What Tears Couples Apart: A Machine Learning Analysis of Union Dissolution in Germany

Bruno Arpino, Marco Le Moglie, and Letizia Mencarini

ABSTRACT   This study contributes to the literature on union dissolution by adopting a machine learning (ML) approach, specifically Random Survival Forests (RSF). We used RSF to analyze data on 2,038 married or cohabiting couples who participated in the German Socio-Economic Panel Survey, and found that RSF had considerably better predictive accuracy than conventional regression models. The man's and the woman's life satisfaction and the woman's percentage of housework were the most important predictors of union dissolution; several other variables (e.g., woman's working hours, being married) also showed substantial predictive power. RSF was able to detect complex patterns of association, and some predictors examined in previous studies showed marginal or null predictive power. Finally, while we found that some personality traits were strongly predictive of union dissolution, no interactions between those traits were evident, possibly reflecting assortative mating by personality traits. From a methodological point of view, the study demonstrates the potential benefits of ML techniques for the analysis of union dissolution and for demographic research in general. Key features of ML include the ability to handle a large number of predictors, the automatic detection of nonlinearities and nonadditivities between predictors and the outcome, generally superior predictive accuracy, and robustness against multicollinearity.

KEYWORDS   Union dissolution • Machine learning • Random Survival Forests • Germany • Socio-Economic Panel Study

Introduction

The literature on union dissolution in Western countries is extensive, and reviews (e.g., Lyngstad and Jalovaara 2010; Mortelmans 2020) highlight how knowledge of the predictors of union dissolution has increased considerably in recent decades. This deepening understanding is consistent with the broader diffusion of the phenomenon itself, but is also due to the emergence of high-quality survey data.

This article offers three contributions to the literature. First, we identify predictors of union dissolution in Germany using a machine learning (ML) approach (Hastie et al. 2009; Molina and Garip 2019)—specifically, Random Survival Forests (RSF). This approach, which does not impose any sharp limits on the number of variables

that can be included, allows us to account simultaneously for the numerous predictors of union dissolution highlighted by previous studies while, at the same time, permitting the exploration of nonlinearities and nonadditivities in the links between these predictors and union dissolution. Second, we focus on the role of both partners' personality traits (PTs) in union dissolution. Previous studies have not considered the PTs of both partners at the same time, nor the interactions of PTs with each other and with other predictors. Implementing this kind of analysis would be complicated with a standard regression-based approach; we show that ML offers a different approach that can handle a high number of PT items (five per partner) and potential interactions among them. More specifically, we examine the importance of PTs relative to other variables in predicting union dissolution and whether pairwise combinations of PTs (within or between partners) are associated with union survival probabilities. Third, we assess predictive accuracy, which typically has been overlooked in previous studies on union dissolution. ML is a useful approach here. For instance, conventional regression models appear to have lower out-of-sample predictive accuracy than RSF and other algorithmic approaches. This is an important advantage of ML approaches because, as Hofman et al. (2017) argue, prediction, like explanation, is crucial for theory building and validation. We also show that insights from ML algorithms can be used to better specify standard regression models.

Our study focuses on Germany and uses data from the German Socio-Economic Panel Study (SOEP) to examine union dissolution among individuals who married or began cohabiting between 1984 and 2015. SOEP is particularly useful for the purposes of our study because it includes most of the variables that previous studies employed to predict union dissolution.

In addition to offering abundant survey data, Germany provides an interesting setting for this analysis. Divorce has long been commonplace in Germany; the current duration-specific divorce rate suggests that more than 40% of marriages dissolve (Vignoli et al. 2018). However, the divorce rate, after decades of increases (Wagner et al. 2015), declined from 329 per 1,000 people in 2003 to about 218 per 1,000 in 2018, according to the German Statistical Office. The dissolution rate for cohabiting individuals has been estimated to be more than double that for married individuals (Andersson 2003; Hiekel and Wagner 2020).

From a methodological point of view, our analyses show the relevance of ML techniques in the study of union dissolution and in demographic research more generally. We show that RSF is able to automatically detect nonlinear links between union dissolution and its predictors. We also demonstrate that the method allows the examination of interactions among a large number of independent variables, thus permitting the exploration of complex patterns within relatively small, survey-based data sets. One of the primary advantages of ML methods is that they do not require the researcher to make assumptions about parametric distributions (Molina and Garip 2019). These advantages are critical, as regression-based results in social science research can be influenced by model specification choices (Young 2009) or the particular link function for the linear predictor (Young and Holsteen 2017). Moreover, some demographers may have been reluctant to use ML approaches because of the methods' black-box nature. Our analyses show how to "open the box"—that is, how to gain substantive insights from the RSF algorithm.

Previous Studies on Union Dissolution

Union dissolution occurs worldwide, but with great heterogeneity within and across countries (Emery 2013; Mortelmans 2020; Wagner and Weiß 2006). With the decline in the prevalence of legal marriage and the spread of cohabitation as a stable form of union, the term "union dissolution" has become an umbrella term for all uncoupling processes, irrespective of legal bond (Mortelmans 2020).

The European research tradition in union dissolution came long after U.S.-centered studies and found heterogeneity in dissolution within Europe, corresponding to variation in the pace of diffusion of the phenomenon. In particular, union dissolution has occurred along a strong north–south gradient; rates have been highest in Scandinavia, while unions tend to be more stable in southern Europe, though dissolution rates have increased there as well (Kalmijn 2007).

In reviewing the previous two decades of research on Europe and the United States, Lyngstad and Jalovaara (2010) offered a classification of the predictors of divorce. Here, in summarizing these predictors, we refer mainly to Europe and, when possible, specifically to Germany.

The first group of predictors comprises the personal characteristics of the members of the couple, the most important being age, age at union formation, union duration, education, personality, and subjective well-being. An individual's age is a better predictor of union dissolution than age at union formation (Lutz et al. 1991). In general, age is negatively correlated with the probability of dissolution (Lyngstad and Jalovaara 2010). The effect of education is rather complex (Härkönen and Dronkers 2006; Lyngstad and Jalovaara 2010; van Damme 2020) and appears to follow a pattern linked to the so-called Goode (1962) hypothesis, which states that education's impact varies with the prevalence of union dissolution in the country. Consistent with this hypothesis, the effect of education on union dissolution is negative in areas with high union dissolution, such as the United States and Scandinavia, but positive in countries such as Italy, where the prevalence of union dissolution, though growing, is relatively low (e.g., Härkönen and Dronkers 2006; Matysiak et al. 2014; Salvini and Vignoli 2011; Vignoli and Ferro 2009). Several studies have highlighted the importance of the interaction between partners' education levels. Homogamy tends to protect unions from breaking up (e.g., Kalmijn 1998; Kalmijn 2003; Mäenpää and Jalovaara 2014), although in Germany, Kraft and Neimann (2009) suggested that homogamy does not increase marital stability but rather higher education per se does, and Blossfeld (2014) found that homogamous marriages rank second in stability after women's educationally upward marriages. Moreover, the effect of education needs to be separated from the effects of partners' income and labor market status.

Regarding partners' personality, most of the literature on its influence on union dissolution relies on the "big five" traits (agreeableness, conscientiousness, extraversion, neuroticism, and openness). In recent studies, associations between PTs and union dissolution have been similar in the United Kingdom, Flanders, and Germany. In particular, a low level of conscientiousness and a high level of openness are significant risk factors, although the relationship between openness and dissolution has weakened as the latter becomes more common and less costly socially and economically (Boertien and Mortelmans 2018; Boertien et al. 2015).

Most of the literature on subjective well-being has investigated how happiness, or overall life satisfaction, differs by marital status (e.g., Oswald 1997) and is influenced by union dissolution (e.g., Gardner and Oswald 2006). To the best of our knowledge, no studies have explicitly assessed the role of partners' overall life satisfaction as a predictor of union dissolution. Past research, however, has found that dissatisfaction with the relationship is an important antecedent of union dissolution, especially among women (e.g., Røsand et al. 2014).

A second group of predictors of union dissolution consists of the partners' economic status and their division of paid and unpaid labor. Empirical evidence across the latter part of the past century indicates that, in many countries, employed married women were more likely to divorce than those who were not employed (e.g., Blossfeld and Müller 2002; Chan and Halpin 2002; Cooke 2004, 2006; De Rose 1992; Lyngstad and Jalovaara 2010; Ozcan and Breen 2012; Poortman 2005a, 2005b; see Vignoli et al. 2018 specifically for Germany). This gave rise to the idea of the so-called "independence effect" (Cooke et al. 2013), in which a married woman's working full-time leads to a higher likelihood of divorce, a phenomenon that appears to dominate the income effect—that is, the positive effect on union stability from the fact that the wife's resources also add to the total resources of the family. However, a new strand of studies suggests, instead, that women's employment does not have a negative effect per se, and that women's paid work becomes detrimental to union stability only if men's contribution to unpaid work within the household is limited (e.g., Mencarini and Vignoli 2017). In light of this, gender equality within couples in the sharing of domestic chores becomes an important positive predictor of union stability (e.g., Cooke et al. 2013; Frisco and Williams 2003; Oláh 2001; Oláh and Gahler 2014; Sigle-Rushton 2010; and, specifically for Germany, Bellani and Esping-Andersen 2020; Bellani et al. 2018). Similar effects are found in terms of earning equality, especially for cohabiting couples and for younger married couples (Ishizuka 2018; Jalovaara 2003; Kalmijn et al. 2007).

A third group of predictors includes the couple's characteristics, such as whether they are married (as opposed to cohabiting), their number of children, and union duration. Rates of union dissolution are generally higher for cohabiting than for married couples, a pattern that appears to be independent of the presence of children (e.g., Andersson 2002; Berrington and Diamond 1999; Liefbroer and Dourleijn 2006). However, the difference seems to be explained (at least in part) by self-selection into cohabitation or marriage (e.g., Svarer 2004). Furthermore, whether a couple cohabited prior to marriage seems to have mixed effects on union stability (Lyngstad and Jalovaara 2010). The risk of dissolution among couples who marry after cohabiting is generally lower among those with children, and is especially low during the period following the birth of the first child. However, this lower risk appears to be partially driven by selection, whereby partners who have little trust in the continuity of their union are less likely to have children (Lyngstad and Jalovaara 2010). Finally, union duration is negatively correlated with the probability of dissolution (Lyngstad and Jalovaara 2010).

A last important variable, which is specific to Germany and included in our analysis, refers to the East–West divide. East Germans, other factors being equal, are more likely than West Germans to divorce (e.g., Boertien and Mortelmans 2018).

The factors listed in the foregoing are the predictors most commonly considered in union dissolution studies. However, specific studies, using particular surveys, have

found other important characteristics linked to union dissolution, such as biological and genetic characteristics, migration and minority status, values and religiosity, and whether the partners' own parents divorced.

Despite the many studies on union dissolution in Europe (including Germany), the literature certainly presents some gaps. First, the reasons behind the often mixed and heterogeneous results are not well settled and may have been driven by model specification issues (i.e., different parametric specifications). As an example of nonlinear effects, Cooke and Gash (2010) used a categorical measure of wives' employment hours and admitted that if they coded this measure as continuous they would have needed different model functional forms for each country considered: namely, linear for West Germany, quadratic for Great Britain, and third-order polynomial for the United States. Another example, quoted in Lyngstad and Jalovaara's (2010) review, regards the wife's income, which often displays a nonlinear relationship with union dissolution.

Second, studies have tested interactions between specific independent variables. For example, Cooke and Gash (2010) found evidence of nonadditive effects (i.e., interactions) between children's age and women's employment, and Mencarini and Vignoli (2017) reported an interaction between women's employment and men's contribution to housework. However, previous studies using parametric models could not consider all (not even many) interactions between independent variables simultaneously. A clear-cut example of the potential difficulties of considering all variables and their possible interactions concerns personality traits: to account for both partners' PTs (10 variables) and all their two-way interactions (25 variables), one would need to include 35 independent variables, which would be very problematic in a regression model. Thus, despite our increasing understanding of union dissolution, there is still a strong need for in-depth studies to disentangle interwoven and matted complexities.

Third, previous studies have focused on explaining findings but have overlooked the importance of outcome predictive accuracy. Addressing these gaps demands that we account for a large set of union dissolution predictors and their interactions, and this may not be feasible with conventional event-history regression models. It is better accomplished using an ML-based approach.

Methods

Our study relies on an ML technique, RSF, because of the advantages it offers, which are outlined below. However, given that RSF is an unconventional approach in the union dissolution literature, we also present results from standard discrete-time event-history logit models ("DT models" henceforth). Our goal is not to directly compare the results from two fundamentally different approaches, but to bring out the different implementation and information content of the new approach. In this regard, it is worth stressing that the conventional DT model and the RSF are completely different approaches by nature. DT models and RSF belong to what Breiman (2001a) refers to, respectively, as *data modeling culture* and *algorithmic modeling culture*. Data modeling culture is based on a "representation of the mechanisms by which nature works," while algorithmic modeling culture "is concerned solely with linking

inputs to outputs" (Berk 2016:327). RSF, like similar algorithms, can be treated as a form of regression analysis (Berk 2016). However, there is a key difference between them and conventional parametric regression models: algorithms do not require for a model of nature to be stated, and the estimation target can be considered the best approximation to the true response surface, from which the researcher can get useful information about how the outcome is related to its predictors.

The two statistical cultures have developed largely in parallel. Data analysis culture has focused almost exclusively on explanation and has overlooked the importance of predictive accuracy, while algorithmic approaches have tended to focus on prediction (for further discussion, see Berk 2016; Hofman et al. 2017). Recently, a rapidly growing movement hybridizing the two cultures has emerged, and has asserted the need to combine explanation and prediction (Hofman et al. 2017) and suggested methods for improving the interpretation of findings from algorithmic approaches (e.g., Zhao and Hastie 2021). We follow these developments by recognizing the importance of prediction in substantive studies. Thus, although our main goal is to examine the possibly complex way in which a large set of predictors may be related to union dissolution, we use an ML technique also to improve the predictive accuracy of union dissolution.

Statistical inference with RSF, as with most ML techniques, is an unsolved problem because of these methods' inductive nature and their use of data snooping (see Berk 2016 for a discussion of this issue and for references to recent approaches to statistical inference with random forests). This is another key difference between algorithmic and model-based research, as the latter typically relies on considerations of statistical significance.

In the following sections, we will first briefly discuss some issues concerning conventional DT models used to analyze union dissolution and then focus on RSF. Note again that the inclusion of both approaches is not done with the aim of directly comparing their results. Instead, reporting results from a conventional DT model serves to clarify the advantages offered by RSF in the study of union dissolution.

DT Models: Issues in the Analysis of Union Dissolution

Previous studies on union dissolution have often relied on DT models, which model the probability of experiencing union dissolution during a given interval as a function of a set of independent variables and of time spent in the risk set (duration). Applying parametric models, such as DT models, may prove difficult for our research goals for several reasons. First, as the preceding literature review suggests, our reanalysis of the predictors of union dissolution has to account for a large number of independent variables. For instance, PTs alone account for 10 independent variables (five for each partner). Second, we are interested in understanding whether characteristics of the same individual or of the partner interact in influencing union dissolution. Several studies have identified important nonlinear and nonadditive effects of predictors of union dissolution that should be considered when specifying the model. This implies that an intractable number of interactions should be included in a regression model. Third, several antecedents of union dissolution are correlated (e.g., partners' education, income, and working hours) and consequently their inclusion may give rise to multicollinearity.

Another limitation of previous studies based on DT models concerns the way model fit is commonly assessed, especially with respect to different choices for the duration, which results in models that differ in their baseline hazard function. Demographers typically choose the best specification of the baseline hazard function according to measures of model fit, like the Akaike information criterion (AIC), for which lower values denote a better fit. Model fit is evaluated using the same sample (*in-sample*) from which the model is estimated, giving rise to possible issues of *overfitting*, meaning that prediction quality can be very good in the specific working sample, but not in others (*out-of-sample*). This, of course, limits external validity (i.e., generalizability to other data sets) (Hofman et al. 2017). We follow the common practice in ML of separating estimation from the evaluation of predictive accuracy (see Berk 2006); that is, DT models will be estimated on a randomly drawn subsample (*training set*), and predictive accuracy will be assessed on the remaining part of the initial sample (*testing set*).

The predictive accuracy of DT models will also be assessed on the basis of the area under the curve (AUC). In the context of survival analyses, time-dependent AUC measures can be calculated for assessing how predictive accuracy varies at different time points. An AUC of .5 is not better than random guessing, while an AUC of 1 indicates perfect predictive accuracy. Because AUC does not distinguish between false positives and false negatives, we assume that the costs of incorrect predictions of union survival are the same as those for incorrect predictions of union dissolution.

Random Survival Forests

RSF belongs to the category of *supervised* ML (SML) techniques.[1] Generally speaking, ML algorithms can be defined as procedures that give computers the ability to learn without being explicitly programmed (Samuel 1959). SML techniques are a subgroup of ML techniques that consist of iterative algorithms and are focused primarily on prediction problems.

RSF is a modification of random forests for survival data, which in turn is an ensemble of single classification trees (Breiman 1996; Breiman et al. 1984). A classification tree (usually known as a regression tree in the case of a continuous outcome) uses a recursive algorithm to describe the relationship between a set of independent variables and a given outcome. At each step, the algorithm splits the data into subsets. Splits can occur between any two subsets of the observed values of any of the predictors. Among all possible splits, the algorithm selects the one that minimizes the prediction error (for more details, see James et al. 2013; Shu 2020). A classification tree captures nonlinearities in covariates by automatically splitting them into different intervals. Another general feature of trees is the automatic detection of interactions. In this context, a two-way interaction between two variables is found if a split in one variable makes a split in the other either systematically less likely or more likely.

---

[1] More formal and detailed discussions of ML techniques are available in the books by Berk (2016), Hastie et al. (2009), and James et al. (2013). These offer comprehensive and accessible accounts of the main ML techniques.

Random forests are among the most popular and best-performing SML techniques (Breiman 1996, 2001b; Cannas and Arpino 2019; Glaeser et al. 2016). Random forests are "ensemble" algorithms based on many multiple trees (Berk 2006). The prediction obtained with random forests is constructed on the basis of hundreds or thousands of distinct trees that differ from one another because each tree is fit on different data obtained as bootstrap samples of the original data. Additionally, each tree is grown using a random subset of variables at each split. A single prediction is obtained by majority vote across all trees: the class predicted by each tree is recorded, and the overall prediction is the most commonly occurring class (James et al. 2013). By relying on a multitude of trees, random forests reduce problems of overfitting and usually show better prediction quality than single trees (Athey and Imbens 2017). However, while some rare examples of demographic studies using single trees exist (Billari et al. 2006; De Rose and Pallara 1997), we are not aware of any implementation of ensemble SML techniques in demographic analyses, with the exception of studies on mortality (e.g., Bitew et al. 2020).

We use a modification of random forests, the RSF, which is an ensemble tree–based method for the analysis of right-censored survival data (Hothorn et al. 2006; Hothorn et al. 2004; Ishwaran et al. 2008). More specifically, we use the algorithm implemented in the package *randomForestSRC* in the open-source environment R (Ishwaran and Kogalur 2014), which builds an ensemble of hazard functions using the log-rank statistic as the default splitting rule.

Like other tree-based approaches, RSF has several advantages (James et al. 2013). It can handle automatically (i.e., without need for recoding, grouping, etc.) continuous, nominal, and ordinal independent variables, and can automatically capture non-linear effects and interactions. Another important attribute of RSF is its ability to use a large number of predictors, even if most of them are correlated.

To assess the performance of the algorithm in predicting survival, we first look at the *out-of-bag* (OOB) error rate. RSF does not need an independent validation data set to get an unbiased estimate of the test set error, as this is estimated internally during the run of the algorithm. More specifically, each tree in the forest is constructed by bootstrapping a sample from the original data. Sampling with replacement implies that for each tree, a random sample of about a third of the observations are excluded from the calculations—OOB observations—giving rise to an immediate source of test data (Berk 2016:192). Then, each OOB observation in the construction of the *k*th tree is dropped down the tree, and the algorithm estimates the percentage of times that the class assigned to OOB observations is not equal to the true class. Finally, the total OOB error is obtained as the average of these estimates across all the trees in the forest. As with the DT models, we assess the predictive accuracy of RSF by also estimating AUCs at different time points.

A disadvantage of RSF (as with all ensemble methods) is that by combining multiple trees, it does not offer a single tree to visualize and interpret: that is, it is difficult to understand how inputs are related to the output. However, several measures can be calculated to ease interpretation. First, it is possible to obtain a measure of variable importance (Zhao and Hastie 2021). We focus on measuring the importance of a variable by its contribution to predictive accuracy. More specifically, the variable importance measure (VIMP) that we calculate for variable X is the difference

between the prediction error when noise in the original variable X is added by permuting its values randomly and the prediction error under the original predictor (Breiman 2001b; Ishwaran 2007; Ishwaran et al. 2008). Large VIMP values are linked to variables with some predictive power (since inducing noise in these variables increases prediction errors), whereas zero or negative VIMP values identify variables that are not predictive of the outcome. Setting thresholds to separate highly predictive variables from others using VIMP is difficult (Ishwaran et al. 2011), and distinguishing variables according to their predictive ability is usually done rather arbitrarily by ranking variables based on their VIMP and considering thresholds where there is a large decrease in VIMP values. We complement the ranking of variables based on VIMP with the inspection of partial dependence plots.

Second, to reveal how each predictor is related to the outcome, one useful solution is to produce *partial dependence plots* (Friedman 2001; Hastie et al. 2009) that show the relationship between a given predictor and the response averaged over the joint values of the other predictors as they are represented in a tree structure. Similarly, *partial dependence co-plots* represent how the predicted outcomes vary as a function of the joint distribution of two or three predictors, thus making it possible to visualize potential interactions. Results from RSF can be used to automatically detect the presence of independent variables whose pairwise interaction may be better investigated using partial dependence co-plots. The approach involves calculating joint VIMPs for all pairs of predictors and has two steps. First, two variables X and Z are paired, and their paired VIMP is calculated, as mentioned earlier, for a single variable. Second, the separated VIMPs for X and Z are also calculated and their values are added up. Then, conditional to a relatively large univariate VIMP for both X and Z, the larger the absolute difference between the paired VIMP and the additive VIMP, the more potentially interesting the interaction between the two variables is.

Data and Variables

We used data from SOEP, a nationally representative ongoing longitudinal study started in 1984. SOEP is well suited for the study of union dissolution for two reasons. First, the length of the study allows us to follow individuals over a long period. Second, it includes information on union dissolution and the main predictors that were identified in past studies.

In order to have information on both partners, we selected, from the original database, women whose partners were also surveyed. In particular, we included women aged 65 years or younger who started their relationship during the observation period (1984 to 2015). The final sample consisted of 18,613 observations, corresponding to 2,038 couples, married or in a cohabiting union, who were observed, on average, for 12 years.

For the dependent variable, we constructed a dummy variable that measured union dissolution and that was equal to 1 when we observed a change in a woman's partner from year $t-1$ to t, and to 0 otherwise. After union dissolution, we stopped following both members of the couple, which means that our sample included no more than one

Table 1 Summary statistics for the independent variables

Variable	Mean	SD	Minimum	Maximum
Personal Characteristics				
Age (woman)	36	9	17	65
Age (man)	39	9	17	65
Age homogamy[a]	0.1			
Older woman[a]	0.2			
Tertiary education (woman)[a]	0.2			
Tertiary education (man)[a]	0.2			
Education homogamy[a]	0.7			
More educated woman[a]	0.1			
Life satisfaction (woman)	7.2	1.6	0	10
Life satisfaction (man)	7.1	1.6	0	10
Health (woman)	2.4	0.8	1	5
Health (man)	2.4	0.8	1	5
Agreeableness (woman)	5.5	0.9	1	7
Conscientiousness (woman)	6.1	0.9	1	7
Extraversion (woman)	5.0	1.2	1	7
Neuroticism (woman)	3.8	1.2	1	7
Openness (woman)	4.6	1.3	1	7
Agreeableness (man)	5.3	1.0	2	7
Conscientiousness (man)	5.9	0.9	1	7
Extraversion (man)	4.8	1.2	1	7
Neuroticism (man)	4.3	1.2	1	7
Openness (man)	4.4	1.2	1	7
Economic Situation				
Labor income (woman)	5.6	2.8	1	10
Labor income (man)	6.0	2.6	1	10
Labor income homogamy[a]	0.1			
Richer woman[a]	0.4			
Percentage of labor income (woman)	34	25	0	100
Unemployment (woman)[a]	0.1			
Unemployment (man)[a]	0.1			
Working hours (woman)	24	18	0	80
Working hours (man)	40	16	0	80
Couples Characteristics				
Married[a]	0.7			
Percentage of housework (woman)	72	22	0	100
No. of children	1.3	1.1	0	7
West Germany[a]	0.8			

Note: Summaries are calculated for the 18,613 observations included in the baseline sample.

[a] For dummy variables (all coded 0/1), we report only the mean (proportion).

union dissolution per individual. During the observation period, 914 couples (45%) split up.

On the basis of the literature review, we considered 28 predictors of union dissolution measured for both members of the couple or at the couple level. Table 1 lists all these variables and provides summary statistics. A description of the operationalization of each variable is given in the online Appendix A, while Appendix B provides the reader with the R code for replicating the results of the analysis introduced in the next section.

Table 2 In-sample and out-of-sample Akaike information criterion (AIC) and area under the curve (AUC) for five discrete-time event-history models

Measure	Linear (DT1)	Quadratic (DT2)	Cubic (DT3)	Dummy Variables (DT4)	Splines (DT5)
In-Sample					
AIC	6,912.805	6,890.300	6,879.962	6,913.825	6,885.830
AUC, year 1	.879	.908	.881	.921	.926
AUC, year 5	.853	.870	.875	.871	.874
AUC, year 15	.758	.781	.785	.778	.784
AUC, year 25	.570	.715	.680	.767	.684
Out-of-Sample					
AIC	3,547.659	3,536.851	3,533.052	3,565.267	3,538.321
AUC, year 1	.512	.516	.517	.518	.515
AUC, year 5	.512	.509	.512	.514	.512
AUC, year 15	.490	.497	.496	.489	.496
AUC, year 25	.566	.578	.576	.570	.571

Notes: DT1 through DT5 are discrete-time event-history models that differ for the specification of the baseline hazard. To determine out-of-sample predictions, the original sample was randomly split into two parts of the same size, with one used as a training data set and the other as a testing data set.

Results

Discrete-Time Event-History Models

We considered five different specifications for duration in the DT models: linear, quadratic, cubic, step function (dummy variables for each year), and b-spline. For each of these model specifications, Table 2 presents the AIC and AUC at different survival times (one, five, 15, and 25 years), both in-sample and out-of-sample. The former was obtained by using the entire sample of couples; for the latter, we randomly split the original sample into two parts of equal size, which were then separately used as training and test sets, and calculated the measures of predictive performance using the test set. Table 2 shows very clearly that while all models were fairly good at making in-sample predictions, they performed very poorly at out-of-sample predictions. In almost all cases, models were only slightly better than a simple random guess (i.e., AUC = .50).

Estimates and the statistical significance of all coefficients were very similar across all the DT models. Thus, in Table 3, we focused on estimates from model DT3, which shows the best predictive accuracy. In this model, only six variables had coefficients that were statistically significant. This model indicates that higher life satisfaction for both partners, as well as older age for women, was significantly associated with a lower probability of union dissolution. On the other hand, older age among men, higher openness scores among men, and higher number of working hours among women were associated with a greater risk of dissolution. That only six independent variables were associated with union dissolution is perhaps due to multicollinearity, which may inflate standard errors. We checked for multicollinearity in model DT3 using several diagnostics available in the R package *mctest* (see Imdadullah et al. 2016 for details on the diagnostics). Results are given in Tables A.2

Table 3 Estimates of a discrete-time hazard model with a cubic specification for the duration

Variable	Estimate	SE	z Value	p Value
Agreeableness (woman)	−0.048	0.054	−0.885	.376
Conscientiousness (woman)	−0.024	0.057	−0.419	.675
Extraversion (woman)	0.001	0.046	0.012	.990
Neuroticism (woman)	−0.028	0.042	−0.669	.503
Openness (woman)	−0.052	0.044	−1.196	.232
Agreeableness (man)	0.030	0.053	0.557	.578
Conscientiousness (man)	−0.040	0.057	−0.712	.476
Extraversion (man)	0.054	0.046	1.177	.239
Neuroticism (man)	0.042	0.044	0.957	.339
Openness (man)	0.106	0.047	2.284	.022*
Tertiary Education (woman)	−0.143	0.191	−0.746	.456
Tertiary Education (man)	−0.176	0.192	−0.917	.359
More Educated Woman	0.208	0.257	0.807	.420
Education Homogamy	−0.039	0.164	−0.238	.812
No. of Children	0.082	0.054	1.525	.127
Age (woman)	−0.027	0.013	−1.981	.048*
Age (man)	0.043	0.012	3.507	.000**
Older Woman	0.226	0.166	1.355	.175
Age Homogamy	0.149	0.179	0.830	.407
Working Hours (woman)	0.008	0.004	1.843	.065†
Working Hours (man)	−0.004	0.004	−1.062	.288
Percentage of Housework (woman)	−0.003	0.002	−1.182	.237
Life Satisfaction (man)	−0.118	0.034	−3.431	.001**
Life Satisfaction (woman)	−0.073	0.035	−2.095	.036*
Unemployment (woman)	−0.147	0.255	−0.578	.563
Unemployment (man)	−0.230	0.258	−0.892	.373
Labor Income (woman)	−0.029	0.041	−0.700	.484
Labor Income (man)	−0.005	0.038	−0.121	.904
Richer Woman	0.127	0.214	0.595	.552
Labor Income Homogamy	0.053	0.198	0.266	.790
Percentage of Labor Income (woman)	−0.001	0.004	−0.237	.813
Health (man)	0.058	0.065	0.901	.368
Health (woman)	−0.017	0.064	−0.271	.786
Married	−0.126	0.115	−1.097	.273
West Germany	0.142	0.130	1.093	.274

Notes: Coefficients for duration coefficients are not shown, but are available upon request. Estimates are obtained by running the regression on the test sample.

†$p<.10$; *$p<.05$; **$p<0.01$

and A.3 of the online Appendix C. We first considered six tests of overall multicollinearity. Four of these tests detected multicollinearity issues (Table A.2). Then we considered seven tests for individual multicollinearity for each of the independent variables (Table A.3); three of these tests detected multicollinearity issues for each of the independent variables.

The DT models presented here can be made more flexible by including specific interactions and nonlinear terms, usually introduced to test specific hypotheses. In the following analysis, we took a different approach by implementing an RSF. We asked the algorithm to automatically detect potentially relevant interactions and

Table 4 Area under the curve (AUC) for Random Survival Forests at different durations

Year	AUC
Year 1	.711
Year 5	.657
Year 15	.630
Year 25	.588

nonlinearities. This approach does not suffer from the multicollinearity issues that, in our data set, were likely to have affected the statistical significance of coefficients in model DT3.

RSF Algorithms

The RSF is based on a multitude of trees. Each tree is based on an iterative algorithm illustrated in Figure A.1 in the online Appendix C. This illustrative tree demonstrates how classification proceeds at each node.

Assessing the Algorithm's Performance

Before the RSF algorithm is run, three important parameters have to be set: the size of the terminal nodes, the number of variables to be randomly selected at each split, and the number of trees to be grown in the forest. For the first two parameters, we implemented the algorithm developed by Ishwaran and Kogalur (2014), which searches for the combination of parameters that minimizes the OOB error rate. The contour plot in Figure A.2 graphically represents the analysis generated through the algorithm. This shows that the OOB error rate was minimized when the size of terminal nodes was set to one and the number of variables randomly selected at each split was four.[2] After setting the values for the first two parameters, we fine-tuned the third (number of trees grown in the forest). Again, the number of trees that are grown can be determined by choosing the value that minimizes the OOB error rate. In our case, the OOB error rate stabilized around 35% when at least 500 trees were employed (see Figure A.3 in the online Appendix C). We chose to use 1,000 trees, which yielded a final OOB error rate of 34%. Note that although it increases computational time, choosing a high number of trees does not create problems in terms of overfitting (Breiman 2001b).

In Table 4 we report the AUC calculated at one, five, 15, and 25 years. The predictive accuracy of the algorithm fell with duration: it was highest at one year and lowest at 25 years. We note that the predictive accuracy (out-of-sample) of RSF was considerably superior to that of DT, consistent with the usual finding that SML improves

---

[2] The algorithm grows a forest of 50 trees for each pair of parameters, and then calculates the OOB error rate associated with that forest. The contour plot graphically shows the level of OOB obtained from each of these calculations.
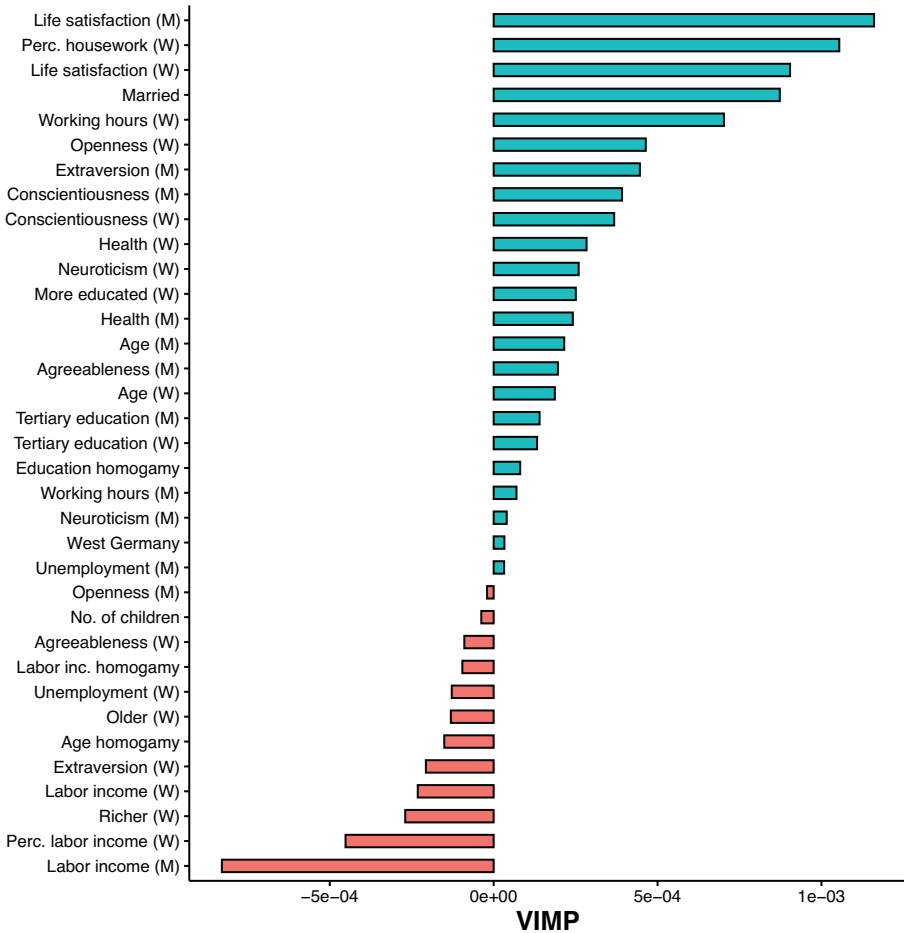
Fig. 1 Variable importance (VIMP) measures for all 35 independent variables included as predictors in the Random Survival Forests

prediction. Nonetheless, the predictive accuracy of RSF was limited despite the use, as input variables, of all the most important predictors of union dissolution identified in the literature.

Variable Importance

Figure 1 shows the VIMP of each variable used in the RSF. Among the variables with the greatest predictive ability, we find the life satisfaction of both partners, woman's percentage of housework, marital status (i.e., married vs. cohabiting), woman's working hours, woman's level of openness, and man's level of extraversion. Variables with the lowest predictive ability include man's and woman's labor income, woman's percentage of labor income, if woman is richer than man, and woman's level of extraversion. Note that the sign of VIMP is not informative of the sign of the
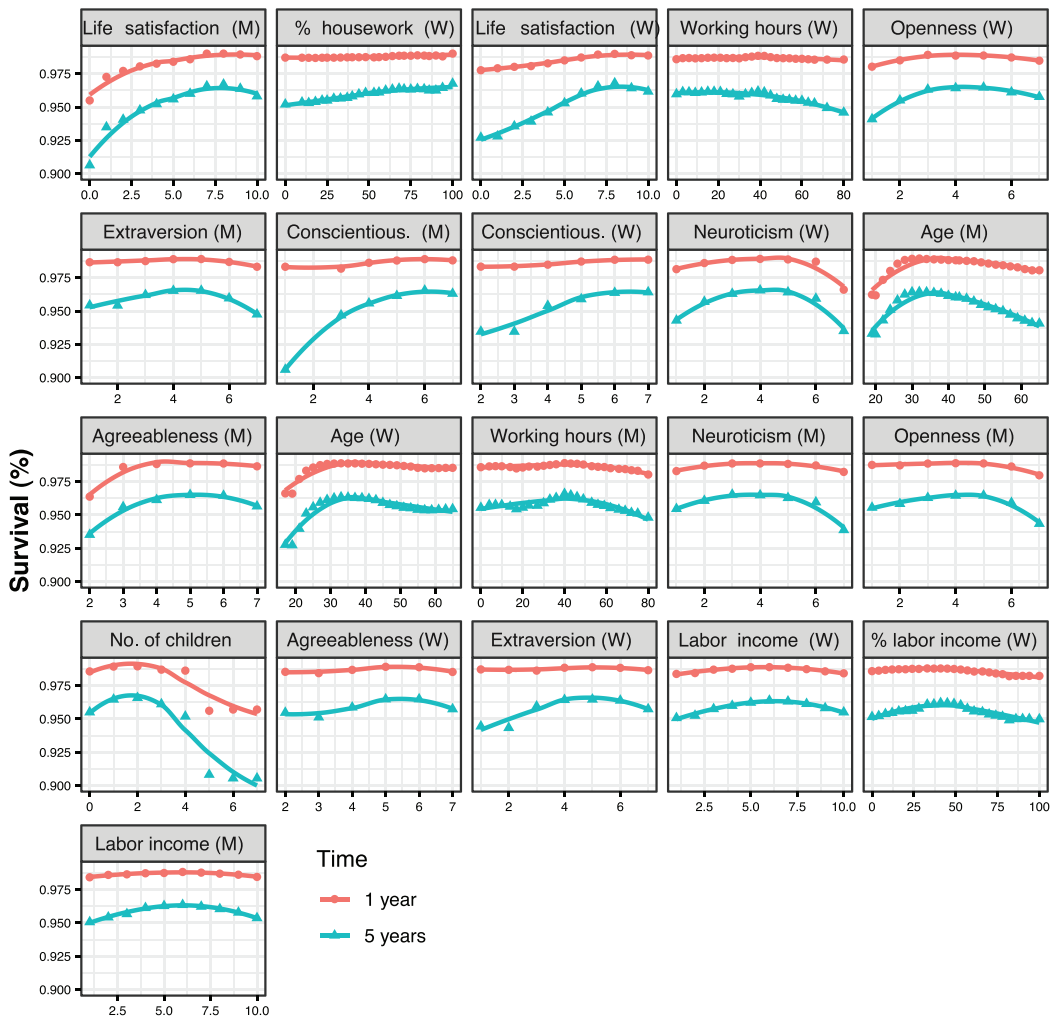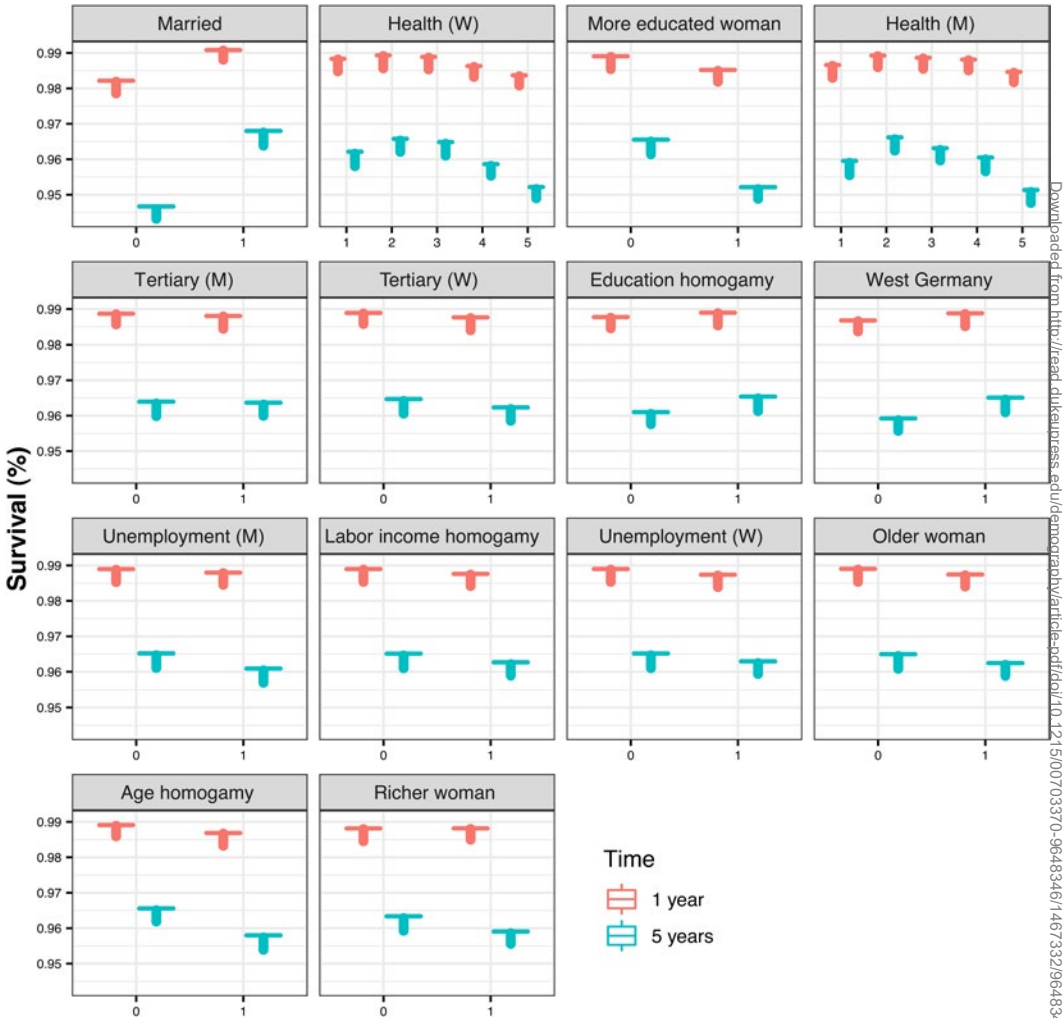
Fig. 2 Partial dependence plots of all continuous independent variables at one and five years. The *x*-axis shows the distribution of the predictor within our sample, while the *y*-axis shows the predicted survival associated with each value of the predictor. Conscientious.=conscientiousness.

association between the predictor and the outcome, and its value is not informative of the strength of the association between a predictor and the outcome.

Interpreting the Relationship Between Independent Variables and Union Dissolution

To gain insight into how each independent variable is related to union dissolution, we created partial dependence plots calculated at one and five years for all continuous (Figure 2) and categorical (Figure 3) predictors, ordered by their VIMP. Each point in both figures represents the average percentage of votes for the "yes" class (union

Fig. 3 Partial dependence plots of all categorical independent variables at one and five years. The *x*-axis shows the distribution of the predictor within our sample, while the *y*-axis shows the predicted survival associated with each value of the predictor.

dissolution) across all observations, given a fixed level of the predictor. Conceptually, this type of plot is similar to a graphical representation of the predicted survival probabilities as a function of a given variable.

Figure 2 illustrates the type and strength of relationship existing between each predictor and union survival probabilities. We found that both partners' levels of overall life satisfaction positively predicted survival: the higher the satisfaction, the higher the survival probability. The variation in survival probabilities, as a function of life satisfaction, was much stronger for couples whose union remained intact after five years. We also note a slight nonlinearity, as survival probabilities peaked when life satisfaction was at eight points and then declined.

CORRECTED PROOFS

From Figure 2 we recognize an important feature of tree-based approaches: their ability to automatically detect complex relationships. For example, survival probabilities varied with both the man's level of extraversion and his age according to nonlinear patterns that might be difficult to properly model with a parametric model. For example, a quadratic specification may only approximate the patterns displayed.

Figure 2 is also informative about how strong the variation in predicted survival probabilities was for different values of a given independent variable. We note a rather strong variation for some variables, such as woman's life satisfaction. In this case, we observe a variation in survival probabilities at five years of about four percentage points if we compare couples in which the woman reported the lowest as opposed to medium-high level of overall life satisfaction (i.e., of eight points).

Figure 3 shows the predicted survival probabilities for each value of each categorical variable, highlighting that RSF does not require grouping categories for this type of variable. We note that survival probabilities varied only slightly for the categorical predictors, with the exception of marriage status, education, and health of both partners. Married couples had substantively higher survival probabilities than cohabiting couples, at both one and five years. In contrast, after five years, couples in which the woman was more educated than the man were less likely to survive compared with other partners' educational pairings. Finally, survival probabilities, especially at five years' duration, seem to vary nonlinearly for partners' health, with the highest values reported for medium levels of health.

To investigate possible interactions among predictors, we calculated joint VIMP values starting from the seven most important predictors (according to the simple VIMP measure presented in Figure 1). The joint VIMP values are reported in Table A.1 in the online Appendix C. Among the interactions worth exploring was that between the woman's and the man's life satisfaction, which had both the highest difference between paired VIMP and additive VIMP and two of the highest univariate VIMP values.

We investigated the pairwise interactions between variables using partial dependence co-plots. In particular, Figures 4, 5, and 6 show the partial dependence co-plots with respect to the couple's survival probability at five years. Figure 4 shows how couples' survival probabilities varied as a function of life satisfaction of both the female and male partners. A complex nonlinear interaction pattern emerged. When man's life satisfaction was high, higher woman's life satisfaction (almost) monotonically increased the union's chances of surviving. But when man's life satisfaction was low, the association between woman's life satisfaction and union survival was negative after a given threshold (about 7.5).

According to the last column of Table A.1 in Appendix C, the interaction between woman's percentage of housework and her working hours was also of potential interest. Figure 5 shows how woman's percentage of housework and her working hours interacted in predicting union survival. The relationship between a woman's percentage of housework and the couple's survival probability is almost linear and positive for low levels of working hours, while it was nonlinear for higher levels of the latter. Unlike in Figure 4, here the distance between the curves varies more along the value of the predictor on the *x*-axis (i.e., woman's percentage of housework). This suggests the presence of a substantial combined effect (interaction) of the two variables on union survival.

Figure 6 shows that the man's level of life satisfaction did slightly interact with woman's working hours in predicting the likelihood of union dissolution. The survival
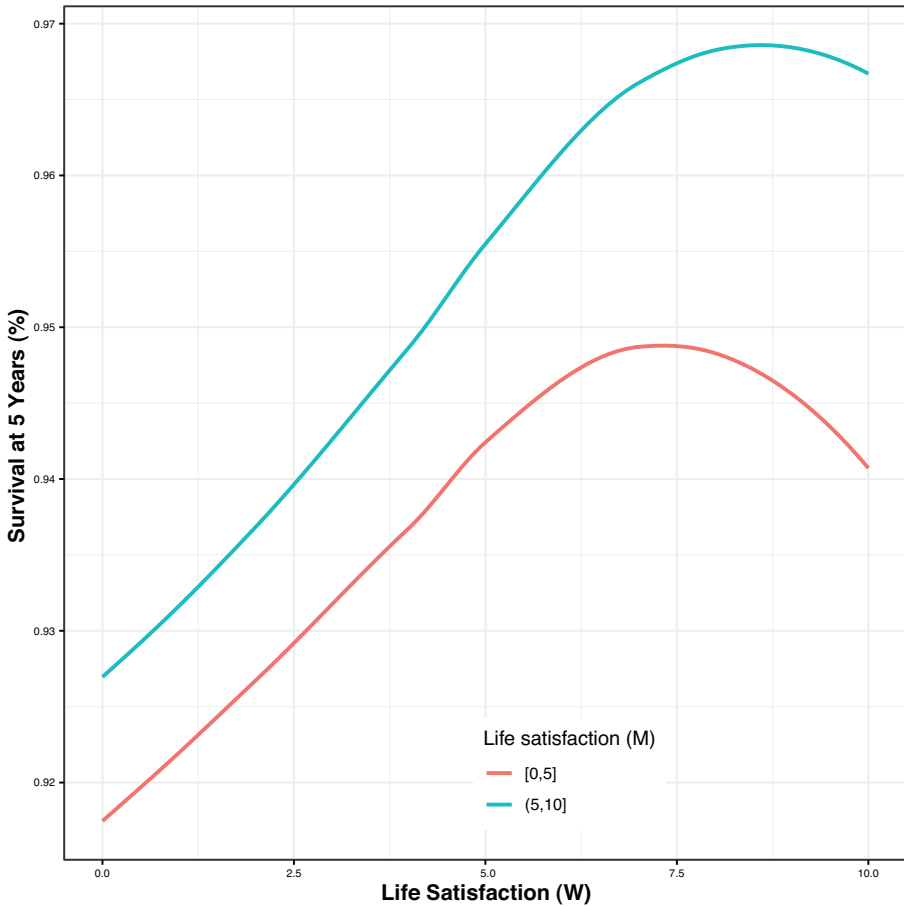
Fig. 4 Partial dependence co-plot: survival probabilities at five years of union duration as a function of a woman's and a man's life satisfaction. The *x*-axis shows the distribution of a woman's level of life satisfaction, while the *y*-axis shows, for each level of a man's life satisfaction, the predicted survival associated with each value of a woman's life satisfaction.

probability for both women who worked up to 36 hours and those who worked more than 36 hours was nonlinearly related to man's life satisfaction. The distance between the two curves tended to widen slightly as the man's life satisfaction increased, signaling that the latter may strengthen the negative association between number of hours women worked and the union's survival probability. Finally, we notice from Table A.3, as well as from additional partial dependence co-plots (available upon request), that none of the interactions between PTs of the same or different members of the couple appeared to be of substantial importance.

Revising the DT Model Using Insights From the RSF

The DT models estimated in the "Assessing the Algorithm's Performance" section could be modified in several ways. The models were not necessarily similar to those
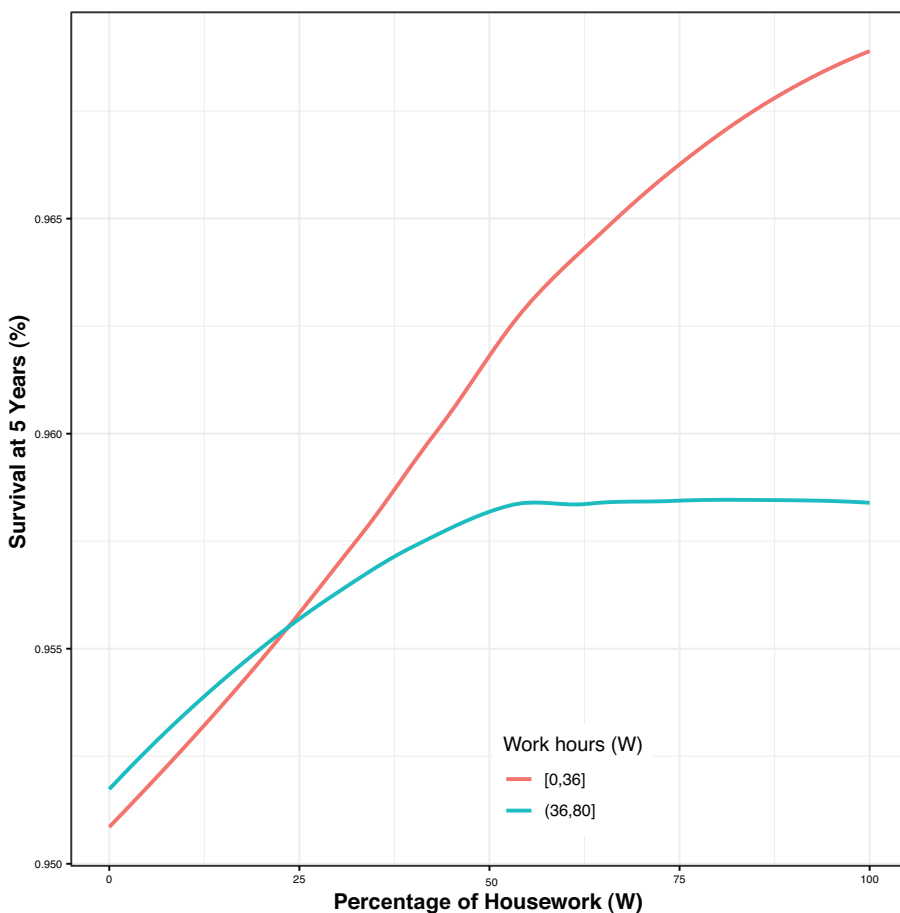
Fig. 5 Partial dependence co-plot: survival probabilities at five years of union duration as a function of a woman's percentage of housework and working hours. The *x*-axis shows the distribution of a woman's percentage of housework, while the *y*-axis shows, for each level of a woman's working hours, the predicted survival associated with each value of a woman's percentage of housework.

that would be commonly used in the framework of a traditional demographic study, in which the goal usually is to estimate the effects of specific independent variables to test a theory, net of a set of controls. Thus, we could select a smaller number of independent variables, which might reduce multicollinearity, though this is not an issue if prediction is the goal of the study. Also, the models might be made more flexible by allowing, for example, for specific nonlinearities. Finally, to improve accuracy, some variable selection approaches could be adopted. In the previous section we showed that RSF is able to pursue the goal of examining the complex pattern of links between a large set of predictors and union dissolution, while at the same time attaining a higher accuracy.

Here we consider a possible approach to improve the reference DT model shown in Table 3 (i.e., DT3) by leveraging insights from the RSF analysis. First, we excluded the variables showing a negative VIMP in Figure 1 (i.e., labor income of both partners,
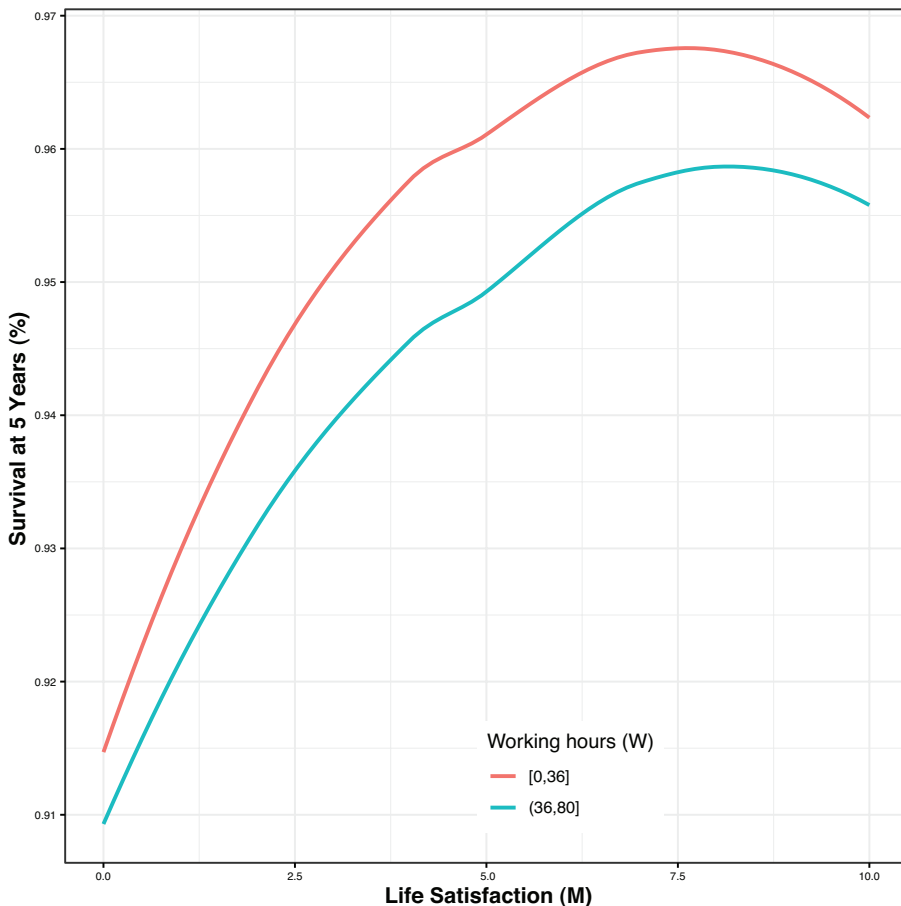
Fig. 6 Partial dependence co-plot: survival probabilities at five years of union duration as a function of a man's life satisfaction and a woman's working hours. The *x*-axis shows the distribution of a man's level of life satisfaction, while the *y*-axis shows, for each level of a woman's working hours, the predicted survival associated with each value of a man's life satisfaction

woman's percentage of labor income, labor income homogamy, woman's unemployment status, woman's extraversion and agreeableness, man's openness, number of children, and the dummy variables for age homogamy and for the woman being either older or richer than the man).

Second, we added a nonlinear term. Specifically, we added a square term for man's extraversion, which was among the seven most important variables, and whose impact on couple survival showed a general nonmonotonic nonlinearity (Figure 2). In particular, when man's level of extraversion was low but increasing, the probability that the couple broke up increased too, while the opposite was true at a medium-to-high level of man's extraversion.

The results of this revised version of model DT3 are presented in Tables A.4 and A.5 of the online appendix. We notice a general improvement of the model in terms of AIC (3,519), which is lower than in the original model (3,533). Additionally, the

coefficient of the new quadratic term for man's extraversion was statistically significant, signaling the importance of considering nonlinearities for such a predictor. Of course, other nonlinear terms suggested by the partial dependence co-plot could also be added. Finally, the predictive power (out-of-sample) of the model slightly improved for all possible survival times, although it remained smaller than that of the RSF.

Discussion and Conclusion

The aim of this article was to contribute to the literature on union dissolution using an ML technique, and in particular RSF applied to longitudinal SOEP data from Germany.

Results from RSF indicate that the most important predictors of union dissolution are both partners' life satisfaction. Previous studies on union dissolution have considered specific dimensions of life satisfaction, such as satisfaction with relationship quality. Our study points to the potential role of overall life satisfaction as a proximate determinant of union dissolution. Although RSF cannot generally identify causal effects, our findings suggest that future studies on union dissolution could benefit from devoting more attention to general life satisfaction, or to satisfaction with more domains.

Relatedly, RSF also detected an interesting and complex nonlinear interaction between partners' life satisfaction, such that woman's life satisfaction was positively associated with union survival when the male partner's well-being was high. However, when the male partner's life satisfaction was low, a high level of woman's life satisfaction was negatively associated with union survival. As with findings on fertility behaviors (Aassve et al. 2016), this result points to the need to consider both partners' subjective well-being in studying union survival.

Our analyses based on RSF were able to account for both partners' PTs and to explore their interactions. Interestingly, we found that, despite the importance of some specific PTs, none of the interactions between partner's PTs were relevant for predicting union survival. This finding should be analyzed in more depth in future studies, to assess, for example, the extent to which individuals form unions by matching specific combinations of PTs and whether assortative mating based on PTs might explain the lack of interactions.

As for the role of specific PTs, we found that the male partner's extraversion had an impact on predictive accuracy and substantial associations with union survival probabilities. This is in line with the few studies on union dissolution that accounted for PTs (Bortien et al. 2015; Bortien and Mortelmans 2018). Additionally, we found that the impact of the man's extraversion on union survival probability showed a nonmonotonic nonlinearity, which substantially affected the predicted probability of union dissolution.

From a methodological point of view, our analyses also showed that traditional discrete-time event-history models have very poor out-of-sample predictive accuracy, not substantially different from random guessing, while the conventional in-sample values were considerably higher. This indicates that previous studies may have suffered from overfitting. ML may suffer from overfitting as well; however, apart from

separating estimation from the evaluation of predictive accuracy, our analyses used RSF, which, being based on a multitude of trees, reduce risks of overfitting.

The RSF offered better predictive accuracy than DT models, confirming past research indicating that algorithmic approaches typically show better performance in this task. This is an important advantage of ML approaches because, as Hofman et al. (2017) argue, predictive accuracy is crucial for theory building and validation. However, the predictive accuracy of RSF remained limited, a finding that is in line with the results of a mass scientific collaboration that found low predictability in life outcomes (Salganik et al. 2020). In other words, union dissolution may be, in part, intrinsically "random" (Hofman et al. 2017), and thus, as in other social science contexts, we have to accept that a portion of phenomena might not be predictable. Additionally, studies have found that selection of variables based on theory or expert judgments does not necessarily improve predictive performance (Filippova et al. 2019). We used a large set of predictors identified in previous studies, but other factors that have not yet been studied might contribute to predicting union dissolution. Future applications of ML techniques could help identify these new predictors.

In conclusion, our results using RSF point to a complex interaction among partners' life satisfaction, and to the absence of interactions among partner's PTs, in predicting union dissolution. They also demonstrated which individuals' and couples' characteristics are highly predictive of union dissolution and which, instead, have weak or null predictive power.

A methodological contribution of this study has been to demonstrate the potential of ML techniques for the analysis of union dissolution and for demographic research more generally. We have shown that RSF are able to handle a large set of predictors, which proved very useful in the study of union dissolution. We also illustrated how RSF permitted exploring nonlinearities and nonadditivities in the links between predictors and union dissolution. Moreover, we found that RSF outperformed standard DT models in terms of out-of-sample predictive accuracy. Finally, we have demonstrated how insights from ML techniques can be integrated into more standard regression analyses.

Our implementation of RSF can be easily applied to different data and topics in demography by adapting the R code we provide. Cesare et al. (2018) noted that demographers have used some ML techniques in the analysis of digital trace data, such as for the (semi)automatic coding of tweets (Karamshuk et al. 2017). We showed how demographers might find ML techniques useful in the context of more "traditional" survey data too. Despite the promise of ML, however, our analyses demonstrate that artificial intelligence cannot replace human judgment entirely (Berk 2006; Lichtenthaler 2018). Machine learning is not a purely data-driven approach, just as regression-based research is not purely theory-driven (for further discussion, see Shu 2020). The use of ML techniques is not, as some have suggested, the "end of theory" (Anderson 2008). Key aspects of the definition, the measurement, and the selection of the variables to be analyzed remain with the researcher. Similarly, the interpretation of empirical results, and their contextualization within the broader literature and with respect to the study setting and the historical period, necessitate decisive human inputs. By combining subject-matter expertise with automatic and semiautomatic computational methods (see, e.g., Blei and Smyth 2017), demographers will be able to leverage the benefits of both the human and the machine.

References

Aassve, A., Arpino, B., & Balbo, N. (2016). It takes two to tango: Couples' happiness and childbearing. *European Journal of Population, 32,* 339–354.

Anderson, C. (2008, June 23). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. Retrieved from http://www.wired.com/2008/06/pb-theory

Andersson, G. (2002). Children's experience of family disruption and family formation: Evidence from 16 FFS countries. *Demographic Research, 7,* 343–364. https://doi.org/10.4054/DemRes.2002.7.7

Andersson, G. (2003). *Dissolution of unions in Europe: A comparative overview* (MPIDR Working Paper No. WP-2003-004). Rostock, Germany: Max Planck Institute for Demographic Research.

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives, 31*(2), 3–32.

Bellani, D., & Esping-Andersen, G. (2020). Gendered time allocation and divorce: A longitudinal analysis of German and American couples. *Family Relations, 69,* 207–226.

Bellani, D., Esping-Andersen, G., & Pessin, L. (2018). When equity matters for marital stability: Comparing German and U.S. couples. *Journal of Social and Personal Relationships, 35,* 1273–1298.

Berk, R. A. (2006). An introduction to ensemble methods for data analysis. *Sociological Methods & Research, 34,* 263–295.

Berk, R. A. (2016). *Statistical learning from a regression perspective* (2nd ed.). Cham, Switzerland: Springer Nature.

Berrington, A., & Diamond, I. (1999). Marital dissolution among the 1958 British birth cohort: The role of cohabitation. *Population Studies, 53,* 19–38.

Billari, F. C., Fürnkranz, J., & Prskawetz, A. (2006). Timing, sequencing, and quantum of life course events: A machine learning approach. *European Journal of Population, 22,* 37–65.

Bitew, F. H., Nyarko, S. H., Potter, L., & Sparks, C. S. (2020). Machine learning approach for predicting under-five mortality determinants in Ethiopia: Evidence from the 2016 Ethiopian Demographic and Health Survey. *Genus, 76,* 37. https://doi.org/10.1186/s41118-020-00106-2

Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences, 114,* 8689–8692.

Blossfeld, G. J. (2014, May). *Educational assortative mating and divorce: A longitudinal analysis of the influences of education on the divorce rate for different educational matches.* Paper presented at the annual meeting of the Population Association of America, Boston, MA.

Blossfeld, H.-P., & Müller, R. (2002). Union disruption in comparative perspective: The role of assortative partner choice and careers of couples. *International Journal of Sociology, 32,* 3–35.

Boertien, D., & Mortelmans, D. (2018). Does the relationship between personality and divorce change over time? A cross-country comparison of marriage cohorts. *Acta Sociologica, 61,* 300–316.

Boertien, D., von Scheve, C., & Park, M. (2015). Can personality explain the educational gradient in divorce? Evidence from a nationally representative panel survey. *Journal of Family Issues, 38,* 1339–1362.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24,* 123–140.

Breiman, L. (2001a). Statistical modeling: Two cultures (with discussion). *Statistical Science, 16,* 199–231.

Breiman, L. (2001b). Random forests. *Machine Learning, 45,* 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees.* Boca Raton, FL: Chapman & Hall/CRC.

Cannas, M., & Arpino, B. (2019). Comparison of machine learning algorithms and covariate balance measures in propensity score matching and weighting. *Biometrical Journal, 61,* 1049–1072.

Cesare, N., Lee, H., McCormick, T., Spiro, E., & Zagheni, E. (2018). Promises and pitfalls of using digital traces for demographic research. *Demography, 55,* 1979–1999.

Chan, T. W., & Halpin, B. (2002). Union dissolution in the United Kingdom. *International Journal of Sociology, 32,* 76–93.

Cooke, L. P. (2004). The gendered division of labor and family outcomes in Germany. *Journal of Marriage and Family, 66,* 1246–1259.

Cooke, L. P. (2006). "Doing" gender in context: Household bargaining and risk of divorce in Germany and the United States. *American Journal of Sociology, 112,* 442–472.

Cooke, L. P., Erola, J., Evertsson, M., Gahler, M., Härkönen, J., Hewitt, B., . . . Trappe, H. (2013). Labor and love: Wives' employment and divorce risk in its socio-political context. *Social Politics: International Studies in Gender, State & Society, 20,* 482–509.

Cooke, L. P., & Gash, V. (2010). Wives' part-time employment and marital stability in Great Britain, West Germany and the United States. *Sociology, 44,* 1091–1108.

De Rose, A. (1992). Socio-economic factors and family size as determinants of marital dissolution in Italy. *European Sociological Review, 8,* 71–91.

De Rose, A., & Pallara, A. (1997). Survival trees: An alternative non-parametric multivariate technique for life history analysis. *European Journal of Population, 13,* 223–241.

Emery, R. E. (Ed.). (2013). *Cultural sociology of divorce: An encyclopedia*. Thousand Oaks, CA: Sage Publications.

Filippova, A., Gilroy, C., Kashyap, R., Kirchner, A., Morgan, A. C., Polimis, K., . . . Wang, T. (2019). Humans in the loop: Incorporating expert and crowd-sourced knowledge for predictions using survey data. *Socius, 5*. https://doi.org/10.1177/2378023118820157

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29,* 1189–1232.

Frisco, M. L., & Williams, K. (2003). Perceived housework equity, marital happiness, and divorce in dual-earner households. *Journal of Family Issues, 24,* 51–73.

Gardner, J., & Oswald, A. J. (2006). Do divorcing couples become happier by breaking up? *Journal of the Royal Statistical Society: Series A, 169,* 319–336.

Glaeser, E. L., Hillis, A., Kominers, S. D., & Luca, M. (2016). Predictive cities crowdsourcing city government: Using tournaments to improve inspection accuracy. *American Economic Review: Papers & Proceedings, 106,* 114–118.

Goode, W. J. (1962). Marital satisfaction and instability: A cross-cultural class analysis of divorce rates. In R. Bendix & S. M. Lipset (Eds.), *Class, status, and power: Social stratification in comparative perspective* (pp. 377–387). New York, NY: Free Press.

Härkönen, J., & Dronkers, J. (2006). Stability and change in the educational gradient of divorce: A comparison of seventeen countries. *European Sociological Review, 22,* 501–517.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer Science+Business Media.

Hiekel, N., & Wagner, M. (2020). Individualized relationship practices and union dissolution: Differences between marriage and cohabitation. *European Sociological Review, 36,* 868–885.

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science, 355,* 486–488.

Hothorn, T., Buhlmann, P., Dudoit, S., Molinaro, A., & van der Laan, M. J. (2006). Survival ensembles. *Biostatistics, 7,* 355–373.

Hothorn, T., Lausen, B., Benner, A., & Radespiel-Tröger, M. (2004). Bagging survival trees. *Statistics in Medicine, 23,* 77–91.

Imdadullah, M., Aslam, M., & Altaf, S. (2016). mctest: An R package for detection of collinearity among regressors. *R Journal, 8*(2), 499–509.

Ishizuka, P. (2018). The economic foundations of cohabiting couples' union transitions. *Demography, 55,* 535–557.

Ishwaran, H. (2007). Variable importance in binary regression trees and forests. *Electronic Journal of Statistics, 1,* 519–537.

Ishwaran, H., & Kogalur, U. B. (2014). *Random forests for survival, regression and classification (RF-SRC),* R Package Version 1.6 [Software]. Retrieved from https://CRAN.R-project.org/package=randomForestSRC

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *Annals of Applied Statistics, 2,* 841–860.

Ishwaran, H., Kogalur, U. B., Chen, X., & Minn, A. J. (2011). Random survival forests for high-dimensional data. *Statistical Analysis and Data Mining, 4,* 115–132.

Jalovaara, M. (2003). The joint effects of marriage partners' socioeconomic positions on the risk of divorce. *Demography, 40,* 67–81.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. New York, NY: Springer Science+Business Media.

Kalmijn, M. (1998). Intermarriage and homogamy: Causes, patterns, trends. *Annual Review of Sociology, 24,* 395–421.

Kalmijn, M. (2003). Union disruption in the Netherlands: Opposing influences of task specialization and assortative mating? *International Journal of Sociology, 33,* 36–64.

Kalmijn, M. (2007). Explaining cross-national differences in marriage, cohabitation, and divorce in Europe, 1990–2000. *Population Studies, 61,* 243–263.

Kalmijn, M., Loeve, A., & Manting, D. (2007). Income dynamics in couples and the dissolution of marriage and cohabitation. *Demography, 44,* 159–179.

Karamshuk, D., Shaw, F., Brownlie, J., & Nishanth, S. (2017). Bridging big data and qualitative methods in the social sciences: A case study of Twitter responses to high profile deaths by suicide. *Online Social Networks and Media, 1,* 33–43.

Kraft, K., & Neimann, S. (2009). *Effect of labor division between wife and husband on the risk of divorce: Evidence from German data* (IZA Discussion Paper No. 4515). Bonn, Germany: Institute of Labor Economics.

Lichtenthaler, U. (2018). Substitute or synthesis: The interplay between human and artificial intelligence. *Research-Technology Management, 61*(5), 12–14.

Liefbroer, A. C., & Dourleijn, E. (2006). Unmarried cohabitation and union stability: Testing the role of diffusion using data from 16 European countries. *Demography, 43,* 203–221.

Lutz, W., Wils, A. B., & Nieminen, M. (1991). The demographic dimensions of divorce: The case of Finland. *Population Studies, 45,* 437–453.

Lyngstad, T. H., & Jalovaara, M. (2010). A review of the antecedents of union dissolution. *Demographic Research, 23,* 257–292. https://doi.org/10.4054/DemRes.2010.23.10

Mäenpää, E., & Jalovaara, M. (2014). Homogamy in socio-economic background and education, and the dissolution of cohabiting unions, *Demographic Research, 30,* 1769–1792. https://doi.org/10.4054/DemRes.2014.30.65

Matysiak, A., Styrc, M., & Vignoli, D. (2014). The educational gradient in marital disruption: A meta-analysis of European research findings. *Population Studies, 68,* 197–215.

Mencarini, L., & Vignoli, D. (2017). Employed women and marital union stability: It helps when men help. *Journal of Family Issues, 39,* 1348–1373.

Molina, M., & Garip, F. (2019). Machine learning for sociology. *Annual Review of Sociology, 45,* 27–45.

Mortelmans, D. (Ed.). (2020). *European studies of population series: Vol. 21. Divorce in Europe: New insights in trends, causes and consequences of relation break-ups*. Cham, Switzerland: Springer Nature.

Oláh, L. S. (2001). Gender and family stability: Dissolution of the first parental union in Sweden and Hungary. *Demographic Research, 4,* 29–96. https://doi.org/10.4054/DemRes.2001.4.2

Oláh, L. S., & Gahler, M. (2014). Gender equality perceptions, division of paid and unpaid work, and partnership dissolution in Sweden. *Social Forces, 93,* 571–594.

Oswald, A. J. (1997). Happiness and economic performance. *Economic Journal, 107,* 1815–1831.

Ozcan, B., & Breen, R. (2012). Marital instability and female labor supply. *Annual Review of Sociology, 38,* 463–481.

Poortman, A.-R. (2005a). How work affects divorce: The mediating role of financial and time pressures. *Journal of Family Issues, 26,* 168–195.

Poortman, A.-R. (2005b). Women's work and divorce: A matter of anticipation? A research note. *European Sociological Review, 21,* 301–309.

Røsand, G. M. B., Slinning, K., Røysamb, E., & Tambs, K. (2014). Relationship dissatisfaction and other risk factors for future relationship dissolution: A population-based study of 18,523 couples. *Social Psychiatry and Psychiatric Epidemiology, 49,* 109–119.

Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., . . . Datta, D. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences, 117,* 8398–8403.

Salvini, S., & Vignoli, D. (2011). Things change: Women's and men's marital disruption dynamics in Italy during a time of social transformations, 1970–2003. *Demographic Research, 24,* 145–174. https://doi.org/10.4054/DemRes.2011.24.5

CORRECTED PROOFS

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development, 3,* 210–229.

Shu, X. (2020). *Knowledge discovery in the social sciences: A data mining approach*. Oakland: University of California Press.

Sigle-Rushton, W. (2010). Men's unpaid work and divorce: Reassessing specialization and trade in British families. *Feminist Economics, 16*(2), 1–26.

Svarer, M. (2004). Is your love in vain? Another look at premarital cohabitation and divorce. *Journal of Human Resources, 39,* 523–535.

van Damme, M. (2020). The negative female educational gradient of union dissolution: Towards an explanation in six European countries. In D. Mortelmans (Ed.), *European studies of population series: Vol. 21. Divorce in Europe: New insights in trends, causes and consequences of relation break-ups* (pp. 93–122). Cham, Switzerland: Springer Nature.

Vignoli, D., & Ferro, I. (2009). Rising marital disruption in Italy and its correlates. *Demographic Research, 20,* 11–36. https://doi.org/10.4054/DemRes.2009.20.4

Vignoli, D., Matysiak, A., Styrc, M., & Tocchioni, V. (2018). The positive impact of women's employment on divorce: Context, selection, or anticipation? *Demographic Research, 38,* 1059–1110. https://doi.org/10.4054/DemRes.2018.38.37

Wagner, M., Schmid, L., & Weiß, B. (2015). Exploring increasing divorce rates in West Germany: Can we explain the iron law of increasing marriage instability? *European Sociological Review, 31,* 211–229.

Wagner, M., & Weiß, B. (2006). On the variation of divorce risks in Europe: Findings from a meta-analysis of European longitudinal studies. *European Sociological Review, 22,* 483–500.

Young, C. (2009). Model uncertainty in sociological research: An application to religion and economic growth. *American Sociological Review, 74,* 380–397.

Young, C., & Holsteen, K. (2017). Model uncertainty and robustness: A computational framework for multimodel analysis. *Sociological Methods & Research, 46,* 3–40.

Zhao, Q., & Hastie, T. (2021). Causal interpretations of black-box models. *Journal of Business & Economic Statistics, 39,* 272–281.

Bruno Arpino (corresponding author)
bruno.arpino@unifi.it

Arpino • Department of Statistics, Computer Science, Applications, University of Florence, Florence, Italy; https://orcid.org/0000-0002-8374-3066

Le Moglie • Department of Economics and Finance, Catholic University, Milan, Italy; https://orcid.org/0000-0003-3123-9818

Mencarini • Department of Social and Political Sciences, and Carlo F. Dondena Centre for Research on Social Dynamics and Public Policy, Bocconi University, Milan, Italy; https://orcid.org/0000-0001-5070-6069