Does Machine Learning Offer Added Value Vis-à-Vis Traditional Statistics? An Exploratory Study on Retirement Decisions Using Data from the Survey of Health, Ageing, and Retirement in Europe (SHARE)

Montserrat González Garibay 1, *, Andrej Srakar 1, Tjaša Bartolj 1 and Jože Sambt 2

1 Institute for Economic Research (IER), SI-1000 Ljubljana, Slovenia; sarakara@ier.si (A.S.); bartoljt@ier.si (T.B.)
2 School of Economics and Business, University of Ljubljana, SI-1000 Ljubljana, Slovenia; joze.sambt@ef.uni-lj.si
* Correspondence: gonzalezmm@ier.si; Tel.: +386-53030868

Abstract: Do machine learning algorithms perform better than statistical survival analysis when predicting retirement decisions? This exploratory article addresses the question by constructing a pseudo-panel with retirement data from the Survey of Health, Ageing, and Retirement in Europe (SHARE). The analysis consists of two methodological steps prompted by the nature of the data. First, a discrete Cox survival model of transitions to retirement with time-dependent covariates is compared to a Cox model without time-dependent covariates and a survival random forest. Second, the best performing model (Cox with time-dependent covariates) is compared to random forests adapted to time-dependent covariates by means of simulations. The results from the analysis do not clearly favor a single method; whereas machine learning algorithms have a stronger predictive power, the variables they use in their predictions do not necessarily display causal relationships with the outcome variable. Therefore, the two methods should be seen as complements rather than substitutes. In addition, simulations shed a new light on the role of some variables—such as education and health—in retirement decisions. This amounts to both substantive and methodological contributions to the literature on the modeling of retirement.

Keywords: machine learning; random forests; retirement; survival analysis; time-dependent covariates

1. Introduction

Do machine learning algorithms offer added value vis-à-vis traditional statistical methods for the prediction of retirement outcomes? This article compares several modeling techniques for retirement patterns in the light of this question. The issue of retirement has been extensively analyzed in the economic and sociological literature over the past several decades. Numerous theoretical and empirical models have been used to assess its likelihood and determinants. Gradually, the widespread availability of longitudinal data—whether from administrative databases or surveys—has opened up new research avenues. These data have been used for both basic and applied research.

Given the large variety of theoretical models having been developed, the research agenda has entered a phase of refinement. The research agenda proposed by Fisher et al. [1] identifies several possible avenues for future research: the incorporation of the influence of variables such as country, gender, socioeconomic status, and health status—and their interaction—on retirement; the critique of current theories; the exploitation of longitudinal data; the construction of a common language for the discipline; interdisciplinary and holistic work; and the incorporation of selection bias and heterogeneity in the analysis. Similarly, as Scharn et al. [2] point out, more research is needed on the differential impact of pension determinants on men and women, the effects of ongoing changes in the labor market, the interaction of different domains (e.g., demography, health, society), and the
impact of policies. Admittedly, as Fisher et al. [1] suggest, most of these issues can best be assessed by relying on complex databases, rich in the number of variables and observations.

The exploitation of large databases is also a recurring topic in the parallel debate on the use of machine learning for the study of social and economic subjects [3–7]. Even though machine learning techniques are used widely across several domains—such as medicine or marketing— their incorporation into economic and social research is still in its infancy. In addition, even though machine learning has been applied across economics and the social sciences, the number of studies comparing the performance of statistical methods with machine learning techniques is rather reduced. A search for these applications in Web of Science with the search query “TS = (“machine learning”) AND (TS = (sociology) OR TS = (“social sciences”) OR TS = (“social science”) OR TS = (“political economy”) OR TS = (“economics”) OR TS = (“econometrics”))”, refined to those records with Economics, Management, Business and Business Finance, Communication, Social Sciences Mathematical Methods, Social Sciences Interdisciplinary, Political Science, Sociology, Social Issues, International Relations, Development studies, Social Work, Demography, Family Studies, History of Social Sciences, Industrial Relations Labor, and Public Administration as their main labels, delivered 259 unique records, of which 140 were applications of machine learning methods to empirical subjects.

Machine learning algorithms have been compared sporadically to mainstream statistics in fields such as human capital [8], econometric models [9,10], asset management [11], monetary policy [12], the film industry [13], unemployment [14], transportation [15], experimental frameworks [16], infrastructure effects [17], health economics [18], audit decision making [19], food inflation [20], qualitative demography [21], the replicability of findings based on the European Social Survey [22], poverty predictions [23], and the forecasting of unemployment in the Euro area [24]. Their results do not point to any specific direction; there seem to be no generalizable findings about the worth of ML with respect to mainstream statistics, and no systematic comparisons have been carried out thus far. To our knowledge, there have been no articles to date using machine learning to address the retirement decision or process.

In this setting, the question arises as to whether the theoretical refinement of retirement studies mentioned above may be complemented with methodological refinements from the machine learning literature. We chose to address three of the research gaps identified by Fisher et al. [1]: the use of longitudinal databases, the need to incorporate more variables into the analysis of retirement, and the introduction of an interdisciplinary dimension.

Concretely, this article asks whether recent advances in machine learning techniques may contribute to the current research on retirement decisions. Though ubiquitous in contemporary computer science and routinely applied in other fields, machine learning is only starting to seep into economic research. This article aims at contributing to this incipient debate by comparing a machine learning algorithm (survival forests) with the current modeling of pension decisions, using simulation techniques. The analysis uses longitudinal data from six out of the seven waves of the Survey on Health, Ageing, and Retirement in Europe (SHARE) [25–31], and includes a large number of variables potentially bearing influence on the retirement process: country, age, gender, health, education, presence of grandchildren, and financial obligations (rent or mortgage payments). Here, a caveat is pertinent: this article’s aims are purely exploratory, given the fact that a single dataset is used; its aim is not to make any general claims about the suitability of machine learning over traditional techniques for the whole area of research, but rather to illuminate new ways to look at the topic of retirement in the light of its current methodological agenda.

In short, the article’s intended contribution to the literature is threefold:

(a) Theoretically, it attempts to shed new light on the retirement puzzle by addressing the interplay of newly identified factors influencing the decision to retire (cf. supra);
(b) Methodologically, it intends to explore a new technique for the study of large databases, and to compare it with the performance of existing techniques;
Empirically, it attempts to probe the actual impact of the newly identified variables in a longitudinal setting.

To answer the research question, Section 2 provides a review of the state of the art in retirement and machine learning research in light of the limitations mentioned above. Section 3 explains the methodological approach, and provides some information on the data to be used. Section 4 discusses the analysis’ results, and Section 5 lists a few concluding remarks.

2. State of the Art

The paradigmatic choices of this article are made in light of the mutual constraints posed by the theories and methodologies in the retirement literature, the available data, and the machine learning algorithms; Figure 1 portrays those choices.

The three panels of Figure 1 display a summary of the theoretical and methodological options available to analyze the retirement data from SHARE (cf. supra). The red rectangles indicate the choices made. First, the combination of the retirement literature with the available data (SHARE, cf. supra) poses the first constraint: the richness of the data makes it necessary to use substantive determinants of retirement rather than formal macroeconomic models to fully utilize the data’s potential. Second, the availability of longitudinal materials makes the SHARE data particularly suitable for event history analysis (EHA) or survival analysis. Third, among the vast landscape of machine learning algorithms, survival forests are chosen above other options. Within the survival forests algorithm, the R package most suited to deal with time-varying covariates is randomForestSRC, as it is the only random forest software developed to include survival analysis [32]. The argumentation for these paradigmatic choices is elaborated on further below.

2.1. Retirement

Following the OECD [33] and Dorn and Souza-Posa [34], retirement is defined as “having a self-described status of retired, regardless of employment status and receipt of a pension”. This definition should be seen in contrast with others based on income or employment status [33]. Retirement is furthermore assumed to be an absorbing state or an event, as opposed to a process [35]. The decision to retire has been studied from several
perspectives, using several statistical techniques. Here, the approaches are broadly divided along two strands: econometrics, and what here is termed “substantive determinants”, encompassing sociological, psychological, and health determinants.

In the first strand, econometrics and microeconomics have centered on the decision to retire from the perspective of the rational individual facing an optimization problem, by modeling concepts such as individual preferences, expected utility, and assets. Studies have followed the traditional economic debates [36] between structural models using concepts such as utility maximization [37–45], along with reduced-form models that address utility maximization functions indirectly by studying sociodemographic explanatory variables. Dahl et al. [46] analyzed the impacts of family situation, the presence of children, age, educational level, being a civil servant, income, spouse characteristics, and industry. Structural models also seem well suited to gauging the effects of specific policy interventions on behavior [46–51]. More recently, Hospido and Zamarro [47] investigated couples’ retirement patterns using the partner’s retirement, age, age difference with partner, education, the presence of children and grandchildren, health, and the partner’s health as the main explanatory variables, using a bivariate probit technique. Other examples of reduced-form models focus on policy discontinuities to assess the influence of policy changes on retirement patterns [49,50], on the impact of systemic factors such as the 2009 financial crisis [48], and on the relationship between financial incentives and retirement [51].

Stock and Wise, in their seminal [52] article outlining the option value model, contrasted the structural model with the reduced-form model and attempted to reconcile them. Their model combines both the structural and reduced-form approaches by adding the time variation (i.e., the individual re-evaluates their choices periodically); it can be used as purely structural (i.e., when all structural utility parameters are estimated) or in a reduced form (when those parameters are assumed).

The option value model has given rise to a sprawling literature; it is mostly used in its reduced form, although structural estimations have taken place as well; Lumsdaine, Stock, and Wise [53], as well as the contributors mentioned by Gruber and Wise [54], estimated the parameters either by regression or by a grid search. In this regard, it has been argued that the structural versions of the model “are rarely estimated successfully” [51]. Examples from the reduced form include Van Sonsbeek [55] and Mazzaferro and Morciano [56], who estimated models within the microsimulation literature. Hanappi et al. [57] and Borsch-Zupan [58] combined their own estimations for some parameters with assumptions for others. Other reduced-form applications of the model include those of Belloni and Alessie [59], Samwick and Wise [60], Samwick [61], and Berkel and Borsch-Zupan [62]. The reduced-form models mentioned above [49,51] are also inspired by Stock and Wise’s [52] option value framework.

In the second strand—situated within the fields of sociology, organizational psychology and health studies—countless studies have been conducted to assess the impact of several variables on retirement, without using a rational choice framework [2,63]. The classification of the variables influencing the decision to retire in review studies from this strand seems to have remained stable throughout the years. Feldman [64] identified four types of factors: individual history, opportunity structures in the career path, organizational factors, and external environment; he further divided those variables into push and pull factors, which either coerce or motivate workers to retire, and which remain widely used [2,35,65,66].

More recently, a similar classification of the determinants of retirement into four levels has been proposed by Topa et al. [63]: individual (income, financial security, and health), job (job satisfaction), work (organizational pressures, workplace timing for retirement, and job stress), and family (family pull); they conducted a meta-analysis of 151 articles dealing with these determinants. Fisher et al. [1] proposed a comprehensive conceptual model in which variables are divided into antecedents (e.g., macroeconomic factors, family, work-related and person-job factors, individual features) and moderators (e.g., context of retirement transition, retirement planning, form of retirement). Scharn et al. [2] defined 12 categories,
giving more attention to personal characteristics: demographic factors, health, lifestyle, social factors, social participation, work characteristics, job demands, contextual factors, financial factors, retirement preferences, macro effects, and birth cohort. Beyond these factors, other issues such as retirement planning have also been researched [67].

The empirical evidence for the existing models points to different directions. Topa et al.’s [63] meta-analysis mentioned above concludes that the best predictor of early retirement is the timing of retirement at the workplace, followed by organizational pressures. Other aspects (e.g., health, financial security, family pull) play a smaller, albeit significant role, even though some individual studies have noted large effects in particular contexts. Sundstrup et al. [68], for example, point in a recent publication to the importance of poor health and physical work demands in explaining early retirement in Denmark. This relates to Scharr et al.’s systematic review [2], which underlines the disparate findings of studies across countries, which may be related to the institutional regime at stake. This is further echoed by Boissonneault et al. [69], who conducted a systematic review of the factors explaining the recent increases in actual retirement ages across the OECD member states; they also pointed to institutional heterogeneity, and highlighted policy changes, education levels, and personal wealth as the factors driving the increase. The effects of policies have also been highlighted by Boot et al. [70].

Until recently, most of the available studies were based on cross-sectional analysis [71–75]. In recent years, however, much progress has been made in the use of longitudinal data. Throughout the past decade, more complex statistical techniques have been applied. Bockermann and Ilmakunnas [76] used probit models to study the impact of working conditions on early retirement. Trentini [77] used logit regressions to analyze SHARE data, and concluded that low levels of education and a stable career drive retirement patterns in Italy; in addition, he identified health as a potential factor driving involuntary retirement. SHARE data and event history analysis were used by Macken et al. to conclude that education plays a role in explaining involuntary labor market exit across Europe [78]. Likewise, Hagan et al. [79] used event history analysis to disentangle the causality of health status and retirement, and found that acute health shocks have a large effect on retirement decisions, possibly mediated by institutional dynamics. Using a stratified Cox regression, a cause-specific Cox model, and the model of Fine and Gray, van der Mark-Reeuwijk demonstrated that workers with poor health were more likely to leave the labor force than workers with good health [80]. By means of multilevel event history analysis, De Preter et al. confirmed the influence of bad health, as well as the presence of grandchildren, as factors speeding retirement [66]. The role of grandparenting was further explored and confirmed by Van Bavel and De Witter—also by means of multilevel event history analysis [81]. The same technique was used by De Breij et al. [82] to highlight the role of low education in explaining early exit from the labor market, and by Radl and Himmelreicher to gauge the influence of spousal behavior on retirement; the authors concluded that a partner’s early retirement pushes both men and women to retire as well, whereas the loss of a partner or a divorce have the opposite effect [83]. Bertogg et al. also used multilevel analysis to explore the interaction of gender norms and retirement patterns, and found that female main earners transition to retirement earlier than secondary earners [84]. Other techniques have also been applied. Using sequence analysis, Hoven et al. produced clusters of workers’ careers, and found that poor socioeconomic circumstances are related to early retirement [85]. Nonparametric estimations were used by Manoli and Weber, who concluded that financial incentives matter little to keep workers at work for longer [50]. Radl further applied piecewise exponential functions to explore the way in which class membership shapes retirement, along with other factors, and found that workers at both ends of the class continuum retire the latest [86]. Hofacker et al. investigated the likelihood of voluntary vs. involuntary retirement by means of multinomial logistic regressions across three countries (England, Germany, and Japan), and found national differences in the ways in which some factors, such as education levels, influence retirement choices [87].
2.2. Machine Learning

The main goal of the scientific field of machine learning—often interchangeably referred to as data mining—is find patterns in data. Following the recent econometrics literature, we opted to focus on the term “machine learning” and leave aside the theretofollowed “big data” and “data mining”. Big data is a concept lacking a uniform definition; it may, for instance, refer to large datasets [3,88], or to data fulfilling several criteria (e.g., volume, velocity, variety, exhaustiveness, resolution, relational nature, flexibility) [89]; due to this murkiness, it will not be discussed here further. Data mining is referred to as the automated process of discovering patterns in data, with a focus on structural patterns, whereas machine learning points to the concrete techniques within data mining that guide the discovery [3,90]. Hassani et al. [91] seem to see both terms as substitutes, and mention several machine learning techniques as applications of data mining. In the context of this paper, analyses are limited to supervised machine learning, understood as those techniques focusing “primarily on prediction problems: given a dataset with data on an outcome \( Y_i \) [such as retirement], which can be discrete or continuous, and some predictors \( X_i \), the goal is to estimate a model on a subset of the data, given the values of the predictors \( X_i \)” [92].

Unsupervised learning, which attempts to find patterns in the data without any guidance from the researcher [93], is beyond the scope of this article.

In spite of the large variety of machine learning approaches, their modus operandi is often uniform; it starts with partitioning data randomly into a training set—on the basis of which a prediction model is estimated (learner)—and a testing set, on which the predictions are tested [94]. Usually, several models are compared, and a choice is validated. Their performance is not only assessed by the accuracy with which they can predict results correctly in the test set, but by more complex criteria. Varian [3] highlights the fact that validation and testing are often combined. Moreover, regularization parameters can be used in order to improve the model’s performance. These parameters often look at one dimension of the machine learning model at stake and vary depending on the kind of algorithm used. Examples include the number of branches (depth) of a decision tree, the number of trees in a random forest, the number of levels in deep learning and neural networks, or the sum of squared beta coefficients in ridge regression (for a simplified overview of the technique and the different types of regularization functions, see Mullainathan and Spiess [93]; for a technical overview, see Hastie et al. [94]).

The assessment of the predictive performance and the regularization usually takes place empirically by applying k-fold validation techniques: the data are divided into k subsets labelled \( s = 1, \ldots, k \); subsequently, the model is fitted using k-1 subsets, predictions are made for subset \( s \), and the associated performance is measured. Then, the regularization parameter is chosen with the best average performance [3]. Regularization parameters may or not may be used at all, with some studies foregoing them [95].

Parameters are also used to measure algorithm and model performance, e.g., the receiver operating characteristic (ROC) graph and the area under the curve (AUC) derived from it, the Brier score, classification errors, the confusion matrix, the Gini coefficient, the kappa statistic, log loss, several variations of the mean errors (e.g., absolute, root, square, log), the Matthews correlation coefficient, and sensitivity and specificity indicators.

Within supervised learning, there are a vast number of available algorithms. Basic forms include “Naïve Bayes” probabilistic classification, “divide-and-conquer”, “separate-and-conquer”, linear and logistic regressions, mistake-driven algorithms, instance-based learning, and clustering. Ensemble methods build upon the basic algorithms and encompass, among others, bagging and boosting techniques, random forests (which combine the two former), and meta-models (“stacking” techniques). Deep learning algorithms lie at the basis of several types of neural networks [96].

Arguably, discussing all of the algorithms would prove impossible. Therefore, some heuristic criteria to define this review’s scope are defined here in light of the retirement literature and the use of longitudinal data referred to in the introduction.
First, the retirement-as-decision-making paradigm, with retirement being a puzzle that needs to be explained, makes it necessary to use algorithms equipped to deal with regression rather than classification problems (which excludes, for example, clustering). In addition, given the current research trends and the availability of longitudinal data, the machine learning algorithm used for the analysis should be adequate to process longitudinal data by means of event history analysis. Second, the use of the SHARE data imposes the further constraint of right censoring. Third, given the time-varying nature of the variables explaining retirement—such as health—the algorithm should be able to accommodate time-varying covariates. We deal with these issues in the following subsections.

2.3. Machine Learning and EHA

As has been pointed out before, the current literature on retirement makes extensive use of EHA, also called “survival analysis” or “hazard models”. Wang, Li, and Reddy [97] provide a comprehensive overview of both “classical” and machine learning EHA: survival trees, Bayesian methods, neural networks, support-vector machines, ensemble learning (e.g., random survival forests, bagging survival trees, and boosting), active learning, transfer learning, and multi-task learning. The latter four fall into the category of advanced machine learning, and according to the authors would be better at dealing with censored information.

Even though econometrics increasingly makes use of machine learning to study phenomena such as companies’ financial distress [98], bankruptcy [99], the microsimulation of pension expenses [100], and employee turnover [101], the field of retirement remains underexplored from the machine learning perspective. As an illustration, a quick search with the Web of Science search engine delivered only six results linked to the economics domain, none of which dealt with the decision to retire. One study among them [95], which used data from a longitudinal survey (SHARE’s US counterpart) to estimate the performance of machine learning algorithms vis-à-vis traditional statistic methods, had a similar setup to this paper’s; the authors compared the performance of different machine learning methods to statistical regressions when explaining several health indicators with social determinants. By contrast, a large body of knowledge on machine learning—and more specifically, on event history analysis—exists in other fields.

Using the bibliometrix R package and the biblioshiny interface, a short search for the main methods of analysis mentioned by Wang et al. [97] within the query results mentioned above rendered the results laid out in Table 1. It should be noted that decision trees were not looked at, given the fact that random forests constitute a more advanced application of the same technique.

| Biblioshiny Most Frequent Words Search Field | Author’s Keywords | Keywords Plus | Abstract |
|---------------------------------------------|-------------------|---------------|----------|
| Neural networks                             | 8                 | 19            | 27       |
| Forest                                      | 14                | 7             | 61       |
| SVM                                         | 7                 | 3             | 28       |
| Bayesian                                    | 8                 | 2             | 23       |
| Boosting                                    | 3                 | 0             | 15       |
| Bagging                                     | 0                 | 0             | 0        |

Given the relative abundance of random forest references over other techniques, the chances of mainstream solutions being created within this algorithm for the concrete problems presented by our data (time-varying covariates) are higher. Moreover, the classification techniques used by random forests (i.e., data splitting) are straightforward, and suited for the exploratory setup of this article.

The basic components of survival forests are survival trees [102], based on the classification and regression trees (CART) technique [103]. CART is a learning algorithm by which
decision trees are created. Decision trees may be roughly defined as the construction of groups of observations (nodes) according to their value on a certain outcome variable; this is done by the successive partition of data along a set of explanatory variables, according to some splitting criterion, such as the sum of squared deviations from the mean. The partition results in a final classification [102].

Survival trees [102] build on the general principles of decision trees and CART, but split the data according to the survival function. In other words, the observations grouped together share a similar survival function. The calculation of the survival functions may be carried out using known techniques, such as exponential or Cox regression models.

Survival and decision trees, in turn, serve as the basis for random forests, in which bootstrap samples are drawn from the data and trees are grown using them. At each node, a random number of explanatory variables is selected. The final outcome is predicted by “averaging the predictions from each individual tree” [102]. The random survival forest (RSF) algorithm is implemented in the randomForestSRC R package [104].

As has been mentioned above for machine learning in general, survival forests have been only sparingly applied to the economic literature [23]. There are, however, a growing number of articles in the medical literature that compare survival forests with mainstream statistical methods. Using the search query TOPIC: (“survival forests” “Cox regression”), a non-exhaustive search was performed on 29 June 2021 on the Web of Science Core Collection; it rendered 32 results, of which the subject of 7 was a comparison of Cox regressions with survival forests. One study was a systematic review comparing the results of machine learning and conventional statistical methods for the prediction of heart failure [105]. A summary of the studies assessing individual models is provided below in Table 2.

### Table 2. Overview of studies comparing Cox regressions and random survival forests.

| Source | Topic                              | Comparison Criterion for Model Performance | Simulations | Best Performance                      |
|--------|------------------------------------|---------------------------------------------|-------------|---------------------------------------|
| [106]  | Cancer survival                    | Harrell’s concordance index (C-index)       | No          | Cox                                   |
| [107]  | Osteoporotic fractures             | C-index                                     | No          | Cox                                   |
| [108]  | Kidney transplants                 | Integrated prediction error curve (IPEC)    | Yes         | Random survival forests               |
| [109]  | Cancer survival                    | C-index                                     | No          | Machine-learning-based models         |
| [110]  | HIV treatment                      | C-index                                     | No          | Random survival forests               |
| [111]  | Bankruptcy                         | C-index                                     | No          | Random survival forests               |
| [112]  | Congenital heart channelopathy     | Brier score                                 | No          | Random survival forests               |

Table 2 above paints a mixed picture; the findings do not favor one model above the other. This coincides with the findings of the systematic review mentioned above [105]. Moreover, it is clear that they mostly rely on the comparison of a single model, and no simulations are carried out. It is also remarkable that all of the articles, with the exception of one, are from the field of medicine.

### 2.4. Random Forests and Retirement Data

The retirement conundrum poses two problems for the use of random forests: right censoring, and time-varying covariates. A censored observation implies that the event
of interest might occur in the future (right censoring) or has occurred in the past (left censoring), but is not observed. In the case of retirement, right censoring means that a subject leaves the observation period without having gone into retirement.

A popular approach to right censoring is to estimate a distribution of censoring by applying the inverse probability of censoring weighting (IPCW) approach [113], which estimates a separate risk function for censoring, defines censoring weights based on that function, and applies the weights to the subsequent analysis in the dataset. Other approaches have been developed [114,115], but their application and incorporation into existing random survival forest software remains limited [116].

Time-varying covariates [117,118] are variables whose values may vary in time—as opposed to invariant or near-invariant variables, such as gender. Examples include the health status of near-pensioners, or the number of children living at home. The analysis of time-varying covariates has been applied only partially to survival forests. Moradian [119] compiled the existing approaches and focused on three methods: Bou-Hamad et al. [99], which models discrete survival for time-invariant covariates; Bou-Hamad et al. [120], which allows for time-varying covariates; and Schmid et al. [121], which is similar to the second method. These methods are relatively simple to implement, as they involve the creation of “pseudo-subjects”, by which single observations are split along the lines of time periods or, in the case of Schmid et al. [121], the period itself is added as a covariate. The separate datasets are then used to build random forests without any survival elements. This approach entails that a single subject may end up “split” across two different nodes of a tree. Moradian [119] introduced variations on the three methods and carried out simulations using the randomForestSRC package.

3. Materials and Methods

To compare the event history analysis and random forests for the modeling of the retirement decision, we used data from waves 1, 2, 4, 5, 6, and 7 of SHARE [25]. We paid attention to two specific features of the retirement data: right censoring, and longitudinal covariates. These two features shape the two analytical steps. In the first step, the need to model time-dependent covariates was assessed by means of a comparison of survival analysis with time-dependent covariates, survival analysis without time-dependent covariates, and survival random forests without time-dependent covariates. In the second step, the best performing model from the first step was compared to the machine learning techniques applied by Moradian [119] to the analysis of time-dependent covariates using simulations.

3.1. The Data

SHARE is an EU-wide panel study encompassing more than 120,000 individuals aged 50+, currently covering 27 European countries and Israel. To date, seven waves of the study have been conducted: 2004, 2006–2007, 2008–2009, 2010–2011, 2013, 2015, and 2017. Information is collected via computer-assisted personal interview (CAPI) questionnaires, physical measurements, and fill-in questionnaires across several modules containing demographic, health, psychological, and socioeconomic information [25]. The 2013 wave (SHARELIFE) adopted a retrospective perspective and recorded life histories, and is therefore left out of this study.

Using waves 1, 2, 4, 5, 6, and 7 of the SHARE data, a dataset was constructed to track the retirement decisions. The study analyzed only those respondents who were working in wave 1 (according to question ep005_), and whose countries participated in all waves taken into consideration. Observations with missing values in their employment status were removed. This produced a dataset of 2882 respondents. A person–period dataset was subsequently constructed, in which each period or interval is represented by a row (since SHARELIFE was not used, moments 3, 4, 5, and 6 correspond to waves 4, 5, 6, and 7, respectively). The retirement status of the respondents was measured at the end of each interval (for the 1–2 interval, the retirement status was measured at moment 2).
In the second step, covariates were added to the first dataset based on the theoretical insights from existing studies—mainly the study by Scharn et al. [2]—and the availability of the data within SHARE: age at moment 1, gender, country, self-perceived health status (sphus), number of grandchildren as a binary indicator (grchild_bin), a self-constructed binary indicator measuring whether the respondent paid rent or mortgage (rent_or_mortgage), and the number of years in education (yedu). The time-dependent covariates (sphus, ngrchild, and rent_or_mortgage) were measured at the beginning of each time interval.

Initially, other variables possibly influencing the retirement decision were added to the dataset—number of children, reading score, writing score, fluency score, numeracy score, the working status of the partner (for those respondents whose partner was in the survey), the effective retirement age by country (retrieved from the OECD database), the main job’s industry and tenure in 2009 (retrieved from SHARE’s third wave), the income decile (calculated by country), mortgage payments, and payment of rent. However, a quick analysis using Cox regressions made clear that the contribution of most variables to the model (as measured by the likelihood ratio test) was none, so they were removed. The number of children and the effective retirement age were the only significant variables, but were kept out so as to avoid collinearity with the number of grandchildren and the country variable, respectively. After all, the country variable accounts for the institutional heterogeneity across countries. In the case of mortgage payments and payment of rent, which were significant individually, the variables were synthetized in a single binary indicator in order to avoid the so-called “spikes at zero”, in which the distribution is affected by the large percentage of variables with zero value and, therefore, the coefficients may become biased. Even though the binarization of the variable implies a loss of information, the application of specific techniques to tackle the spikes at zero would have deviated from the main focus of the model. The number of grandchildren was binarized following the same logic.

Table 3 presents the descriptive statistics for the variables. A short description of the variables is provided in Appendix A. For the time-dependent covariates, only the first measurement is provided.

Table 3. Descriptive analysis of covariates, by survival status at the end of the observation period (N = 2882).

| Covariate | Censored (N = 884) | Retired (N = 1998) | Total (N = 2882) |
|-----------|--------------------|--------------------|-----------------|
| Gender    |                    |                    |                 |
| Male      | 437 (49.4%)        | 1111 (55.6%)       | 1548 (53.7%)    |
| Female    | 447 (50.6%)        | 887 (44.4%)        | 1334 (46.3%)    |
| Country   |                    |                    |                 |
| Austria   | 12 (1.4%)          | 101 (5.1%)         | 113 (3.9%)      |
| Germany   | 92 (10.4%)         | 196 (9.8%)         | 288 (10.0%)     |
| Sweden    | 145 (16.4%)        | 358 (17.9%)        | 503 (17.5%)     |
| Netherlands | 165 (18.7%)      | 156 (7.8%)         | 321 (11.1%)     |
| Spain     | 33 (3.7%)          | 109 (5.5%)         | 142 (4.9%)      |
| Italy     | 39 (4.4%)          | 166 (8.3%)         | 205 (7.1%)      |
| France    | 95 (10.7%)         | 252 (12.6%)        | 347 (12.0%)     |
| Denmark   | 110 (12.4%)        | 214 (10.7%)        | 324 (11.2%)     |
| Switzerland | 69 (7.8%)        | 124 (6.2%)         | 193 (6.7%)      |
| Belgium   | 124 (14.0%)        | 322 (16.1%)        | 446 (15.5%)     |
Table 3. Cont.

| Covariate                  | Censored (N = 884) | Retired (N = 1998) | Total (N = 2882) |
|----------------------------|--------------------|--------------------|------------------|
| Self-perceived health status in 2004 |                    |                    |                  |
| Poor                      | 9 (1.0%)           | 29 (1.5%)          | 38 (1.3%)        |
| Fair                      | 72 (8.1%)          | 174 (8.7%)         | 246 (8.5%)       |
| Good                      | 310 (35.1%)        | 800 (40.0%)        | 1110 (38.5%)     |
| Very good                 | 278 (31.4%)        | 587 (29.4%)        | 865 (30.0%)      |
| Excellent                 | 215 (24.3%)        | 408 (20.4%)        | 623 (21.6%)      |
| Exit wave                 |                    |                    |                  |
| Mean (SD)                 | 3.85 (1.63)        | 3.35 (1.23)        | 3.50 (1.38)      |
| Median (Min, Max)         | 4.00 (2.00, 6.00)  | 3.00 (2.00, 6.00)  | 3.00 (2.00, 6.00) |
| Age in 2004               |                    |                    |                  |
| Mean (SD)                 | 52.3 (4.01)        | 56.9 (4.22)        | 55.5 (4.67)      |
| Median (Min, Max)         | 52.0 (38.0, 76.0)  | 57.0 (47.0, 82.0)  | 55.0 (38.0, 82.0) |
| Grandchildren in 2004     |                    |                    |                  |
| Yes                       | 669 (75.7%)        | 1084 (54.3%)       | 1753 (60.8%)     |
| No                        | 215 (24.3%)        | 914 (45.7%)        | 1129 (39.2%)     |
| Paid rent or mortgage in 2004 |                   |                    |                  |
| No                        | 231 (26.1%)        | 828 (41.4%)        | 1059 (36.7%)     |
| Yes                       | 653 (73.9%)        | 1170 (58.6%)       | 1823 (63.3%)     |
| Years in education in 2004 |                   |                    |                  |
| Mean (SD)                 | 12.6 (3.71)        | 12.0 (3.81)        | 12.2 (3.79)      |
| Median (Min, Max)         | 13.0 (0, 21.0)     | 12.0 (0, 21.0)     | 12.0 (0, 21.0)   |

From Table 3, some differences are clear in the distribution of the variables across the censored and retired groups. The majority of the retired group tend to be males. Across countries, larger proportions of Belgians, Italians, French, Spanish, and Austrians can be found among the retired, whereas the opposite is true for the Dutch, Germans, Swedish, Danish, and Swiss. There do not seem to be very large differences in health status, save for the higher proportion of those in excellent health among the censored group. The mean observation time of both categories and the years in education equally fluctuate around the same values, but the mean age is four years higher in the retired group. The censored group also tend to be less likely to have grandchildren, and more likely to pay rent or a mortgage.

3.2. Cox Regressions

The first analyses were conducted using Cox regressions. The Cox model is “equivalent to conditional logistic regression, with conditioning at times where events occur” [122]. The model is composed of two main elements: the baseline hazard, which is expressed as a function of time, and which can be seen as the functional equivalent of the intercept in a linear regression; and the exponentiated vector containing the linear predictors and their corresponding coefficients. This is called a semi-parametric model, as the baseline hazard function does not need to be estimated. Arguably, this is what makes the model rather popular in life sciences and, increasingly, in other disciplines. The generic form of the model is as follows:

$$h(t|X) = h_0e^{\beta X}$$  \hspace{1cm} (1)
In Equation (1), \( h(t) \) represents the hazard of an event occurring at time \( t \), \( h_0 \) is the baseline hazard, and \( e^{\beta X} \) is the exponentiated vector of coefficients \( \beta \) and predictors \( X \). The model’s results are therefore reported in terms of hazard ratios.

There are three assumptions underpinning the model: First, the proportional hazards assumption states that the ratio between the hazards of different categories within a variable should remain constant over time. Second, censoring of observations (i.e., observations that disappear from the data without experiencing the event of interest) is assumed to be random, but cannot be tested. Third, the continuous predictors are assumed to have a linear form [123].

Before applying the model to our data, the first and third assumptions were tested with a model defined by the following equation:

\[
    h(t) = h_0 e^{\beta_1 \text{gender} + \beta_2 \text{country} + \beta_3 \text{age} + \beta_4 \text{sphus} + \beta_5 \text{grchild_bin} + \beta_6 \text{rent_or_mortgage} + \beta_7 \text{yedu}} \tag{2}
\]

where the term \( h(t) \) stands for the hazard of retirement at time \( t \). This is equal to the baseline hazard \( h_0 \), multiplied by the exponentiated sum of the covariates’ values (gender, country, age, sphus, grchild_bin, rent_or_mortgage, yedu), multiplied by the \( \beta \) coefficients. The linearity assumption was met, but the Cox model’s proportional hazards assumption turned out to be problematic. The assumption was tested by means of the cox.zph option of R’s survival package [124], which produces a series of chi-squared tests that, if significant, point to violations of the assumption. In addition, the function includes a plot of the beta(t) coefficient, which should have a slope close to zero, and which allows for a graphical analysis.

The results of the test for the variables described in the previous section are provided in Table 4. After the first analysis, two variables (gender and age) displayed violations; in order to tackle the problem, they were stratified—in other words, a baseline hazard was calculated for each combination of the variable’s categories. In addition, the age variable was aggregated into four categories in order to ensure that each sample within the simulations (cf. infra) contained sufficient observations from each category. However, this also uncovered a significant violation for the country variable.

Table 4. Results of the test for the Cox proportional hazards assumption, SHARE data (N = 2882).

| Variable          | Before | After 1 | After 2 |
|-------------------|--------|---------|---------|
|                   | Chi-Squared | p       | Chi-Squared | p       | Chi-Squared | p       |
| country           | 12.8925    | 0.1675  | 18.1786  | 0.033   |
| gender            | 8.7456     | 0.0031  |          |         |
| age               | 42.5487    | <0.0001 |          |         |
| sphus             | 0.0466     | 0.8297  | 0.4119   | 0.521   | 0.38866    | 0.53    |
| grchild_bin       | 0.4912     | 0.4834  | 0.0931   | 0.760   | 0.02468    | 0.88    |
| yedu              | 0.6002     | 0.4385  | 0.0573   | 0.811   | 0.03024    | 0.86    |
| rent_or_mortgage  | 0.0159     | 0.8996  | 0.4463   | 0.504   | 0.00726    | 0.93    |
| Global            | 54.9196    | <0.0001 | 18.8434  | 0.128   | 0.52460    | 0.97    |

The above situation posed a problem: the stratification of all three variables implied the creation of 80 different strata, each with its own baseline hazard. Some of these strata would unavoidably lack observations while conducting the simulations (cf. infra). Therefore, we opted to stratify only the country and age variables. The former was needed to make abstraction of the institutional context, whereas the deviation of the beta(t) coefficient for the age variable was larger than in the case of gender, where the deviation was minimal.

It should be noted that, whereas this solution enables the use of the model for those variables that do not violate the assumption, the influence of the stratified vari-
ables on the outcome cannot be known, as they are “internalized” in the Cox regression’s baseline hazard.

The model’s basic function after the stratification is given as follows:

\[
h_{ij}(t) = h_{0ij}e^{\beta_1 \text{gender} + \beta_2 \text{sphus} + \beta_3 \text{grchild}_\text{bin} + \beta_4 \text{rent}_\text{or}_\text{mortgage} + \beta_5 \text{yedu}}
\]  

where \( h_{ij}(t) \) is the hazard function, \( h_{0ij} \) is the baseline hazard, and \( \sum_{i=1}^{\infty} \beta_i X_i \) are the exponentiated linear predictors. The \( i \) and \( j \) indices denote different baseline hazards for the different country and age strata, respectively (cf. supra).

Two types of Cox regression were carried out using the R package survival: First, the retirement event was modelled using the formula given above. In the second variation, the need to use time-dependent covariates, following the approach outlined by Therneau et al. [115], was probed. Herein, the hazard is dependent on the value of the time-dependent covariates at a certain point in time; algebraically, it can be represented as follows:

\[
h_{ij}(t) = h_{0ij}e^{\beta_1 \text{gender} + \beta_2 \text{sphus}(t) + \beta_3 \text{grchild}_\text{bin}(t) + \beta_4 \text{rent}_\text{or}_\text{mortgage}(t) + \beta_5 \text{yedu}}
\]  

In analogy to Equation (3), \( h_{ij}(t) \) is the hazard function, \( h_{0ij} \) is the baseline hazard, and \( \text{gender, sphus, grchild}_\text{bin, rent}_\text{or}_\text{mortgage, and yedu} \) are the predictors, multiplied by the respective \( \beta \) coefficients. The \( i \) and \( j \) indices denote different baseline hazards for the different country and age strata, respectively (cf. supra). The only change is that the hazard at a time \( t \) is no longer dependent only on a single value of the covariates, but also on the value of the time-dependent covariates at the same point.

For validation and comparison purposes, 500 simulations were carried out along the lines discussed below. For each simulation, prediction errors were calculated using the receiver operating characteristic (ROC) graph and the area under the curve (AUC), which is derived from the ROC. The ROC curve plots the cumulative percentage of false positives identified by an algorithm (sensitivity) against the cumulative percentage of true positives (specificity). If the curve overlaps with the diagonal, the model is no better at predicting outcomes than a random algorithm. The AUC amounts to the percentage of the graph’s surface situated under the ROC curve [125]. When used in this context, the ROC and AUC are always calculated with reference to a certain estimation horizon. When the ROC is calculated for survival models, it is calculated along the lines of incidental sensitivity and dynamic specificity (I–D curve) [126].

It should be noted here that the AUC is not the only available metric to compare model performance; both Harrell’s C-index and the AUC are commonly used in the literature for assessing the predictive performance of survival models (cf. supra Table 2). In addition, the Brier score [127] and Nagelkerke’s R\(^2\) [128] are used to compare the overall performance [122,129]. Here, the AUC was chosen above the C-index, as described by Blanche et al. [130], who provided a mathematical demonstration that the C-index does not always reach a maximum value for the model with the best predictions, whereas the AUC does not display this problem. Brier’s score was not used, as it requires a separate estimation of the censoring function [131–133], and Nagelkerke’s R was not used, as it maximizes the disadvantage of outcomes that do not maximize the predicted probability when this probability approaches 0 or 1 [122].

3.3. Machine Learning Algorithms

The approach towards machine learning algorithms followed the same broad guidelines as the regressions. In a first step, a random survival forest was grown without using time-dependent covariates, using the randomSurvivalForest package from R. The forests were created according to the following formula:

\[
\text{retired} = \text{gender} + \text{strata(\text{age})} + \text{strata(\text{country})} + \text{sphus} + \text{grchild}_\text{bin} + \text{rent}_\text{or}_\text{mortgage} + \text{yedu}
\]  

The above formula indicates that the random forests for the classification of the binary variable retired should use the variables gender, stratified age, stratified country, sphus, grchild_bin, rent_or_mortgage, and yedu.

Random survival forests do not deal with time-dependent covariates. Hence, an alternative approach was found in the techniques of Moradian (cf. supra), who compiled and tested several existing methods that use ordinary classification forests (i.e., without a time dimension), and introduced the survival dimension in two ways: by means of a prediction horizon, and by taking time periods into account as variables.

Moradian makes a distinction between two time dimensions in survival analysis. On the one hand, there is \( t \), or the moment at which covariates are measured, while on the other there is the moment in time (later than \( t \)) for which an outcome variable is being measured and predicted, i.e., the estimation horizon. For the subject matter of this article, the \((t,u)\) matrix looks as shown in Table 5.

**Table 5.** Estimation problems \((t,u)\) in Moradian [119].

| \( t \) | \( u \) |
|-------|-------|
| \( 1 \) | o     | o     | o     | o     | o     |
| \( 2 \) | o     | o     | o     | o     |       |
| \( 3 \) | o     | o     | o     |       |       |
| \( 4 \) | o     | o     |       |       |       |
| \( 5 \) | o     |       |       |       |       |

Source: adapted from Moradian [119].

At \( t = 1 \), given the information available at the moment, predictions can be made for \( u = [2,6] \) for those respondents who have survived up to \( u \).

The estimation problem posed by the combination of estimation horizons and measurement moments is modelled by Moradian in five different ways, of which three are included in the present article. First, the separate method creates an independent estimation for each \((t,u)\) combination; for instance, the combination \((2,4)\) encompasses all of those respondents surviving until at least the fourth wave, and contains information on covariates measured during the second wave. The event of interest (retirement transition) is measured at \( u \), i.e., the fourth wave. In terms of random forests, this method requires 15 forests—one for each \((t,u)\) combination.

In the second approach, which draws from the insights of Schmid et al. [121], all \( t \) are pooled for a given \( u \). For instance, for \( u = 4 \), all variables from \( t = [1,3] \) are used. In addition, the \( t \) itself is used as a variable. A single respondent has one line per \( t \), containing the status regarding retirement, the time at which the status was recorded, and the covariates’ values.

In the third approach, which borrows from landmark analysis, all \((t,u)\) combinations are stacked and analyzed in a single dataset, with \( u \) becoming a variable as well.

To compare Moradian’s three variants of random forest algorithms with the Cox regressions, a simulation strategy was devised, following the general principles outlined by Morris et al. [134]. The data-generating mechanism, given the large amount of observations available, was resampling with replacement, by which 500 independent samples were obtained using the SHARE data described above.

After careful consideration, the resampling strategy was chosen above random data generation due to the presence of very few measurement points, which made it difficult to assign a distribution to the survival and hazard functions, and to produce realistic random data from which the same insights could be drawn as from our data of interest. The samples contained 50% of all observations in the dataset (for datasets with more than one line per observation, the number of unique observations was taken as the basis). Within each sample, further sampling was conducted to separate test and training datasets at a ratio of 50%. The test data were further divided into five datasets along the \( u \) dimension: the first dataset contained all observations, the second one contained those observations surviving
until measuring moment 2 or beyond, etc. For each dataset, the corresponding method was applied to the training dataset, and linear predicted values were obtained for each \((t,u)\) combination using the five test datasets. The AUC was obtained for each of those combinations, and its distribution was compared across methods.

Across the simulations, we chose not to vary any of the regularization parameters used for random forests. In this, we followed the example of earlier comparative studies \([95,106–108,110,111]\). This was done for the sake of simplicity, and to keep the focus on the contrast between machine learning and the Cox regressions. The number of trees per forest was set at 100, after an initial run of all models using the training datasets. A plot of the out-of-bag (OOB) error rates in light of the number of trees using the plot.rfsrc option from the rfsrc package \([32,135]\) showed that the error rate reached a stable minimum for almost all of the models after the 80th iteration.

4. Results

The results from the first analytical step, in which the need for time-dependent covariates was tested, make clear that the performance of the Cox model including time-dependent covariates is clearly superior to both the Cox model without time-dependent covariates and the random forest. Table 6 provides the coefficients estimated from the Cox baseline models with time-dependent covariates, as described by Equation (4).

Table 6. Summary of Cox regression of retirement with time-dependent covariates.

|              | Hazard (95%CI) | \(p\)      |
|--------------|---------------|------------|
| gender       | 1.00 (0.92–1.1) | 0.8971     |
| sphus        | 0.90 (0.86–0.94) | <0.0001    |
| grchild_bin  | 1.28 (1.16–1.41) | 0.000000   |
| yedu         | 0.98 (0.97–0.99) | <0.01      |
| rent_or_mortgage | 0.82 (0.74–0.91) | <0.001     |

Likelihood ratio test = 71.17 on 5 df, \(p = < 0.0001; n = 7213,\) number of events = 1998.

The influence of the different variables occurs along the expected lines: the probability of experiencing a transition to retirement decreases with a better health status. Having grandchildren makes the probability rise again, as opposed to the payment of rent or mortgage; those respondents with financial obligations regarding real estate are 18% less likely to be retired than those without. Similarly, the more years a respondent has spent in education, the less likely they are to retire. Gender does not influence the outcome significantly.

A comparison of the performance of the Cox model with time-dependent covariates, a random survival forest, and a Cox model without time-dependent covariates—with the AUC as the criterion—is shown in Table 7. It is clear that taking into account changes in the variables across time does offer a certain advantage in terms of predictive power. The summary of the Cox model without time-dependent covariates is not provided here, as the coefficients are very similar to the first model’s.

Table 7. Area under the curve (AUC) for Cox models and RSF with and without time-dependent covariates (TDCs), by time intervals (\(N = 7213\)).

| Time | Cox—TDCs | Cox—No TDCs | RSF—No TDCs |
|------|-----------|-------------|--------------|
| \(T = 2\) | 0.5331 | 0.4985 | 0.5211 |
| \(T = 3\) | 0.6131 | 0.5556 | 0.5706 |
| \(T = 4\) | 0.6842 | 0.5463 | 0.5370 |
| \(T = 5\) | 0.5900 | 0.4390 | 0.4878 |
| \(T = 6\) | 0.5333 | 0.5030 | 0.5162 |
| Average | 0.5909 | 0.5084 | 0.5265 |
Across the board, the Cox regressions with TDCs cover a larger AUC than their counterparts. The Cox regressions without TDCs perform worst, whereas the RSF without TDCs do slightly better. In other words, the results justify the incorporation of a time dimension to the model.

In a second step, simulations were carried out. First, 500 simulations of the Cox regressions defined in Equation (4) (cf. supra) were conducted using the whole sample (as opposed to the partitioned test and training datasets) to check the variability of the coefficients vis-à-vis the single regression conducted above; their summarized results are provided in Table 8 below.

### Table 8. Results from simulated Cox regressions of retirement ($n = 500$).

| Variable      | Average Coefficient | Average p-Value | Median p-Value |
|---------------|---------------------|-----------------|----------------|
| gender        | 1.0009              | 0.6066          | 0.6131         |
| sphus         | -0.09056            | 0.0270          | 0.0053         |
| grchild_bin   | 1.2829              | 0.0034          | 0.0004         |
| yedu          | -0.9822             | 0.1119          | 0.4912         |
| rent_or_mortgage | -0.8223           | 0.0349          | 0.0093         |

The simulations confirm the findings from the regression conducted above, although the significance of the education variable disappears in the simulation, putting the results of the single regression somewhat into perspective.

Second, separate simulations were carried out in, which the Cox regressions were executed along the random forest algorithms following the formula given by Equation (4). For the latter, the individual influence of variables was assessed with the variable importance (VIMP), defined as the extent to which the random forest classification algorithm uses a variable when classifying a case in a certain category. Consequently, those variables that split a tree in the beginning of the recursive partition process (i.e., “higher”) are considered to be more important [104]. It should be noted that the use of strata in RSF did not affect the results. For the sake of brevity, the VIMP scores are omitted; they are available from the authors upon request.

From the analysis of the VIMP, no crystal-clear picture emerged across methods and estimation horizons, except for prominent roles for age and country, which scored first in nearly all forests. The other factors seem rather unstable.

Further insights can be derived from the comparison of the AUCs obtained from the simulated Cox regressions and the random forests, summarized in Figures 2–6. Each figure corresponds to a row from Table 5. Figure 2, for instance, represents the estimation problems when $t = 1$ and the estimation horizon $u$ goes from 2 to 6.

The Cox regressions provide the lowest prediction accuracy overall, and have the highest variance. The stacked method outperforms the rest when the prediction horizon $u$ is below 5, across all $t$. However, when $u$ is 5 or 6, the advantage vanishes, and the emerging picture is rather murky, with the stacked method still prevailing for a few cases. The separate method seems to be the second most accurate. It should also be noted that the variability of all four methods increases as the prediction horizon widens.

However, a caveat is needed: Within each simulation, the test datasets are not independent from one another, as they are constructed by successively dropping the observations that do not survive beyond $(u - 1)$ (cf. supra). Consequently, for the Cox regressions and the stacked variant, the AUC remains constant for each $u$ across the $t$ dimension (see, for instance, the two last squares in the figure above), as the predictions derived from the training dataset are always the same.
Figure 2. AUC comparison of simulated Cox regressions and random forests (N = 500) when t = 1: (a) AUC comparison for t = 1, u = 2; (b) AUC comparison for t = 1, u = 3; (c) AUC comparison for t = 1, u = 4; (d) AUC comparison for t = 1, u = 5; (e) AUC comparison for t = 1, u = 6.

Figure 3. AUC comparison of simulated Cox regressions and random forests (N = 500) when t = 2: (a) AUC comparison for t = 2, u = 3; (b) AUC comparison for t = 2, u = 4; (c) AUC comparison for t = 2, u = 5; (d) AUC comparison for t = 2, u = 6.
5. Discussion

This article used random forest machine learning algorithms to address transitions to retirement, which have been in the past tackled with classical statistical methods. Using survey data from SHARE, a pseudo-panel was constructed, in which retirement transitions were tracked for 2882 respondents. Those decisions were further inserted into a Cox survival model and compared to several alternatives using simulations. From the models, we can draw several conclusions—both substantive and methodological.
When contrasted with the state of the art, this article’s findings are similar to those of other studies comparing machine learning and statistical methods (cf. supra, Table 2): Random forests seem to have an edge vis-à-vis Cox regressions in predictive performance. Likewise, the substantive findings on the decision to retire are broadly in line with the findings of those authors highlighting the impact of health on retirement decisions: poor health is linked to earlier retirement [66,79,136]. The role of grandchildren, as explored by Van Bavel and De Winter [81], is also confirmed: grandchildren are likely to give workers a motivation to retire. With regard to education, our findings differ from the existing literature, which underlines its importance [66,77,82]; conversely, we found no significant influence of the variable when carrying out simulations.

In the context of the above findings, the use of simulations raises an important caveat: As pointed out in Scharn et al.’s [2] systematic review (cf. supra), sociodemographic factors and health tend to play important roles in the decision to retire. However, the simulation approach makes clear that the influence of individual factors should not be taken for granted by using individual comparisons. De Preter et al. [66], who used the same survey in a similar context, concluded that grandchildren, health, and education level all influence the timing of retirement. Nevertheless, our simulation approach suggests that, after taking institutional differences into account, the education variable may bear a non-significant influence on the decision to retire; This highlights the usefulness of simulations—whereas in single-regression studies confidence intervals can be useful in discerning the influence of a covariate on a certain phenomenon, simulations complement this process. Their contribution should not be underestimated, especially as the availability of very large databases increases.

The addition of time-dependent covariates to the modelling of retirement decisions clearly increases the model’s predictive performance, albeit sparingly in this case. Moreover, random forests have some edge when compared to Cox regressions’ predictive capabilities, at least in the context of the SHARE data. There is, however, a considerable difference between the different random forest methods used, depending on the prediction horizon at stake. This also echoes Moradian’s [119] simulation results, albeit with a caveat: the stacked method’s performance improved for further prediction horizons in her approach; this implies that there is potential for the further development of random forest algorithms capable of accommodating the increasing complexity of survey and administrative data.

The explanatory power of random forests remains limited. The advantage of the random forests algorithm vis-à-vis the Cox regressions in terms of prediction, along with the limited information about each variable’s role provided by the VIMP, do not fully offset the detailed insights offered by the latter in terms of causality and the relative importance of different variables in explaining the outcome. All in all, when prediction is the central goal of the analytical process in question, machine learning algorithms are preferred. However, if causal processes are to be addressed as well, Cox regressions remain essential to the analysis. An important caveat here is that the use of stratification to tackle violations of the proportional hazards assumption may diminish the attractiveness of Cox regressions, as we lose the ability to look into the baseline hazard for the stratified variables. The use of simulations, widespread in the statistical and informatics literature, may provide further insights.

Simulations arguably constitute an innovative way to further the research agenda on economic and social topics such as retirement, but are certainly not the only one. Other elements of the machine learning literature that may enrich the discussion include the addition of more layers of complexity to the comparisons of “classical” statistical methods with machine learning, e.g., the variation of hyperparameters, the use of larger datasets and artificially generated data, robustness analyses across different datasets, or the use of different splitting rules for the construction of random forests. The extension of our type of analysis to other machine learning methods adequate for survival analysis (e.g., support-vector machines, artificial neural networks) offers another promising avenue of research.
This article started by posing the question of whether machine learning offers added value to the methods used traditionally for the study of retirement. The answer is definitely not clear-cut. Even though the machine learning algorithm we used seems to be better at predicting whether someone will retire or not, the fine-grained analysis offered by the statistical methods is lacking. In this sense, the methods may arguably be seen as complements rather than substitutes, with statistical methods helping to identify the factors driving phenomena, and machine learning helping to forecast future developments based on those factors.

**Author Contributions:** Conceptualization, M.G.G., A.S. and J.S.; methodology, M.G.G., A.S. and J.S.; software, M.G.G.; validation, M.G.G.; formal analysis, M.G.G.; investigation, M.G.G.; data curation, M.G.G.; writing—original draft preparation, M.G.G.; writing—review and editing, T.B., J.S. and A.S.; visualization, M.G.G.; supervision, A.S. and J.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the project “Upgrade of Analytical Models Concerning the Pension Scheme”, financed by the European Social Fund and the Slovenian Ministry of Labour, Family, Social Affairs, and Equal Opportunities. The SHARE data collection was funded by the European Commission, DG RTD through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646), and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782), as well as by DG Employment, Social Affairs and Inclusion through VS 2015/0195, VS 2016/0135, VS 2018/0285, VS 2019/0332, and VS 2020/0313. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the US National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C, RAG052527A), and from various national funding sources is gratefully acknowledged (see [www.share-project.org](http://www.share-project.org)).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** This paper uses data from SHARE waves 1, 2, 4, 5, 6, and 7 (DOIs: 10.6103/SHARE.w1.710, 10.6103/SHARE.w2.710, 10.6103/SHARE.w4.710, 10.6103/SHARE.w5.710, 10.6103/SHARE.w6.710, and 10.6103/SHARE.w7.711); see Börsch-Supan et al. [25] for methodological details.

**Acknowledgments:** The authors wish to thank Joost Bollens for his useful feedback and comments. All omissions or mistakes remain the authors’.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A

#### Table A1. Variable descriptions.

| Variable Name | Description | Variable Type | Values | Time Dimension |
|---------------|-------------|---------------|--------|----------------|
| country       | Country where the respondent resides | Nominal       | Austria  
Germany  
Sweden  
Netherlands  
Spain  
France  
Denmark  
Switzerland  
Belgium | None |
Table A1. Cont.

| Variable Name      | Description                                                                 | Variable Type | Values                        | Time Dimension |
|--------------------|------------------------------------------------------------------------------|---------------|-------------------------------|----------------|
| age                | Respondent’s age at the beginning of the measurements (2004)                 | Continuous    | 38–82                         | Not included in survey |
| gender             | Respondent’s gender                                                         | Binary        | 1 = Male, 2 = Female          | None           |
| sphus              | Self-perceived health status                                                 | Ordinal       | 1 = Poor, 2 = Fair, 3 = Good, 4 = Very good, 5 = Excellent | Yes            |
| grchild_bin        | Presence of grandchildren, constructed on the basis of ngrchild from the original survey data | Binary        | 0 = No grandchildren, 1 = One or more grandchildren | Yes            |
| rent_or_mortgage   | Constructed on the basis of the "Owner, tenant or rent free" (otrf) and "Mortgage" (mort) variables from the original SHARE data. | Binary        | 0 = Respondent pays no rent or mortgage, 1 = Respondent pays rent or mortgage | Yes            |
| yedu               | Number of years in education in 2004                                         | Continuous    | 0–21                          | No             |

References

1. Fisher, G.G.; Chaffee, D.S.; Sonnega, A. Retirement Timing: A Review and Recommendations for Future Research. *Work. Aging Retire.* 2016, 2, 230–261. [CrossRef]
2. Scharn, M.; Sewdas, R.; Boot, C.R.L.; Huisman, M.; Lindeboom, M.; Van Der Beek, A.J. Domains and Determinants of Retirement Timing: A Systematic Review of Longitudinal Studies. *BMC Public Health* 2018, 18, 1083. [CrossRef] [PubMed]
3. Varian, H.R. Big Data: New Tricks for Econometrics. *J. Econ. Perspect.* 2014, 28, 3–28. [CrossRef]
4. Athey, S. Machine Learning and Causal Inference for Policy Evaluation. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD’15, Sydney, Australia, 10–13 August 2015; pp. 5–6. [CrossRef] [PubMed]
5. Athey, S.; Imbens, G.W. Machine Learning Methods That Economists Should Know About. *Annu. Rev. Econ.* 2019, 11, 685–725. [CrossRef]
6. Hindman, M. Building Better Models: Prediction, Replication, and Machine Learning in the Social Sciences. *Ann. Am. Acad. Political Soc. Sci.* 2015, 659, 48–62. [CrossRef]
7. Boelaert, J.; Ollion, E. The Great Regression Machine Learning, Econometrics, and the Future of Quantitative Social Sciences. *Rev. Fr. Sociol.* 2018, 59, 475–506. [CrossRef]
8. MacLeod, W.B. Viewpoint: The Human Capital Approach to Inference. *Can. J. Econ.* 2017, 50, 5–39. [CrossRef]
9. Kim, D.; Shin, S. The Economic Explainability of Machine Learning and Standard Econometric Models—An Application to the US Mortgage Risk. *Int. J. Strateg. Prop. Manag.* 2021, 25, 396–412. [CrossRef]
10. Hansen, J.V.; Nelson, R.D. Forecasting and Recombining Time-Series Components by Using Neural Networks. *J. Oper. Res. Soc.* 2003, 54, 307–317. [CrossRef]
11. Cerniglia, J.A.; Fabozzi, F.T. Selecting Computational Models for Asset Management: Financial Econometrics versus Machine Learning—Is There a Conflict? *J. Portf. Manag.* 2020, 47, 107–118. [CrossRef]
12. Sofianos, E.; Gogas, P.; Papadimitriou, T. Mind the Gap: Forecasting Euro-Area Output Gaps with Machine Learning. *Appl. Econ. Lett.* 2021, 1–5. [CrossRef]
13. Liu, Y.; Xie, T. Machine Learning versus Econometrics: Prediction of Box Office. *Appl. Econ. Lett.* 2019, 26, 124–130. [CrossRef]
14. Gogas, P.; Papadimitriou, T.; Sofianos, E. Forecasting Unemployment in the Euro Area with Machine Learning. *J. Forecast.* 2021. [CrossRef]
15. Plakandaras, V.; Papadimitriou, T.; Gogas, P. Forecasting Transportation Demand for the US Market. *Transp. Res. Part A Policy Pract.* 2019, 126, 195–214. [CrossRef]
16. Ramsey, S.M.; Bergtold, J.S. Examining Inferences from Neural Network Estimators of Binary Choice Processes: Marginal Effects, and Willingness-to-Pay. *Comput. Econ.* 2021, 58, 1137–1165. [CrossRef]
17. Steinkraus, A. Estimating Treatment Effects With Artificial Neural Nets—A Comparison to Synthetic Control Method. *Econ. Bull.* 2019, *39*, 2778–2791.

18. Arora, P.; Boyne, D.; Slater, J.J.; Gupta, A.; Brenner, D.R.; Druzdzel, M.J. Bayesian Networks for Risk Prediction Using Real-World Data: A Tool for Precision Medicine. *Value Health* 2019, *22*, 439–445. [CrossRef]

19. Hansen, J.V.; McDonald, J.B.; Stice, J.D. Artificial Intelligence and Generalized Qualitative Response Models—An Empirical Test on 2 Audit Decision-Making Domains. *Decis. Sci.* 1992, *23*, 708–723. [CrossRef]

20. Malhotra, A. A Hybrid Econometric-Machine Learning Approach for Relative Importance Analysis: Prioritizing Food Policy. *Eurasian Econ. Res. Rev.* 2021, *11*, 549–581. [CrossRef]

21. Chakrabarti, P.; Frye, M. A Mixed-Methods Framework for Analyzing Text Data: Integrating Computational Techniques with Qualitative Methods in Demography. *Demogr. Res.* 2017, *37*, 1351–1382. [CrossRef]

22. Vijayakumar, R.; Cheung, M.W.-L. Assessing Replicability of Machine Learning Results: An Introduction to Methods on Predictive Accuracy in Social Sciences. *Soc. Sci. Comput. Rev.* 2021, *39*, 768–801. [CrossRef]

23. Sohnensen, T.P.; Stender, N. Is Random Forest a Superior Methodology for Predicting Poverty? An Empirical Assessment. *Poverty Public Policy* 2017, *9*, 118–133. [CrossRef]

24. Gogas, P.; Papadimitriou, T. Machine Learning in Economics and Finance. *Comput. Econ.* 2021, *57*, 1–4. [CrossRef]

25. Börsch-Supan, A.; Brandt, M.; Hunkler, C.; Kneip, T.; Korbmacher, J.; Malter, F.; Schaan, B.; Stuck, S.; Zuber, S. SHARE Central Coordination Team Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *Int. J. Epidemiol.* 2013, *42*, 992–1001. [CrossRef] [PubMed]

26. Börsch-Supan, A. *Survey of Health, Ageing and Retirement in Europe (SHARE)* Wave 1; Release Version: 7.1.0; SHARE-ERIC: Munich, Germany, 2020.

27. Börsch-Supan, A. *Survey of Health, Ageing and Retirement in Europe (SHARE)* Wave 4; Release Version: 7.1.0; SHARE-ERIC: Munich, Germany, 2020.

28. Börsch-Supan, A. *Survey of Health, Ageing and Retirement in Europe (SHARE)* Wave 5; Release Version: 7.1.0; SHARE-ERIC: Munich, Germany, 2020.

29. Börsch-Supan, A. *Survey of Health, Ageing and Retirement in Europe (SHARE)* Wave 6; Release Version: 7.1.0; SHARE-ERIC: Munich, Germany, 2020.

30. Börsch-Supan, A. *Survey of Health, Ageing and Retirement in Europe (SHARE)* Wave 7; Release Version: 7.1.0; SHARE-ERIC: Munich, Germany, 2020.

31. Börsch-Supan, A. *Survey of Health, Ageing and Retirement in Europe (SHARE)* Wave 2; Release Version: 7.1.0; SHARE-ERIC: Munich, Germany, 2020.

32. Ishwaran, H.; Kogalur, U.B. Package ‘RandomForestSRC’; R Foundation for Statistical Computing: Vienna, Austria, 2021.

33. OECD. *The Transition from Work to Retirement*; Social Policy Studies: Paris, France, 1995.

34. Dorn, D.; Sousa-Poza, A. *The Determinants of Early Retirement in Switzerland*. *Swiss J. Econ. Stat.* 2005, *141*, 247–283.

35. Wang, M.; Wanberg, C.R. 100 Years of Applied Psychology Research on Individual Careers: From Career Management to Retirement. *J. Appl. Psychol.* 2017, *102*, 546–563. [CrossRef]

36. Heckman, J.J.; Vytlacil, E. *Structural Equations, Treatment Effects and Econometric Policy Evaluation*; NBER Working Papers Series; NBER: Cambridge, MA, USA, 2005.

37. Gustman, A.L.; Steinmeier, T.L. *How Changes in Social Security Affect Recent Retirement Trends*; NBER Working Papers Series; NBER: Cambridge, MA, USA, 2008.

38. Berkovec, J.; Stern, S. *Job Exit Behavior of Older Men*. *Econometrica* 1991, *59*, 189–210. [CrossRef]

39. Gustman, A.L.; Steinmeier, T.L. *A Structural Retirement Model*; NBER Working Papers Series; NBER: Cambridge, MA, USA, 1983.

40. Gustman, A.L.; Steinmeier, T.L. *Retirement in a Family Context: A Structural Model for Husbands and Wives*; NBER Working Papers Series; NBER: Cambridge, MA, USA, 1994.

41. Gustman, A.L.; Steinmeier, T.L. Retirement in Dual-Career Families: A Structural Model. *J. Labor Econ.* 2000, *18*, 503–545. [CrossRef]

42. Gustman, A.L.; Steinmeier, T.L. *The Social Security Early Entitlement Age in a Structural Model of Retirement and Wealth*; NBER Working Papers; NBER: Cambridge, MA, USA, 2002.

43. Casanova, M. Happy Together: A Structural Model of Couples’ Joint Retirement Choices. Available online: *http://www.econ. ucla.edu/casanova/Files/Casanova_joint_ret.pdf* (accessed on 29 July 2019).

44. Rust, J.; Phelan, C. How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets. *Econometrica* 1997, *65*, 781–831. [CrossRef]

45. Laun, T.; Markussen, S.; Christian, T.; Wallenius, J. Journal of Economic Dynamics & Control, Longevity and Retirement Reform. *J. Econ. Dyn. Control*. 2019, *103*, 123–157. [CrossRef]

46. Dahl, S.-A.; Nilsen, O.A.; Vaage, K. *Work or Retirement? Exit Routs for Norwegian Elderly*; IZA Discussion Papers: Bonn, Germany, 1999.

47. Hospido, L.; Zamarron, G. *Retirement Patterns of Couples in Europe*. *IZA J. Eur. Labor Stud.* 2014, *3*, 12. [CrossRef]

48. Sirven, N.; Barnay, T. Expectations, Loss Aversion and Retirement Decisions in the Context of the 2009 Crisis in Europe. *Int. J. Manpow.* 2017, *38*, 25–44. [CrossRef]

49. Börsch-Supan, A.; Brandt, M.; Hunkler, C.; Kneip, T.; Korbmacher, J.; Malter, F.; Schaan, B.; Stuck, S.; Zuber, S. SHARE Central Coordination Team Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *Int. J. Epidemiol.* 2013, *42*, 992–1001. [CrossRef] [PubMed]
49. Manoli, D.; Mullen, K.J.; Wagner, M. Policy Variation, Labor Supply -Elasticities, and a Structural Model of Retirement. Econ. Inq. 2015, 53, 1702–1717. [CrossRef]
50. Manoli, D.S.; Weber, A. Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions; NBER Working Papers Series; NBER: Cambridge, MA, USA, 2011.
51. Asch, B.; Haider, S.J.; Zissimopoulos, J. Financial Incentives and Retirement: Evidence from Federal Civil Service Workers. J. Public Econ. 2005, 89, 427–440. [CrossRef]
52. Stock, B.Y.H.; Wise, D.A. Pensions, the Option Value of Work, and Retirement. Econometrica 1990, 58, 1151–1180. [CrossRef]
53. Lumdsdaine, R.L.; Stock, J.H.; Wise, D.A. Three Models of Retirement: Computational Complexity versus Predictive Validity. In Topics in the Economics of Aging; Wise, D.A., Ed.; University of Chicago Press: Chicago, IL, USA, 1992; pp. 21–60.
54. Gruber, J.; Wise, D.A. Social Security Programs and Retirement Around the World: Micro Estimation; NBER Working Papers Series; NBER: Cambridge, MA, USA, 2002.
55. van Sonsbeek, J.-M. Micro Simulations on the Effects of Ageing-Related Policy Measures: The Social Affairs Department of the Netherlands Ageing and Pensions Model. Int. J. Microsimul. 2011, 4, 72–99. [CrossRef]
56. Mazzufero, C.; Morciano, M. CAPP DYN: A Dynamic Micro-Simulation Model for the Italian Social Security System; CAPPaper: Singapore, 2012.
57. Hanappi, T.; Hofer, H.; Müllbacher, S.; Winter-Ebmer, R. IHS Microsimulation Model for Retirement Behaviour in Austria; Final Report; Institute for Advanced Studies: Vienna, Austria, 2012.
58. Boersch-Supan, A.H. Incentive Effects of Social Security under an Uncertain Disability Option. In Themes in the Economics of Aging; Wise, D.A., Ed.; University of Chicago Press: Chicago, IL, USA, 2001; pp. 281–310.
59. Belloni, M.; Alessie, R. Retirement Choices in Italy: What an Option Value Model Tells Us. Oxf. Bull. Econ. Stat. 2013, 75, 499–527. [CrossRef]
60. Samwick, A.; Wise, D.A. Option Value Estimation with Health and Retirement Study Data. In Labor Markets and Firm Benefit Policies in Japan and the United States; Ogura, S., Tachibanaki, T., Wise, D.A., Eds.; University of Chicago Press: Chicago, IL, USA, 2003; pp. 205–228.
61. Samwick, A.A. New Evidence on Pensions, Social Security, and the Timing of Retirement; NBER Working Papers Series; NBER: Cambridge, MA, USA, 1998.
62. Berkel, B.; Boersch-Supan, A. Pension Reform in Germany: The Impact on Retirement Decisions; NBER Working Papers Series; NBER: Cambridge, MA, USA, 2003.
63. Topa, G.; Depolo, M.; Alcover, C.M. Early Retirement: A Meta-Analysis of Its Antecedent and Subsequent Correlates. Front. Psychol. 2018, 8, 2157. [CrossRef]
64. Feldman, D.C. The Decision to Retire Early: A Review and Conceptualization. Acad. Manag. Rev. 1994, 19, 285–311. [CrossRef]
65. Kim, M.; Beehr, T.A. Retirement from Three Perspectives: Individuals, Organizations, and Society. In Aging and Work in the 21st Century; Shultz, K.S., Adams, G.A., Eds.; Routledge: New York, NY, USA, 2019; pp. 273–291.
66. De Preter, H.; Van Looy, D.; Mortelmans, D. Individual and Institutional Push and Pull Factors as Predictors of Retirement Timing in Europe: A Multilevel Analysis. J. Aging Stud. 2013, 27, 299–307. [CrossRef]
67. Topa, G.; Antonio, J.; Depolo, M.; Alcover, C.M. Antecedents and Consequences of Retirement Planning and Decision-Making: A Meta-Analysis and Model. J. Vocat. Behav. 2009, 75, 38–55. [CrossRef]
68. Sundstrup, E.; Thorsen, S.V.; Rugulies, R.; Larsen, M.; Thomassen, K.; Andersen, L.L. Importance of the Working Environment for Early Retirement: Prospective Cohort Study with Register Follow-Up. Int. J. Environ. Res. Public Health 2021, 18, 9817. [CrossRef]
69. Boissonneault, M.; Mulders, J.O.; Turek, K.; Carriere, Y. A Systematic Review of Causes of Recent Increases in Ages of Labor Market Exit in OECD Countries. PLoS ONE 2020, 15, e0231897. [CrossRef]
70. Boot, C.R.L.; Scharn, M.; Van Der Beek, A.J.; Andersen, L.L.; Elbers, C.T.M.; Lindeboom, M. Effects of Early Retirement Policy Changes on Working until Retirement: Natural Experiment. Int. J. Environ. Res. Public Health 2019, 16, 3895. [CrossRef]
71. Fischer, J.A.V.; Sousa-Poza, A. The Institutional Determinants of Early Retirement in Europe; Discussion Papers; Department of Economics, University of St. Gallen: St. Gallen, Switzerland, 2006.
72. Van Den Berg, T.; Schuring, M.; Avendano, M.; Mackenbach, J.; Burdorf, A. The Impact of Ill Health on Exit from Paid Employment in Europe among Older Workers. Occup. Environ. Med. 2010, 67, 845–852. [CrossRef]
73. Fleischmann, M. Should I Stay or Should I Go? In A Workplace Perspective on Older Ersons’ Labour Market Participation; Erasmus Universiteit Rotterdam: Rotterdam, The Netherlands, 2014.
74. de Wind, A.; Geusken, G.A.; Ybema, J.F.; Bongers, P.M.; Van Der Beek, A.J. The Role of Ability, Motivation, and Opportunity to Work in the Transition from Work to Early Retirement—Testing and Optimizing the Early Retirement Model. Scand. J. Work. Environ. Health 2015, 41, 24–35. [CrossRef] [PubMed]
75. Van Der Zwaan, G.L.; Hengel, K.M.O.; Sewdas, R.; De Wind, A.; Steenbeek, R. The Role of Personal Characteristics, Work Environment and Context in Working beyond Retirement: A Mixed-Methods Study. Int. Arch. Occup. Environ. Health 2019, 92, 535–549. [CrossRef] [PubMed]
76. Böckerman, P.; Ilmakunnas, P. The Journal of the Economics of Ageing Do Good Working Conditions Make You Work Longer? Analyzing Retirement Decisions Using Linked Survey and Register Data. J. Econ. Ageing 2020, 17, 100192. [CrossRef]
77. Trentini, M. Retirement Timing in Italy: Rising Age and the Advantages of a Stable Working Career. Ageing Soc. 2021, 41, 1878–1896. [CrossRef]
