: Informational Constraints and Private Contracting*", The American Economic Review 76: 437-454.

Persson T, Tabellini G., and Trebbi F, (2001), "*Electoral Rules and Corruption*", CEPR Discussion Paper Series n°2741

Rose D. and 14 others, (1991), "*Micro-social change in Britain: an outline of the role and objectives of the British Household Panel Survey*", Working papers of the ESRC Research Centre on Micro-social Change, Paper I, Colchester, University of Essex.

Rosenbaum P.R. and Rubin D.B., (1983), "*The central role of the propensity score in observational studies for causal effects*", Biometrika 70: 41-55.

———, (1985a), "*Reducing Bias in Observational Studies Using the Subclassification on the Propensity Scores*", Journal of the American Statistical Association, vol.79: 516-524.

———, (1985b), "*Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity*", American Statistician, vol.39(1): 33-38.

Rubin D., (1977), "*Assignment to a Tretment Group on the Basis of a Covariate*", Journal of Educational Statistics, vol.2: 1-26.

———, (1979), "*Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observation Studies*", Journal of the American Statistical Association, vol. 74: 318-328.

Sianesi B. (2001), "*An Evaluation of the Active Labor Market Programs in Sweden*", mimeo UCL

Thornton A, (1977), "*Children and marital stability*", Journal of Marriage and the Family 39: 531-540.

Vuri D. (2001), "*Fertility and Divorce*", European University Institute Working Paper n°5.

Waite L., Booth A. and Edwards J., (1986), "*Children and marital happiness: why the negative relationship?*", Journal of Family Issues 7: 131-148.

Waite L. and Lillard L., (1991), "*Children and marital disruption*", American Journal of Sociology 96: 930-53.

Waite L., Haggstrom G. and Kanouse D., (1985), "*Parenthood and marital stability*", American Sociological Review 50: 850-57.

Weiss, Y and Willis, J.R., (1985), "*Match Quality, New Information and Marital Dissolution*", Journal of Labor Economics 15: S293-S329.

9 Statistical Appendix

9.1 Estimating of the propensity score

Dehejia and Wahba (1998) propose the following algorithm:

1. start with a parsimonious probit (or logit) function with linear covariates to estimate the propensity score;

2. rank all observations by the estimated propensity score (from the lowest to the highest);

3. in order to match treated and controls over the common support of X_i, the control units with an estimated propensity score less than the minimum or greater than the maximum estimated propensity score for treated units are discarded;

4. split the sample in 5 blocks of equal score range;

5. within each block test that the average propensity scores of treated and control couples do not differ;

6. if the test fails in one interval, split the interval and test again. Continue until the test is satisfied in all the blocks;

7. within each block test the *balancing property,* i.e. the means of each covariate do not differ between treated and control units. If the means of one or more covariates differ, interactions and higher order terms can be added and blocks divided into finer blocks until the balance is achieved.

9.2 The blocking estimators

This method is based on the same stratification procedure used for estimating the propensity score, i.e. the strata are chosen so that the balancing property is satisfied. Then, within these blocks indexed by q, the average difference in marital status between the treatment and the control couples is computed:

$$\widehat{\beta}_q = \frac{\sum_{i \in I(q), C_i=1} D_i}{\sum_{i \in I(q)} C_i} - \frac{\sum_{i \in I(q), C_i=0} D_i}{\sum_{i \in I(q)} (1-C_i)} = \frac{\sum_{i \in I(q), C_i=1} D_i}{N_q^T} - \frac{\sum i \in I\left(q\right), C_i = 0 D_i}{N_q^C}$$

where N_q^T and N_q^C are the numbers of treated and controls in block q, and $I\left(q\right)$ is the indicator function for the couple i being in the block q; then, in order to extend this result to the entire population of treated, the weighted average of these differences is computed:

$$\widehat{\beta}_{|C_i=1} = \sum_{q=1}^{Q} \widehat{\beta}_q \frac{\sum_{i \in I(q)} C_i}{\sum_{\forall i} C_i}$$

where the weights in each stratum are the fraction of treated couples in each block. The standard errors for this estimator reported in tables 4-6 have been computed by boot-strapping with 200 repetitions.

An alternative is a linear regression or covariance adjustment techniques within each block. In this case, the treatment effect β_q is obtained by regressing D_i on the treatment and other covariates *within* each block, and then computing a weighted average exactly as outlined before (the weights are identical).The advantage of the "regression in matching" is that controlling again on the covariates X_i should help to eliminate the remaining within block-differences, although the results should no change when the covariates are well balanced.

9.3 The matching estimator

In the *nearest-match method,* each treated unit is matched to the control unit(s) with the closest propensity score. In symbols, the treated couple i is matched to that non-treated couple j such that:

$$p(X_i) - p(X_j) = \min_{k \in \{C_i=0\}} \{| \, p(X_i) - p(X_k) \, |\}$$

None of the treated couples is discarded in the nearest-match method because it is always possible to find a matched control even if it is far away from the treated couple.

The *radius method* of matching consists in matching each treated to the control couple(s) whose propensity score is within a $\delta-$radius chosen by the researcher. In symbols:

$$\delta > p(X_i) - p(X_j) = \min_{k \in \{C_i=0\}} \{| \, p(X_i) - p(X_k) \, |\}$$

If a treated couple has no control couples within a $\delta-$radius, this couple is discarded. Hence, switching from the nearest-match to the radius match one improves the quality of the matches but ends up using less observations and thus generate less precise estimates.

The average difference in marital outcomes of the sub-group of treated and the sub-group of matched comparisons is used to calculate the effect of having children *versus* not having children on marital dissolution. Formally, the matching estimator is:

$$\widehat{\beta}^M_{|C_i=1} = \frac{1}{N^T} \left[\sum_{i \in T} D_i - \sum_{i \in C} \omega_i^C D_i \right] \tag{7}$$

where T and C denote the sets of treated and matched control couples respectively and ω_i^C is the number of times a particular control $i \in C$ is used in the matching with a treated couple. Therefore, the average treatment effect is simply given by the average of the outcome in case of treatment minus the weighted average of the outcome in case of no treatment, with appropriate weights for repeated observations.[21] The standard error for this estimator reported in tables 4-6 have been computed by boot-strapping with 200 repetitions.

As for the stratification estimator, an alternative is to regress, over the sample of pairs, the divorce outcome on the treatment indicator of fertility and the covariates, with appropriate weights for repeated units (the logic for computing the weights is the same as described above). Note that, while in the stratification method the regression was simply an OLS estimation within blocks, in the matching method a weighted least square regression (WLS) is performed, where the weights are one for the treated and the number of times each control couples is used in the matching for the controls (see equation (7)).

[21] Note that N^T is equal to the number of all treated units in the nearest match methods, and to the number of the treated units for whom at least one matched control could be found in the radius method.

Table 1: Sample selection procedure

	Germany	USA	UK
(1) Couples married (only 1st marriage)	2110	2156	2658
(2) Couples lost in 5 years	462	329	549
Total (1)-(2)	1648	2187	2109
Couples where wife is <45 years old	874	1622	1157
...of whom divorced	*46*	*246*	*145*
Couples with total records	848	1580	923

Table 2: Descriptive statistics of observable covariates for all the couples and by children aged 0-18 (sample size 3351)

Variable	All couples		W. children 0-18		W/o children 0-18		
	mean	st.dev	mean	st.dev	mean	st.dev.	t-stat.
# children 0-18	1.51	1.12	-	-	-	-	-
# children 0-6	0.65	0.81	-	-	-	-	-
# children 6-18	0.87	1.02	-	-	-	-	-
having children 0-18	0.77	0.42	-	-	-	-	-
having children 0-6	0.45	0.49	-	-	-	-	-
having children 6-18	0.50	0.50	-	-	-	-	-
Duration of marriage	15.16	6.29	16.58	5.74	9.85	5.32	25.7
Husband's age	35.46	7.02	36.11	6.52	33.19	8.15	10.16
Wife's age	32.99	6.23	33.63	5.82	30.74	7.06	11.33
Mean age	34.22	6.41	34.87	5.95	31.97	7.37	11.09
Husband's education	12.89	2.86	12.77	2.8	13.32	3.02	4.63
Wife's education	12.60	2.68	12.46	2.65	13.11	2.72	5.87
Mean Education	12.75	2.42	12.61	2.37	13.21	2.53	6.01
Family gross income	71426	73291	70425	73758	74924	71574	1.48
Husband's wage	49331	53977	51249	56861	42566	41594	3.79
Wife's wage	15803	24434	12834	22154	26170	28821	13.49
Roman Catholic	0.18	0.38	0.18	0.38	0.18	0.38	0.42
No religion	0.08	0.26	0.07	0.25	0.1	0.3	3.2
Protestant	0.34	0.47	0.35	0.47	0.29	0.45	2.91
Germany	0.25	0.43	0.25	0.43	0.25	0.43	0.46
UK	0.27	0.44	0.27	0.44	0.45	0.5	2.28
USA	0.47	0.5	0.47	0.5	0.31	0.46	1.64

Notes: see note in table 4.

Table 3: Descriptive statistics of observable covariates by children aged 0-6

Variable	With children 0-6		Without children 0-6		
	mean	st.dev	mean	st.dev.	t-stat.
Duration of marriage	13.80	4.80	16.51	7.22	12.18
Husband's age	33.01	5.54	37.51	7.46	19.53
Wife's age	30.68	4.99	34.93	6.50	20.92
Mean age	31.84	5.01	36.22	6.76	20.95
Husband's education	12.98	2.85	12.82	2.87	1.49
Wife's education	12.70	2.51	12.52	2.82	1.85
Mean Education	12.84	2.33	12.67	2.49	1.9
Family gross income	63481	70866	78086	74633	5.77
Husband's wage	48102	58390	50363	49967	1.18
Wife's wage	10924	22843	19893	24973	10.76
Roman Catholic	0.18	0.38	0.18	0.38	0.24
No Religion	0.08	0.27	0.076	0.26	0.33
Protestant	0.34	0.47	0.33	0.47	0.62
Germany	0.23	0.42	0.27	0.44	2.29
UK	0.27	0.44	0.28	0.45	1.07
USA	0.50	0.5	0.45	0.5	2.9

Notes: see note in table 4.

Table 4: Descriptive statistics of observable covariates by children aged 6-18

Variable	With children 6-18		Without children 6-18		
	mean	st.dev	mean	st.dev.	t-stat.
Duration of marriage	19.71	4.40	10.81	448	55.25
Husband's age	38.87	5.57	32.01	6.64	32.38
Wife's age	36.23	4.65	29.71	5.9	35.53
Mean age	37.45	4.85	30.86	6.02	35.39
Husband's education	12.65	2.91	13.14	2.8	5.01
Wife's education	12.22	2.71	12.99	2.6	8.38
Mean Education	12.43	2.44	13.07	2.36	7.62
Family gross income	76368	74537	66429	71685	3.93
Husband's wage	54905	55731	43632	51520	5.96
Wife's wage	13579	20074	18053	27992	5.32
Roman Catholic	0.19	0.39	0.17	0.38	0.83
No Religion	0.10	0.30	0.10	0.30	5.0
Protestant	0.37	0.48	0.31	0.46	3.82
Germany	0.27	0.44	0.23	0.42	3.23
UK	.25	0.43	0.3	0.46	2.97
USA	0.47	0.5	0.47	0.5	0.17

Data legend: Mean Age: average age of the partners; Duration: duration of the marriage; Mean Education: average education of the partners; Family gross income: total income of the household ; Wife or husband's wage: real earnings; Roman Catholic: 1 if both partners catholic; Protestant: 1 if both partners protestant; No religion: 1 if both partners are atheist; Germany: 1 if german; USA: 1 if american; UK: 1 if british.

Notes: : monetary variables are in EURO (in 1990 for USA and Germany and in 1992 for UK). Note that all the explanatory variables refer to characteristics in 1990 (in 1992 for the British sample)

Table 5: Parametric Estimates of the effect of different measures of fertility on marital dissolution

		OLS	Probit
		(1)	(2)
(1)binary treatment[a]	having children 0-18	0.036	0.032
		(0.014)	(0.012)
	having children 0-6	-0.021	-0.014
		(0.012)	(0.010)
	having children 6-18	0.080	0.076
		(0.013)	(0.013)
(2) multi-valued treatment[b]	# children 0-18	0.019	0.019
		(0.005)	(0.004)
	# children 0-6	-0.013	-0.008
		(0.007)	(0.006)
	# children 6-18	0.033	0.033
		(0.006)	(0.005)

Notes: In column 2 (probit) marginal effects are reported .

[a]Least square regression: marital dissolution dummy on a costant, a fertility treatment indicator, duration of marriage, mean age, mean age squared, mean education, log of household income, log of wife's labor earnings, catholic dummy, protestant dummy, atheist dummy, German dummy, American dummy

[b]Least square regression: divorce dummy on a fertility multi-valued treatment indicator on the same covariates as in [a].

Table 6: Propensity score estimates of the effect of the presence of children between 0 and 18 years old on marital dissolution (ATT)

	Unadjusted	Adjusted[a]	N.obs.[b]
		OLS	
	(1)	(2)	(3)
Stratification:			
based on quintiles	0.020	0.004	2622
	(0.016)	(0.004)	
Matching:			
Nearest Match	0.019	0.011	2622
	(0.008)	(0.010)	
Radius: δ<0.0001	0.006	0.008	394
	(0.022)	(0.025)	
Radius: δ<0.0005	0.004	0.009	1518
	(0.010)	(0.013)	
Radius: δ<0.001	0.008	0.014	2186
	(0.008)	(0.011)	

Notes: Coefficients on the binary variable "Having children between 0 and 18 years old or not" are reported. Boot-strapped standard errors in parentheses.

Propensity scores are estimated using the probit model, with the following specification:

$\Pr(C_i=1)=F($marital duration, mean age, mean age^2, mean education, log of household total income, log of wife's labor earnings, catholic dummy, protestant dummy, atheist dummy, German dummy, American dummy

[a]Regression coefficients from linear regression of marital dissolution dummy on fertility indicator and all variables that enter the Probit, estimated by OLS (or Probit) on stratified sample and WLS on the matched sample, the weights on each control reflecting the number of times it is used in the matching.

[b]Number of observations refers to the actual number of comparison and treatment couples used for the stratification and matching estimators; namely, for the stratification estimator all treated couples and those comparison couples whose estimated propensity score is greated than the minimum, and less than the maximum estimated propensity score for the treatment group; for the matching estimator, all treated couples for whom a match close "enough" has been found.

Table 7: Propensity score estimates of the effect of the presence of children between 0 and 6 years old on marital dissolution (ATT)

	Unadjusted	Adjusted[a]	N.obs.[b]
	OLS		
	(1)	(2)	(3)
Stratification:			
based on quintiles	-0.003	-0.005	1527
	(0.015)	(0.015)	
Matching:			
Nearest Match	-0.016	-0.026	1527
	(0.012)	(0.010)	
Radius: δ<0.0001	-0.010	-0.013	440
	(0.020)	(0.018)	
Radius: δ<0.0005	-0.001	-0.0014	1096
	(0.002)	(0.0015)	
Radius: δ<0.001	-0.005	-0.004	1283
	(0.010)	(0.005)	

Notes: Coefficients on the binary variable "Having children between 0 and 6 years old or not" are reported.

Boot-strapped standard errors in parentheses.

Propensity scores are estimated using the probit model, with the following specification:

Pr(C_i=1)=F(duration of marriage, mean age, mean age^2, mean education, log of household total income, log of

wife's labor earnings, catholic dummy, protestant dummy, atheist dummy, German dummy, American dummy

[a]Regression coefficients from linear regression of marital dissolution dummy on fertility indicator and all variables that enter the Probit, estimated by OLS (or Probit) on stratified sample and WLS on the matched sample, the weights on each control reflecting the number of times it is used in the matching.

[b]Number of observations refers to the actual number of comparison and treatment couples used for the stratification and matching estimators; namely, for the stratification estimator all treated couples and those comparison couples whose estimated propensity score is greated than the minimum, and less than the maximum estimated propensity score for the treatment group; for the matching estimator, all treated couples for whom a match close "enough" has been found.

Table 8: Propensity score estimates of the effect of the presence of children between 6 and 18 years old on marital dissolution (ATT)

	Unadjusted	Adjusted[a]	N.obs.[b]
		OLS	
	(1)	(2)	(3)
Stratification:			
based on quintiles	0.031	0.020	1695
	(0.013)	(0.014)	
Matching:			
Nearest Match	0.050	0.039	1695
	(0.013)	(0.014)	
Radius: δ<0.0001	0.018	0.010	266
	(0.024)	(0.024)	
Radius: δ<0.0005	0.037	0.034	920
	(0.012)	(0.013)	
Radius: δ<0.001	0.043	0.037	1317
	(0.009)	(0.016)	

Notes: Coefficients on the binary variable "Having children between 6 and 18 years old or not" are reported. Boot-strapped standard errors in parentheses.

Propensity scores are estimated using the probit model, with the following specification:

$\Pr(C_i=1)=F$(duration of marriage, mean age, mean age^2, mean education, log of household total income, log of

wife's labor earnings, catholic dummy, protestant dummy, atheist dummy, German dummy, American dummy

[a]Regression coefficients from linear regression of marital dissolution dummy on fertility indicator and all variables that enter the Probit, estimated by OLS (or Probit) on stratified sample and WLS on the matched sample, the weights on each control reflecting the number of times it is used in the matching.

[b]Number of observations refers to the actual number of comparison and treatment couples used for the stratification and matching estimators; namely, for the stratification estimator all treated couples and those comparison couples whose estimated propensity score is greated than the minimum, and less than the maximum estimated propensity score for the treatment group; for the matching estimator, all treated couples for whom a match close "enough" has been found.

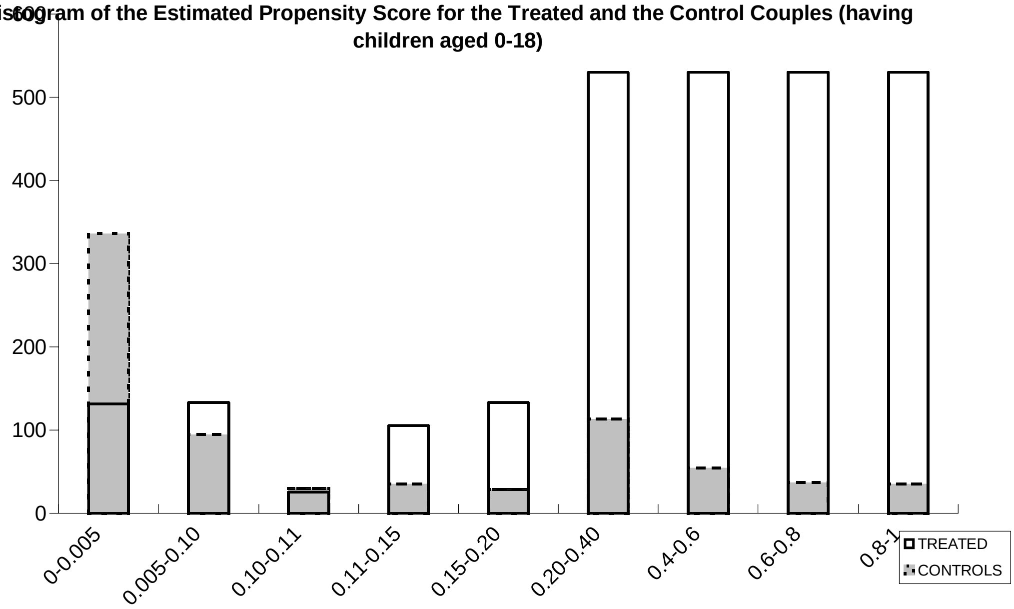
Table 9: Propensity score estimates of the effect of number of young children on marital dissolution

	# children 0-18	# children 0-6[a]	# children 6-18
From 0 to 1 child	0.042	-0.066	0.143
	(0.026)	(0.028)	(0.041)
From 1 to 2 children	0.048	-0.049	-0.006
	(0.032)	(0.015)	(0.016)
From 2 to 3 children	0.026	-0.037	0.077
	(0.045)	(0.012)	(0.019)
From 3 to 4(+) children	0.003	-	0.008
	(0.016)	-	(0.009)
Average effect	0.031	-0.060	0.070
	(0.015)	(0.023)	(0.018)

Note: Bootstrapped standard errors in parentheses. The covariates X included in the regression are the same listed in table 4. [a] For the treatment "number of children aged 0-6", the maximum number of children considered is 3 or more than 3.

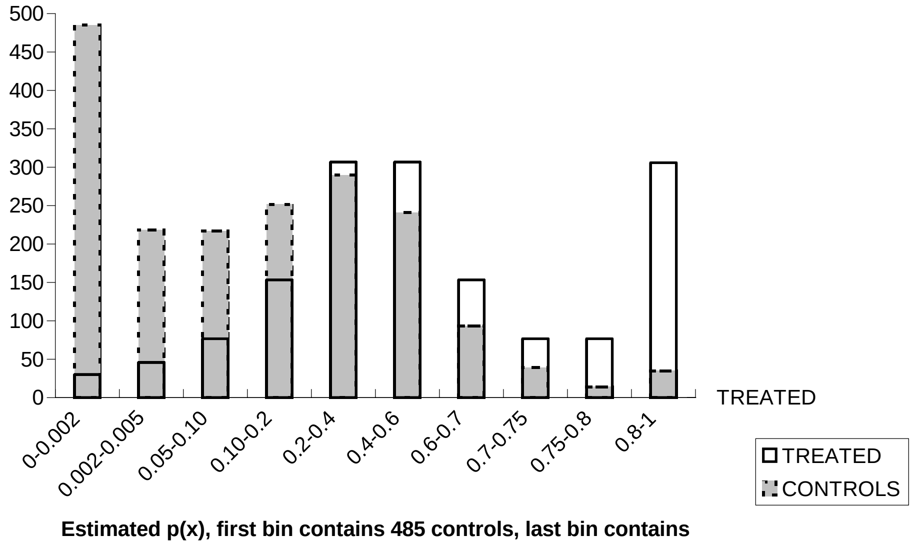
FIGURE 1

Histogram of the Estimated Propensity Score for the Treated and the Control Couples (having children aged 0-18)

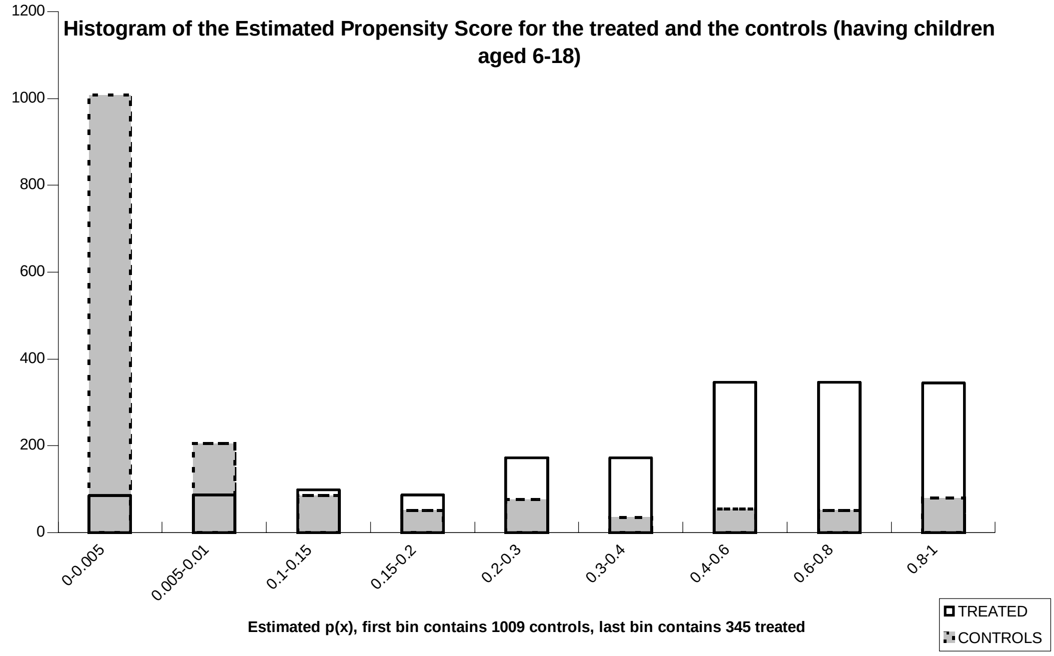


Estimated p(x), first bin contains 336 control units, last bin contains 530 treated units

Histogram of the Estimated Propensity Score for the Treated and the Controls (children aged 0-6)



Estimated p(x), first bin contains 485 controls, last bin contains 306 treated


Histogram of the Estimated Propensity Score for the treated and the controls (having children aged 6-18)
1200
1000
800
600
400
200
0
0-0.005
0.005-0.01
0.1-0.15
0.15-0.2
0.2-0.3
0.3-0.4
0.4-0.6
0.6-0.8
0.8-1
Estimated p(x), first bin contains 1009 controls, last bin contains 345 treated
TREATED
CONTROLS