Labour Economics 102 (2026) 102915

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Predicting retirement and Social Security claiming decisions using machine learning

Alexander Kwon[1], Lilia Maliar[*]

Department of Economics, Graduate Center, CUNY, 365 5th Ave, New York City, 10016, NY, USA

ARTICLE INFO

JEL classification:
C53
H55
J14
J26

Keywords:
Retirement and Social Security
Machine learning
Prediction
Gradient boosted tree
Shapley values

ABSTRACT

Accurate forecasts of retirement and Social Security (SS) claiming behavior are important for projecting the long-run costs of public pension and social-insurance systems. We study whether machine learning methods, applied to a rich set of household-level variables, can improve predictions of retirement and SS claiming decisions. Our model predicts the share of SS beneficiaries with an error of less than 1 percentage point, while the benchmark model used by the Social Security Administration (SSA) has an error of more than 3 percentage points. A conservative calculation implies that this difference corresponds to about 35.16 billion dollars per year in aggregate benefits. In a 2020 forecasting exercise using models trained on pre-2018 data, our model retains strong predictive performance during COVID-19 and outperforms the SSA benchmark. We also find that our model performs better across groups defined by expected SS benefits, with especially large gains among individuals with high expected benefits. In addition, the variables selected by our model differ from those emphasized in the SSA benchmark. Using Shapley values, we show that variable contributions are often nonlinear.

1. Introduction

Accurate forecasts of retirement and Social Security claiming behavior are important for projecting the costs of public pension and social-insurance systems. This issue is becoming increasingly important as aging populations place growing fiscal pressure on these systems around the world. In the OECD, for example, the old-age dependency ratio increased from 19 percent in 1980 to 31 percent in 2023 and is projected to reach 52 percent by 2060 (OECD, 2025). These forecasting problems also speak to a broader question in applied economics: whether data-driven methods can improve forecasts from standard parametric benchmark models, and whether the gains come from flexible functional forms, richer variable selection, or both. We address these questions for retirement and Social Security claiming in the United States, where the Social Security Administration's benchmark forecasting model relies on a pre-specified set of variables and standard probit and logit specifications.

The U.S. Social Security system provides an important context for this question. According to the most recent projections by the U.S. Department of the Treasury, the Old-Age and Survivors Insurance (OASI) Trust Fund is projected to be depleted around 2033.[2] Therefore, it is important to obtain accurate forecasts of individual retirement and Social Security claiming decisions to correctly predict the date of depletion. Accurate prediction is also important for policy discussions about the long-run solvency of the system.

In this paper, we use a machine learning technique called gradient boosted trees to obtain accurate predictions of retirement and Social Security claiming probabilities across individuals. We use gradient boosted trees because it can handle datasets with a large number of variables. Moreover, given that not all variables might be important for prediction, we need a machine learning model that is robust to the inclusion of unimportant variables. The gradient boosted trees can easily work with such large datasets in which not all variables might be important. In our comparison to other widely used machine learning models, such as random forest, neural networks, and Lasso estimation, we observe that the gradient boosted trees method delivers the highest performance on our test data; see Appendix A for our comparison results.

Another advantage of using machine learning, rather than using classical econometric tools or economic theory, is that it allows researchers to avoid making assumptions about which variables are important for prediction. The machine learning model itself can discern between important and unimportant variables for prediction, thus allowing us to avoid a subjective researcher's decision.

* Corresponding author.
E-mail addresses: akwon@ycp.edu (A. Kwon), lmaliar@gc.cuny.edu (L. Maliar).
[1] Co-author.
[2] See Trustees Report, Department of Treasury of 2025, https://www.ssa.gov/oact/TR/2025/trTOC.html

https://doi.org/10.1016/j.labeco.2026.102915
Received 9 November 2024; Received in revised form 1 June 2026; Accepted 4 June 2026
Available online 26 June 2026

To determine if machine learning can improve prediction, we compare our model's performance with that of the Social Security Administration's (SSA) 'Model of Income in the Near Term' (MINT); see Smith et al. (2021). MINT is a dynamic microsimulation model used as a tool to project future retirement incomes, as well as the distributional effects of Social Security reform proposals. The MINT model encompasses various economic and demographic projections, such as employment, individual earnings, periods of unemployment, and contributions to pension plans.

Using the Health and Retirement Survey (HRS) data from 1992 to 2018, we identify which variables are important for predicting retirement and Social Security claiming decisions, respectively. Although a large number of previous studies discuss the determinants of retirement and Social Security claiming decisions, there is no consensus regarding the variables that are important for prediction. Using machine learning, we are able to select the variables that are important for prediction out of a dataset that contains over 580 variables. We demonstrate that the gradient boosted trees method outperforms the commonly used logit and probit estimation methods in predicting retirement and Social Security claiming decisions.

The results of our analysis show that, for the retirement case, the variables related to respondents' job histories and retiree health insurance coverage are selected as important in our model, but are not used in the MINT. Conversely, the MINT incorporates demographic characteristics and education, which our model finds less important for prediction. In the case of Social Security, our analysis places greater emphasis on job history, health insurance, and birthplace, while the MINT incorporates variables related to self-employment status, health measures, household income and wealth, and private pension-plan variables. These disparities provide insights into areas where improvements in prediction can be made for each case.

To compare the performance of the two models, we use the area under a receiver operating characteristic curve (AUC). For the retirement case, the gradient boosted trees model using our selected variables has an AUC of 0.8407, while the probit estimation used by the MINT has an AUC of 0.7510. A significantly higher AUC in our analysis suggests that machine learning can help improve retirement prediction. For the Social Security case, the gradient boosted trees model using our selected variables has an AUC of 0.9729, compared to 0.9397 for the MINT logistic model. The improvement is therefore relatively larger in the retirement case, while for Social Security, the benchmark already performs well, and the gain is smaller in absolute terms.

To better measure the significance of improvements in prediction, we perform a two-step prediction exercise similar to that of the MINT. We first predict retirement and then predict Social Security beneficiary status using the predicted retirement. Our model yields less than 1 percentage point of error in the percentage of Social Security beneficiaries, while the MINT yields more than 3 percentage points of error. Using a conservative comparison, the difference between a 3.41 percentage-point error for the MINT and a 0.71 percentage-point error for our model implies a 2.7 percentage-point gap in predicted benefit receipt. Scaling this gap by the number of beneficiaries and the average monthly benefit, the implied discrepancy is $35.16 billion per year, which is approximately 2.5% of total Old-Age and Survivors Insurance benefit payments and 2.7% of total retirement benefit payments in 2025.

We further analyze where our method predicts better than the MINT and vice versa. We find that the gradient boosted trees model performs better in transitional cases, most notably for individuals between age 62 and the Full Retirement Age, and in cases that may involve more complex household or work decisions, such as having a DC pension or working longer hours in a second job. In contrast, the MINT performs best for respondents who are already receiving Social Security and for cases that more closely follow standard age-based or institutional patterns.

We also study the dollar cost of prediction errors across the distribution of expected Social Security benefits. We find that the gradient boosted tree model yields lower misclassification costs than the MINT in every benefit quintile, with the difference becoming much larger at higher benefit levels. In the top quintile, the average annual cost of misclassification is $2876 under gradient boosting (GB), compared with $4552 under the MINT. Moreover, the signed-cost results show that the MINT tends to over-predict claims much more strongly among individuals with high expected benefits, precisely where each mistake is most costly in dollar terms.

Next, we predict the Social Security beneficiary rate between 2018 and 2020 using the same two-step procedure, training the models on data before 2018. We consider the COVID-19 pandemic, which hit the U.S. in 2020, as an exogenous shock and use this exercise to assess the external validity of our model. We find that both models continue to perform reasonably well despite the unusual conditions in 2020, but the gradient boosted tree model performs better in this comparison. The GB prediction is closer to the observed beneficiary rate, and its AUC is substantially higher than that of the benchmark MINT model.

We also perform several robustness checks on our analysis. There are three potential drawbacks of our analysis. First, the information in our non-selected variables, which are used by the MINT, could already be reflected in our selected variables. To check this possibility, we regress the selected variables on non-selected variables and keep the residuals. These residuals show the leftover variation in the selected variables that is orthogonal to the non-selected variables. We then use the residuals as predictors in the gradient boosted trees. This allows us to see whether the residual variation in the selected variables helps prediction. If so, this implies that selected variables carry important information independently of non-selected variables, rather than relying on them.

Second, our results can be confounded by common time effects, such as business cycle fluctuations, policy reforms, or cohort-wide shocks that affect all individuals surveyed in the same interview period. For example, during an economic crisis, many individuals may decide to retire earlier due to uncertainty or job loss, while at the same time earnings, employment status, and related variables could shift systematically across the population. If such time-specific shocks are not accounted for, the model may mistakenly attribute these effects to individual-level predictors. To address this, we perform a robustness check by subtracting each variable from its mean across individuals in the same interview period (demeaning). This procedure eliminates any common time effect of a given interview period. If the model trained on demeaned variables continues to outperform the benchmark, this indicates that our results are not driven by common time effects.

Third, our results might be influenced by a difference in the definitions of retirement between our model and the MINT model. We define retirement based on a question where respondents can indicate that they consider themselves completely retired. In contrast, the MINT model classifies individuals as retired if they work fewer than 20 h per week. Following the MINT definition, we redefine retirement in our model and evaluate its performance.

In these sensitivity checks, we continue to find that our method outperforms the MINT model. These results provide suggestive evidence that, first, our selected variables contain information that is important for prediction independently of the non-selected variables used in the MINT. Second, common time effects do not account for the better performance of our model compared to the MINT. Third, the improved accuracy in our predictions is not attributable to differences in the definition of retirement.

Finally, using Shapley values from cooperative game theory, we show how the selected variables contribute to the predictions. We find that the contributions of the continuous variables are highly nonlinear. For the retirement case, the results point to hours worked, number of jobs over the employment history, financial variables, age, replacement

rate, and Social Security income as important contributors to prediction. The retirement SHAP patterns also provide suggestive evidence of job transition before retirement and are consistent with the idea that liquidity constraints may affect retirement decisions. For the Social Security case, birth cohort and age in months are among the most important contributors to prediction, together with retirement status, spouse-related claiming variables, and health insurance variables.

The present paper contributes to the literature that identifies the incentives of retirement and Social Security claiming decisions by giving suggestive evidence on what variables are important for prediction and how they contribute to predictions. The incentives are well documented in Chapter 8 of the *Handbook of Economics of Population Aging* by Blundell et al. (2016). According to Blundell et al. (2016), the incentives for retirement and claiming Social Security are closely related to each other since the decisions are not independent of each other.

The paper also contributes to the growing literature that applies machine learning to predictive tasks in economics (Ash et al., 2025; Bluwstein et al., 2023; Barbaglia et al., 2021). For example, recent work has used these techniques to successfully predict financial crises (Fouliard et al., 2020) and consumer defaults (Albanesi and Vamossy, 2019). In a similar context, this paper demonstrates that machine learning can be used to substantially improve forecasts of individual retirement and Social Security claiming decisions.

The paper is organized as follows: Section 2 presents the methodology of our analysis. Section 3 describes the data set used. Sections 4 and 5 discuss the numerical results. Section 6 demonstrates the robustness of our results. Section 7 compares our results about important variables with the previous literature. Section 8 concludes.

2. Methodology of our analysis

In this section, we describe the methodology used in this paper. We employ the following two-step procedure in our analysis. First, we fit a gradient boosted tree with all available variables to select variables important for prediction. These variables are subject to further data cleaning for the interpretability of the gradient boosted trees model related to those variables. Second, we re-estimate the gradient boosted trees with the cleaned variables, with all other variables, and check how each variable that is selected as important contributes to the prediction. In Appendix A, we compare the performance of other machine learning models, including random forest, Lasso, and neural networks, and show that the gradient boosted tree outperforms other models based on AUC with 5-fold cross-validation.

MINT The SSA's Modeling Income in the Near Term (MINT) is a stochastic dynamic microsimulation model that provides insight into the potential impacts of Social Security reforms. By projecting related variables for individuals using various economic theories and econometric tools, the MINT provides policymakers with a tool for counterfactual analysis to examine the distributional consequences of proposed changes to the social insurance program.

Version 8 of the MINT begins with micro-level data of actual and projected individuals born between 1905 and 2068. Its core sample is built from Survey of Income and Program Participation (SIPP) panels that are linked to Social Security administrative records for individuals born before 1980. For the extended cohorts—those born from 1980 to 2068—MINT draws synthetic 31-year-olds from SSA's POLISIM demographic model and pairs each of them with the same-age SIPP respondent. Then, the MINT uses econometric and rule-based modules to project life outcomes until each person dies or until 2099 (Smith et al., 2021; Smith and Favreault, 2019).

Although SIPP provides the starting point, the MINT incorporates data or parameter estimates from several other sources whenever they offer richer information for a particular problem. Specifically, the MINT uses the HRS data to estimate behavioral coefficients, employing probit estimation for predicting retirement probabilities and logistic regression for Social Security take-up.

Gradient boosted trees. We follow Hastie et al. (2009) to describe the gradient boosted trees.[3] A tree partitions the space of all joint predictor variable values into disjoint regions R_j, as represented by the terminal nodes of the tree, where $j \in \{1, \dots, J\}$ with J being the number of terminal nodes. If a data point x falls into the terminal nodes j (i.e., $x \in R_j$), then the decision tree predicts γ_j. That is, if f is a predictive model, then $f(x) = \gamma_j$.

A tree, $T(x;\Theta)$, of x with parameters Θ can be expressed as

$$T(x;\Theta) = \sum_{j=1}^{J} \gamma_j I(x \in R_j), \tag{1}$$

where $\Theta = \{R_j, \gamma_j\}_{j=1}^{J}$; $I(\cdot)$ is an indicator function; γ_j is the predicted value with $I(x \in R_j) = 1$.

The parameters Θ are found by minimizing a loss function $\sum_{j=1}^{J} \sum_{x_j \in R_j} L(y_j, \gamma_j)$,

$$\hat{\Theta} = \arg\min_{\Theta} \sum_{j=1}^{J} \sum_{x_j \in R_j} L(y_j, \gamma_j), \tag{2}$$

where y_j is the true value.

A boosted tree model is a sum of trees,

$$f_M(x) = \sum_{m=1}^{M} T(x;\Theta_m), \tag{3}$$

where M is the number of trees, and m is a tree index. The next tree is searched in a step-wise iterative method: given the collection of $m-1$ trees $f_{m-1}(x)$, the next tree is decided by solving

$$\hat{\Theta}_m = \arg\min_{\Theta_m} \sum_{i=1}^{N} L(y_i, f_{m-1}(x_i) + T(x_i;\Theta_m)) \tag{4}$$

for the region set and constants $\Theta_m = \{R_{j,m}, \gamma_{j,m}\}_{j,m=1}^{J_m}$ of the next tree, considering all data points $i = 1, \dots, N$. By using the Taylor approximation on L, the gradient-boosted tree updates the previous $f_{m-1}(x)$ using the gradient of the loss function L with respect to $f_{m-1}(x)$.

Selecting important variables. We consider three ways of selecting variables. A variable is selected as important if it enters at least two of the three different lists of important variables. In this way, we are being conservative. The three important variable lists are based on three different criteria: a mean decrease in impurity, a permutation-based method and Shapley values. The three lists contain the top 25 important variables for each criterion.

The mean decrease in impurity takes the weighted average of the decrease of the splitting criteria when a certain variable is included. *The permutation-based method* randomly shuffles the value of a variable of interest and checks the loss of the model. If the loss increases more as we shuffle a certain variable, that variable is considered to be more important. *The Shapley value method* compares the mean absolute value of Shapley values of a variable.

Shapley value.[4] Let S be a subset of the feature space F, x_S represent values of input features in set S, and $f_S(x_S)$ represent the model trained with feature S. A Shapley value for a single observation is the weighted average of all possible differences of a prediction model that includes a variable v ($f_{S\cup\{v\}}(x_{S\cup\{v\}})$) with the one that excludes

[3] See Hastie et al. (2009), Section 10.9

[4] Shapley values attribute contributions to a model's prediction, not causal effects. They are conditional on the trained model, the joint distribution of the features, and correlations among them. A causal interpretation would require a correctly specified structural model or assumptions that can help identify causal relationships.

that variable v ($f_S(x_S)$). In particular, a Shapley value of variable v is defined as

$$\phi_v = \sum_{S \subseteq F \setminus \{v\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S\cup\{v\}}(x_{S\cup\{v\}}) - f_S(x_S)]. \quad (5)$$

Intuitively, if we think of the prediction as a game and features as players, Shapley values can be thought of as a weighted average marginal contribution of a single player to every possible subset of players; see Kumar et al. (2020). For a strictly linear model, $Y_i = \sum_{j=0}^{k} \beta_j x_j + \epsilon$, the Shapley values for variable v for observation i simplifies to $\phi_v^i = \beta_v(x_v^i - \mu_v)$, where μ_v is the mean of v (Qin et al., 2025).

Shapley values are calculated with a machine learning technique SHapley Additive exPlanations (SHAP). In particular, since we use gradient boosted trees as our model, we employ the TreeSHAP. Sundararajan and Najmi (2020) show that the desirable properties of Shapley values, which are presented in Lundberg and Lee (2017), might not hold when a computational method uses conditional expectations, which is the case of TreeSHAP. The inability to assign zero Shapley values to variables that do not contribute might be problematic for our analysis (failure of dummy). However, this problem is circumvented by only considering the top 25 important variables when we select variables. Since the variables that do not contribute will be placed at the bottom of the list of important variables, we expect that the top 25 important variables list will not contain any unimportant variables.

Comparison of different lists. The three different lists do not have the same interpretation and all the lists have their advantages and disadvantages. Below, we explain the difference in interpretation and the advantages and disadvantages of each method. We then explain why we selected a variable if it enters at least two of the three lists.

First, the impurity-based list is based on the training data and shows variables important for making predictions (Molnar, 2022). However, the important variables based on the training data might not be important in the test data. Moreover, the impurity based method is known to favor high cardinality variables over low cardinality variables, such as binary features. The permutation-based and Shapley value methods do not have these problems. Thus, finding the intersection of the impurity-based list and the other lists can alleviate the problem that the impurity-based list possesses.

Second, the permutation-based list should be interpreted as an increase in loss when a variable is shuffled randomly. There are a couple of potential problems with permutation-based variable importance. First, the permutation process might yield unrealistic data that we might not observe in reality. For instance, let one variable be the child's age and the second variable be the father's age. If we shuffle the father's age, it is possible to obtain a shuffled data point such as the father's age being 20 and the child's age being 20 (only thinking of the biological father). Training the model with this unrealistic data point might create bias in computing the true variable importance. The TreeSHAP uses the conditional expectation for calculating the Shapley values. This implies that the TreeSHAP uses the given data instead of using unrealistic data when computing the variable importance. Thus, taking an intersection between the permutation-based method and the Shapley value method can alleviate the problem of the permutation-based method potentially relying on unrealistic data points.

Another possible problem with the permutation-based method is that if a variable has a high correlation with another variable, the importance can be split between the importance measures of the two variables, leading to a decreased importance measure for a highly important variable. The selection process in this paper truncates the important variable lists to the top 25 for each criterion. This might let us be unaware of important variables that are ranked lower than the top 25 due to a high correlation with the other variables. To test if our selection procedure selects important variables, we compare model performance (AUC) using the full set of variables against performance using only the selected variables. We find that the AUC only slightly decreases when using only the selected variables compared to the case of using all the available variables. Therefore, we expect that most of the highly important variables are included.

Finally, the Shapley values do not show the difference of the predicted values before and after removing a certain variable but represent the contribution of a variable to the difference between the actual and mean/baseline prediction (Molnar, 2022). Thus, if one wants to know what variables are important for correct prediction, Shapley values might not be the values one should rely on. Instead, the permutation-based method might be more suitable for this case. Therefore, considering an intersection between the permutation-based method and Shapley values method is similar to combining the two interpretations: on one hand, we want to select variables that have high contributions to the prediction, but on the other hand, we want to choose variables important for the correct prediction.

When using the TreeSHAP, there are further potential problems; see Sundararajan and Najmi (2020). The TreeSHAP uses the conditional expectation in the context of a decision tree or an ensemble tree which makes some of the desirable properties of the Shapley values not hold: dummy, linearity, symmetry, and demand monotonicity.[5] We expect that our analysis is not affected by the violation of the desirable properties since our analysis needs a measure of importance that captures the contribution of a variable to the prediction of the model, and does not need the properties to always hold.

Second step. Before we fit the gradient boosted tree, we clean the data again with the selected variables. We do the second cleaning because the variables in our dataset contain reasons for missing data, which are not numerical values. For example, if a respondent refuses to answer a survey question, which leads to the related variable to have a missing value, the variable is usually assigned with a character ".r", and ".d" is assigned if the respondent "don't know". Previously, a large negative value was assigned to each missing value, such as −83 or −96. If there is a systematic pattern for the missing values that differs from the non-missing values, the gradient boosted trees will split the data space between missing and non-missing values if it helps with prediction. In other words, we are giving a new category for each reason for "missing"; see Hastie et al. (2009), page 311.

Before the second step, we check each selected variable that contains missing values and either assign a reasonable value to the missing entries or drop the observation if the reason for the missing data is unknown. For example, the variable named *r14peni_n* represents the number of private pensions currently receiving. One of the possible responses, indicated by ".z", means that there is no private pension income for the respondent. Previously, this value was replaced with a large negative number, but in the second step, we replace it with 0. This process is important because we want to interpret the feature contribution. With the re-cleaned data, we fit a gradient-boosted tree again and use Shapley values to see how the selected variables contribute to the prediction of the dependent variable.

3. Data

Our data come from the Health and Retirement Study (HRS), specifically the *RAND HRS Longitudinal File 2018*. This dataset is a cleaned version of the HRS Core and Exit interviews processed by the RAND organization. The HRS data contain samples of the old-aged population in the U.S. We only include individuals whose age is equal to or over 55. We include waves from 4 to 14. Waves 1, 2 and 3 are not included because they have different interview terms depending on the

[5] This could be problematic since Shapley values are the unique payment attribution method that satisfies a subset of desirable properties including dummy, linearity, and symmetry. This implies that, in some cases, the TreeSHAP is producing something else than the theoretical Shapley values.

individual birth cohort. Using waves 4 to 14 allows us to keep the data-cleaning process consistent throughout the waves. We clean the sample separately for the retirement and Social Security cases.

For the retirement case, the dependent variable is a dummy variable that is 1 if the respondent considers themselves as completely retired. We pool the data as follows. First, we get the non-retired from wave 14. The implicit assumption is that individuals who consider themselves completely retired do not rejoin the labor force. However, that may not be true. For example, some individuals report that they retired at a certain wave but say that they are not or partly retired in later waves. We include these individuals in our analysis but categorize them as retired in the wave they first report as retired.

Next, we obtained retired individuals from each wave. To be specific, individuals who report that they consider themselves completely retired and report their retirement year between 2016 and 2018 are considered as individuals who retired in wave 14. We do this process for each wave. We construct the data in this way because variables usually ask about the status of the individual in the previous 2 years. Thus, wave 14 variables should be related to individuals who either retired in wave 14 or who did not.

However, we use previous-wave data for earnings, hours worked, and self-employment status. This is because retirement mechanically lowers earnings and hours worked, and individuals are usually no longer self-employed once they retire. For the same reason, we also use previous-wave values for some household asset variables. Further details are provided in Appendix B.

We drop individuals who receive Social Security Disability Insurance (SSDI) or Supplemental Security Income (SSI) benefits because the rules for those programs are different with OASI. We exclude variables that directly contain information about retirement. We further exclude variables that contain errors, and we exclude variables that are not in all waves from 4 to 14. There are 580 variables left in the sample. Finally, we only keep individuals who are married. We include only married individuals in our sample for two reasons. First, to ensure comparability with the MINT model, which estimates retirement and Social Security take-up equations separately by marital status. Second, many variables related to the spouse—such as spousal earnings, spousal retirement status, and spousal age—have a category of missing reason for unmarried respondents. Including only married individuals helps us avoid imputation for missing spousal information. More details about the data cleaning process are presented in Appendix B.

The sample consists of 7501 respondents, and 3973 consider themselves completely retired, while 3528 do not. When we do the second step, we re-clean the data by assigning specific values for missing values or dropping observations that contain missing values that cannot be assigned an intuitive number. In the re-cleaned sample, there are 6869 respondents, and 3700 consider themselves completely retired. For the Social Security case, we use the same set of variables but exclude two more variables that directly contain the information about Social Security. There are 7501 observations, and 5748 receive Social Security. The re-cleaned sample for the second step includes 5472 observations receiving Social Security out of 7108.

4. Numerical results

In this section, we present our numerical results. The hyperparameters used for the gradient boosted trees are reported in Appendix C. In Section 4.1, we present the selected variables as a result of our analysis. Section 4.2 compares our gradient boosted trees with the MINT models for both the retirement case and Social Security case. In Section 4.3, we perform the two-step prediction used by the MINT models with our gradient boosted trees method. In Appendix D, we predict early and delayed claiming using our method.

4.1. Selected variables

Table 1 shows the selected variables for the retirement and Social Security cases and compares them with those selected by the MINT. During the selection process, a variable that is highly correlated with another variable is selectively dropped, as we assume that the two variables contain similar information. A detailed selection process is described in Appendix E.

For the retirement case, we find that variables related to respondents' job histories and retiree health insurance coverage are selected as important in our analysis, but are not used in the MINT. In contrast, demographic characteristics and education-related variables are not selected as important for prediction in our model, while the MINT incorporates those variables.

However, we should be careful interpreting the results. They do not imply that demographic characteristics and education have no causal effect on retirement decisions. Instead, they show that when it comes to prediction, the aforementioned variables are less important than the set of selected variables. Because the gradient-boosted tree is not looking for a causal relationship, we acknowledge the fact that the selected variables might already contain information about demographic and educational information. A further discussion and sensitivity results are provided in Section 6.

Job-history variables can be informative about retirement through several channels. First, hours worked in the previous interview period provide a signal of labor supply transition that often precedes full retirement. In particular, reductions in hours may reflect a transition into partial retirement, so lagged hours help distinguish gradual exits from abrupt stops. Second, the number of jobs held over an individual's employment history can proxy for both economic preparedness and preferences over work. A long history of job changes may indicate a greater mobility or a weaker attachment to a given employer, characteristics associated with earlier retirement, whereas, alternatively, frequent job changes may reflect instability or limited savings that delay retirement.

For the Social Security case, our method selects variables related to job history and health insurance as important, but they are not used in the MINT. In contrast, self-employment status, health measures, household income and wealth, and private pension-plan variables are not selected as important in our analysis, despite playing a central role in the MINT.

Medicare eligibility and employer-provided health insurance are considered to be important variables in numerous previous studies (Rust and Phelan, 1997; French and Jones, 2011). Blundell et al. (2016) point out that the provision of health insurance by the employer is an important factor that explains US labor supply decisions. Workers who are not yet eligible for Medicare may be unable to retire because leaving employment can entail the loss of health insurance. Through this retirement–claiming link, Medicare eligibility becomes an important predictor of Social Security claiming. Moreover, the same logic can explain why the variable that indicates whether the individual is covered by employer-provided health insurance (namely, *Covered by employer HI*) and the indicator variable for whether the employer-provided health insurance covers retirees (namely, *Employer-provided health plan covers retirees*) are selected as important variables in our analysis.

The variable named *Industry code of the longest job* is included as an important variable for predicting Social Security claiming. Lifetime labor supply behavior can be different depending on the industry code, which is based on the 1980 census code. For instance, those who work in "Professional and related services", which includes physicians, legal services, education services, social services, engineering, architectural, accounting services, and more, might have a different retirement age than those in the "Agriculture, forestry, fishing" sector.

A comparison of the selected variables for retirement and Social Security claiming reveals overlap but also clear differences in the

emphasis. Age and job-history variables matter in both cases: age in months, respondent birth cohort, and hours worked on the main job (*Hours worked (job 1)*) appear in both models. At the same time, the retirement model places relatively more weight on health and financial variables, including health limitations in work (*Health limits work*), non-housing financial wealth, household capital income, household value of other debt, pension income, and current Social Security receipt. By contrast, the Social Security case places greater emphasis on demographic or human-capital-related variables, such as birthplace, earnings, industry code of the longest-reported job, and education. Since retirement and Social Security claiming are not independent, it is possible that the key variables in each case are also important in the other. Our analysis, however, shows that the importance of key variables differs between predicting retirement and Social Security claiming.

In the next section, we use the selected variables in Table 1 to fit the gradient-boosted tree. To both the retirement and Social Security cases, we add the replacement rate and the logarithm of earnings to be comparable with the MINT. The replacement rate, which is the percentage of how much your benefits will make up your earnings, summarizes the rules related to Social Security. Thus, it can be an important variable to know the effect of Social Security rules on decisions. If a replacement rate and log earnings are not important in prediction, then the gradient-boosted tree will not pick those variables to split. If they are important, then a better prediction will be possible. We find that including the two extra variables does not quantitatively change the results.

4.2. Predictive performance comparison

In this section, we discuss the predictive performance of the gradient boosted trees model. The results for the retirement prediction case are summarized in Table 2. To establish a baseline for comparison, we first replicate the probit model employed by the MINT. The predictive accuracy of this benchmark model is reported in the first column of Table 2. The MINT probit model produces an area under the ROC curve (AUC) of 0.7510. The second column contains the results when we use the MINT variables but use gradient boosted trees for the predictive model. The AUC increases slightly to 0.7565.

If we use all the available variables (580) and use gradient boosted trees as the predictive model, the AUC increases significantly to 0.8376, which is reported in the third column. If we further clean the selected 18 variables and re-estimate the gradient boosted trees model, the AUC slightly decreases to 0.8341. Although we observe a slight decrease in the AUC after cleaning the selected variables, it remains significantly higher than the benchmark model's AUC.

We find that our selected variables produce higher predictive performance than the MINT variables. To compare our approach with the MINT model using a similar number of variables, we use only the selected 18 variables to train a probit model and gradient boosted trees, and report the AUC in the 5th and 6th columns. With an identical probit specification, the AUC rises from 0.7510 when we use the MINT variables to 0.8220 when we employ our 18 selected variables, which is reported in the 5th column of Table 2. The results also indicate that these 18 variables capture most of the predictive content in the full feature set. With the same gradient-boosted trees specification, the AUC is essentially unchanged, and in fact rises slightly from 0.8341 with 580 variables to 0.8407 with only 18 variables (6th Column). In the last column, since the MINT variables do not include variables related to Social Security, we exclude two variables, specifically *Receives Social Security* and *Social Security income*, and retrain the gradient-boosted trees. The resulting AUC is still 0.8363, which is higher than the MINT probit model. Standard errors are obtained from 1000 bootstrap replications of the test set (Yadlowsky et al., 2025).

Table 3 compares the performance of each model for the Social Security case. The MINT logistic model already yields a high AUC of 0.9397, indicating that it can accurately distinguish between individuals receiving Social Security and those who do not. Replacing the logit with gradient-boosted trees on the same set of variables (2nd column) produces a comparable AUC of 0.9341. When we use all 578 available variables (3rd column),[6] the gradient-boosted trees model's AUC increases to 0.9724, and it remains very similar (0.9636, 4th column) after cleaning the selected 16 variables.

The 5th column of Table 3 suggests that our selected 16 variables contain more information about Social Security beneficiary status than the MINT variables, even though the gain in AUC is small. Using the same logistic regression model, our selected 16 variables generates a higher AUC (0.9670) than the MINT variables (0.9397). If we employ only the selected 16 variables, the AUC of the gradient boosted trees model is 0.9729 (6th column), and omitting both replacement rate and log-earnings, which are arbitrarily included to be comparable with the MINT, barely affects performance (the AUC is equal to 0.9728, 7th column). In summary, the MINT logistic model has high predictive accuracy, and using our methods improves predictive performance, but the increase is quantitatively small.

The contrast between the retirement and Social Security results provides insight into the information relevant for prediction in each context. In the case of retirement, gains in predictive performance arise mainly from variable selection rather than from the use of more flexible models. Replacing the MINT probit with gradient boosted trees while holding the MINT covariates fixed yields only a negligible improvement, whereas changing the variable set substantially raises predictive accuracy even under a probit specification. Allowing for a more flexible model provides an additional gain, but that gain is modest relative to the contribution of variable selection. For Social Security claiming, by contrast, the benchmark MINT logit already achieves a high AUC, so the scope for further gains is limited. Our methods still improve prediction, but the increase is modest in absolute terms. These results suggest that the benchmark model's main limitation in the retirement case lies in the information set, while Social Security claiming is already captured relatively well by the benchmark specification. They also suggest that estimating a standard probit or logit model on the variables selected by the machine-learning procedure may provide a useful alternative that preserves interpretability while improving predictive performance.

Another notable result is that a small subset of selected variables captures nearly all of the predictive content in the full feature set. In both the retirement and Social Security cases, models estimated on the selected variables perform almost identically to models estimated on all available variables. This suggests that most of the incremental predictive power comes from a limited set of variables rather than from contributions from hundreds of variables. In summary, the improvement over the benchmark is not driven solely by a high-dimensional model, but by a relatively compact set of predictors.

Overall, we find that our set of variables and trained gradient boosted tree model performs relatively better than the benchmark model, especially for predicting retirement status. However, it is not clear how much better our model is compared to the benchmark model. Therefore, in Section 4.3, we perform a two-step prediction used by the SSA to predict the Social Security claiming decision and compare the error rates between the two models.

4.3. Two-step prediction

In this section, we perform a two-step prediction procedure to check how much the prediction can be improved when we do the prediction sequentially. We follow the MINT model and predict retirement first, and then predict the Social Security status using the selected variables and predicted retirement status with the gradient boosted trees. To

[6] For the Social Security case we exclude two variables that directly record Social Security receipt and income (*Receives Social Security* and *Social Security income*) to avoid circular prediction. The variable count is therefore 578 rather than the 580 used for the retirement case in Table 2; see Section 3.

Table 1
A comparison of selected variables between the MINT and our analysis.

Category	MINT	Retirement	Social security
Age	Age dummies(52 to 63) Born 1938–1942 Born 1943–1954 Born 1955 or later	Age in months Respondent birth cohort	Age in months Respondent birth cohort
Health	Fair/poor health	Health limits work	
Income/Wealth	Per capita family wealth/AWI Homeowner	Non-housing financial wealth (net) Household capital income Value of other debt	
Earnings	Logarithm of earnings		Earnings/prev. wave
Retirement	Retired		Retired
Job history		Hours worked (job 1) Hours worked (job 2) Number of jobs over -employment history	Hours worked (job 1) Industry code of longest job Years worked
Self-employment	Self-employed	Household capital income (included)	
Social Security benefits		Receives Social Security Social Security income	
Private pension	DB plan DC plan CB plan	Pension income (indicator) Number of pensions currently receiving	
Health insurance		Employer-provided health plan covers retiree	Employer-provided health plan covers retirees Covered by employer HI Medicare
Household joint decision	Age difference to spouse Log spouse earnings Present value of spouse lifetime earnings divided by the cohort average Spouse black Spouse Hispanic Spouse two year age dummies (45–66) Spouse age 67 or higher Spouse DB plan Spouse DC plan Spouse CB plan Spouse self-employed	Spouse birth cohort Spouse retirement satisfaction	Spouse age in months Receives Social Security (spouse)
Demographics	Black Hispanic Foreign born		Birth place
Education	Less than high school Some college College More than college		Education (years)
Related to Social Security rules	Replacement rate and squared New age 64 to 68		

Notes: This table compares the sets of predictor variables used in the MINT with those identified as most important by our machine learning analysis. The column named "MINT" lists key variables used in the MINT model to predict retirement and Social Security claiming. The column named "Retirement" lists the top variables selected by the gradient boosted tree model for predicting retirement status. The column named "Social Security" lists the top variables selected by the gradient boosted tree model for predicting Social Security beneficiary status. The "Category" column groups related variables for easier comparison.

Table 2
Out-of-sample predictive accuracy (AUC) for alternative retirement probability models.

	MINT variables		All 580 variables		Selected 18		
	Probit	GB	GB	GB (cleaned)[a]	Probit	GB	GB without SS
AUC	0.7510	0.7565	0.8376	0.8341	0.8220	0.8407	0.8363
(s.e.)	(0.0103)	(0.0101)	(0.0068)	(0.0087)	(0.0126)	(0.0086)	(0.0088)

Notes: The table reports the area under the ROC curve (AUC), a standard summary of predictive accuracy. Columns (1)–(2) present AUC values for models estimated with the baseline MINT variables; Columns (3)–(4) show results when all available variables are included; Columns (5)–(7) report AUCs for models that use only the subset of variables selected by our feature-selection procedure. GB stands for *gradient boosted trees*. "GB without SS" means the selected-variable model excluding Social Security (SS) measures. S.E. stands for standard error and is computed from 1000 bootstrap replications of the test data.

[a] "Cleaned" stands for using cleaned selected variables before fitting the model.

make the two-step exercise comparable to the MINT specification and to avoid a circular prediction of retirement and Social Security receipt, we exclude Social Security-related variables, specifically *Receives Social Security* and *Social Security income*, from the retirement prediction step. We consider two specifications. First, we use all available variables after cleaning the selected variables. Second, we use only the selected variables to compare our method with the MINT model using a similar number of variables. The results are summarized in Table 4.

As shown in the first and second columns of Table 4, our model incurs an error of less than 1 percentage point on the test data. To be specific, the error is 0.71 percentage point when using all variables, and 0.47 percentage point when using only the selected variables.

Table 3
Out-of-sample predictive accuracy (AUC) for Social Security Claiming probability models.

	MINT variables		All 578 variables		Selected 16		
	Logistic	GB	GB	GB (cleaned)†	Logistic	GB	GB w/o RR & ln earn.
AUC	0.9397	0.9341	0.9724	0.9736	0.9670	0.9729	0.9728
(s.e.)	(0.0065)	(0.0068)	(0.0033)	(0.0037)	(0.0072)	(0.0039)	(0.0037)

Notes: The table reports the area under the ROC curve (AUC), a standard summary of predictive accuracy. Columns (1)–(2) present AUC values for models estimated with the baseline MINT variables; Columns (3)–(4) show results when all available variables are included; Columns (5)–(7) report AUCs for models that use only the subset of variables selected by our feature-selection procedure. GB stands for *gradient boosted trees*. "GB w/o RR & ln earn." means the selected-variable model excluding replacement rate and log earnings. s.e. stands for standard error and is computed from 1000 bootstrap replications of the test data. "Cleaned" stands for using cleaned selected variables before fitting the model.

Table 4
Two-step prediction of Social Security beneficiary status.

Observed beneficiary rate (%)	75.14		
	GB (all vars)	GB (selected vars)	MINT (Probit/Logit)
Predicted beneficiary rate (%)	75.85	75.61	78.53
Observed minus predicted (%)	−0.71	−0.47	−3.39
Mean bootstrap error	−0.74	−0.49	−3.41
Bootstrap standard error	0.52	0.53	0.64
AUC	0.9774	0.9721	0.9423

Notes: The table shows the average beneficiary share actually observed and that predicted by each model in a two-step procedure (first retirement, then claiming). The row "Observed minus predicted" reports the difference between the observed and predicted beneficiary rates in the original test sample. 1000 bootstrap replications on the test data are performed to obtain standard errors. "Mean bootstrap error" is the average difference (observed minus predicted, in percentage points) across resamples; negative values imply over-prediction. AUC is the area under the ROC curve for the two-step prediction for Social Security status. "GB (all vars)" denotes the model using gradient-boosted trees estimated on all available variables. "GB (selected vars)" denotes the model using gradient-boosted trees estimated on the selected variables. "MINT (Probit/Logit)" is the original sequential module in the SSA's MINT.

This provides evidence that our selected variables contain important information for prediction. However, as presented in the third column of Table 4, the MINT model overpredicts the beneficiary rate and generates an error of 3.39 percentage points in the test data.

To know the significance of the improvement of prediction accuracy, we perform a simple calculation. Based on a report published by the SSA in May 2025, the number of people receiving OASI for retirement benefits was approximately 55.6 million in May 2025.[7] Being conservative, an approximately 2.7 percentage-point difference (3.39 vs. 0.71) in prediction implies misclassifying 1.5 million people. The average monthly benefit was around 1950 dollars for OASI beneficiaries in May 2025. If we multiply 1.5 million by 1950, it amounts to 2.93 billion dollars. If we convert it to a one-year period, the 2.7 percentage-point difference amounts to 35.16 billion dollars, which is approximately 2.5 percentage points of the total benefit payments under Old-Age and Survivors Insurance and 2.7 percentage points of total retirement benefit payments in 2025.

We also predict the beneficiary rate between 2018 and 2020 using the same two-step procedure by training the models on data prior to 2018. Given that 2020 is the year when COVID-19 hit the U.S., there is a possibility that our model, trained with data before 2018, might predict poorly on 2020 data, indicating low external validity. Table 5 shows that both models continue to perform reasonably well, but the gradient boosted tree model performs better in this comparison. The observed beneficiary rate in 2020 is 26.54 percent. The MINT model predicts 25.17 percent, while the gradient boosted tree model predicts 26.08 percent, which is much closer to the observed rate. The bootstrap mean difference is 1.39 percentage points for the MINT model and 0.46 percentage point for the gradient boosted tree model, with corresponding standard errors of 0.76 and 0.67. In addition, the AUC is substantially higher for the gradient boosted tree model (0.9527) than for the MINT model (0.8693). Thus, even in the presence of the unusual conditions of 2020, the gradient boosted tree model maintains strong predictive performance and appears to generalize better than the benchmark MINT model.

[7] Thereportisnamed"MonthlyStatisticalSnapshot".

Table 5
Two-step prediction for 2020 data.

	Beneficiary rate (%)	AUC	Bootstrap	
			Mean Diff. (%)	SE (%)
Observed	26.54			
Predicted with MINT	25.17	0.8693	1.39	0.76
Predicted with GB	26.08	0.9527	0.46	0.67

Notes: The table reports the observed beneficiary share and the shares predicted by each model in the two-step procedure, where retirement is predicted first and Social Security claiming is predicted second. Standard errors are based on 1000 bootstrap replications using the test data. Mean Diff. reports the average difference between the observed and predicted beneficiary shares, measured in percentage points. Positive values indicate underprediction, while negative values indicate overprediction. The gradient boosted tree model uses only the selected variables.

In Section 6, we check the robustness of our results against three possible problems. First, the important information in the non-selected variables from the MINT model might already be included in the selected variables. Second, the time effects that are not considered in our model might cause a problem. Third, the definition of retirement in the MINT model is whether the individual worked less than 20 h, which is different from our definition. This might be the cause of the better performance of our model. We explore these issues in Section 6.

4.4. Where each model excels and the cost heterogeneity by benefit payments

In this section, we disaggregate the results of the two-step prediction to understand the specific characteristics of individuals for whom each model provides a more accurate prediction. Table 6 reports the means and standard deviations (in parentheses) of key variables across three prediction groups: cases where GB trees outperformed MINT, cases where the MINT outperformed GB, and cases where both models made the same prediction. The last column contains the results of a simple two-sample t-test assuming that the two groups are independent. A positive t-statistic indicates that the mean of the variable is significantly higher in the subpopulation where the gradient boosted tree model outperforms the MINT, whereas a negative t-statistic indicates the

opposite. In the table, we first present t-statistics in the order of their absolute values.

In summary, we find evidence that GB outperforms the MINT for individuals just before their Full Retirement Age (FRA), whereas the MINT performs best for individuals who are already receiving Social Security. This pattern indicates that the gradient boosted tree model may be better suited to transitional life stages in which incentives around retirement and claiming decisions are complex.

The most statistically significant variable favoring the gradient boosted tree model is *Between 62 and FRA* (t = 3.40). GB also performs relatively better for respondents with DC pensions (t = 2.94) and for those who report working more hours in a second job (t = 2.40). One possible interpretation is that these cases reflect patterns that are less standard or less directly captured in the data. Working nontrivial hours in a second job may indicate a labor-supply arrangement that is less well represented by the benchmark specification. Likewise, DC pension arrangements may be less transparent to the researcher than DB pensions. Unlike DB pensions, which are often tied to a formula based on salary and years of service, DC pensions are closer to retirement savings accounts. As a result, the incentives they create may be less directly inferred from the observed variables, even if they are clear to the individual holder.

Furthermore, while the large standard deviations warrant caution, there is a tendency for the gradient boosted tree model to better predict outcomes for individuals with higher total wealth, per capita wealth, earnings, and non-housing financial wealth. One possible explanation for this pattern is that individuals with higher income or wealth tend to depend less on public programs, such as Social Security or Medicare, which may lead to more accurate predictions from the nonlinear, data-driven model.

In contrast, the MINT outperforms GB most clearly for current Social Security beneficiaries (t = −9.61). The MINT also performs relatively better for respondents who are at or after FRA (t = −2.31) and for those younger than age 62 (t = −1.65). These patterns suggest that the MINT performs better when claiming behavior follows more standard age-based or institutional patterns.

Table 7 reports the average annual dollar cost of a prediction error in each quintile of expected Social Security benefits. Expected Social Security benefits are calculated as follows: For respondents receiving a positive amount of Social Security, we use the actual benefit reported in the variable recording Social Security income (*r14isret*). For everyone else, we approximate the future benefit from their most recent two-wave average earnings and apply the benefit formula to the average earnings. Multiplying this benefit figure by an indicator for whether the model's prediction was wrong gives a per-person cost. The table presents the mean of that cost within each benefit quintile for GB and MINT.

We find that GB yields lower misclassification costs than the MINT in every quintile of expected benefits, with the gap widening at higher benefit levels. In the bottom half of the distribution (Q1–Q2), the two models differ by less than $100, but above the median the gap becomes larger: in Q3, an error costs $678.22 under GB versus $839.61 under the MINT, in Q4 the cost is $1107.20 under GB versus $1483.85 under the MINT, and in Q5 the cost is $2876.26 under GB versus $4551.81 under the MINT. These differences suggest that, for individuals with higher expected benefits, selecting GB over MINT can substantially reduce the dollar cost of misclassification. However, sampling variation should be taken into account.

Finally, Table 8 reports the signed cost of misclassification. A positive value indicates that the model failed to identify an individual *truly receiving* Social Security, which can be interpreted as a potential underprediction of benefit receipt. A negative value indicates that the model predicted benefits for someone *not yet receiving* Social Security, which can be interpreted as a potential overprediction.

Table 8 reveals a clear pattern across the benefit distribution. In the bottom two quintiles, both models have positive signed costs, implying that errors are concentrated among individuals who actually receive benefits but are predicted not to. Starting in Q3, however, the signed costs turn negative for both models, indicating a shift towards overprediction. This pattern is much stronger for the MINT than for GB. In Q4, the signed cost is −344.03 for GB but −1041.16 for the MINT, and in Q5 it is essentially zero for GB (2.91) but remains strongly negative for the MINT (−2507.16). These results suggest that the MINT tends to overpredict claims much more strongly among individuals with high expected benefits, precisely where each prediction mistake is most costly in dollar terms. By contrast, the gradient boosted tree model is much closer to balanced in the top quintile, which helps explain why its overall misclassification costs are substantially lower in the upper tail of the benefit distribution.

4.5. Limitations of purely predictive models

Although our gradient-boosted tree dramatically improves out-of-sample fit relative to the SSA's MINT model, it remains a reduced form predictor rather than a decision rule. As Athey (2017) emphasizes, off-the-shelf machine-learning tools work best when the environment generating the data is stable and units do not interact or "interfere" with each other; when either assumption is violated, "pure prediction methods" can mislead policy makers because they do not reveal the causal effect of an intervention.

Recent evidence shows the danger. For example, Kiefer and Mayock (2025) find that mortgage-default models with excellent in-sample performance fail when securitization rules or macro conditions shift. This can be viewed as recent empirical evidence of the Lucas critique (Lucas, 1976), which suggests that reduced form models from historical data are not policy-invariant.

In this context, our study potentially has limitations in counterfactual analysis. Policy reforms—such as raising the full retirement age or altering benefit formulas—may likewise alter the mapping between individual characteristics and retirement or claiming behavior. This is because, as policy reforms are implemented, the regime changes, and the expectations or optimal strategies of individuals may change accordingly. Consequently, we interpret our estimates strictly as predictive probabilities; evaluating counterfactual reforms will require either econometric techniques that estimate heterogeneous causal effects or structural models that embed economic primitives.

The main contribution of this paper is to obtain highly accurate predictions of retirement and Social Security claiming decisions using machine learning on a dataset with high dimensionality. Such prediction analysis is valuable in dynamic microsimulation models, such as the MINT, because predicting each status is important for accurate simulation. Furthermore, we study what data reveal about the patterns of retirement and Social Security claiming in Section 5. These stylized facts are descriptive rather than causal, but they can guide the formulation of new hypotheses.

However, the direct implementation of our model within the MINT microsimulation model requires additional development. Embedding our method inside the MINT would first require that every predictor be either available in the simulation or projected by a new sub-module. Several variables such as job-history related variables are not currently simulated by the MINT. However, if the variables are available, our method can be used for predicting retirement and Social Security beneficiary status. Further analysis is needed to develop a method to project or simulate all the variables needed.

5. Contributions to prediction

In this section, we study how distinct variables affect prediction. In Section 5.1, we report the contributions for the retirement case. In Section 5.2, we report the Social Security beneficiary status case. In Appendix F, we employ only the selected variables for each case and look at the contributions of each variable. We perform this analysis to check if the contribution pattern changes if we use only the selected variables. We find that the overall contribution patterns are similar.

Table 6
Variable means by model-prediction comparison group (mean; SD in parentheses).

Prediction Group	GB better	MINT better	Same prediction	t stat
Receives Social Security	0.27 (0.44)	0.93 (0.26)	0.77 (0.42)	−9.61
Between 62 and FRA	0.73 (0.44)	0.38 (0.49)	0.27 (0.44)	3.40
DC	0.2 (0.4)	0.03 (0.19)	0.21 (0.41)	2.94
Hours worked (job 1)	30.63 (20.07)	34.03 (24.27)	26.97 (20.51)	−0.67
Hours Worked (job 2)	2.04 (6.94)	0.14 (0.74)	0.77 (4.02)	2.40
At or after FRA	0.03 (0.16)	0.21 (0.41)	0.42 (0.49)	−2.31
Before 62	0.24 (0.43)	0.41 (0.5)	0.31 (0.46)	−1.65
Retire	0.18 (0.38)	0.31 (0.47)	0.54 (0.5)	−1.36
Fair/poor health	0.2 (0.4)	0.1 (0.31)	0.13 (0.33)	1.35
Born 1955 or later	0.15 (0.36)	0.07 (0.26)	0.21 (0.41)	1.32
Born 1943–1954	0.85 (0.36)	0.93 (0.26)	0.31 (0.46)	−1.32
Covered by Employer HI	0.54 (0.5)	0.41 (0.5)	0.33 (0.47)	1.20
More than college	0.43 (0.5)	0.31 (0.47)	0.27 (0.45)	1.16
Spouse age in months	746.63 (62.63)	730.72 (70.6)	780.97 (91.05)	1.07
Receives Social Security (spouse)	0.38 (0.49)	0.28 (0.45)	0.76 (0.43)	1.03
Less than high school	0.18 (0.38)	0.1 (0.31)	0.17 (0.37)	1.02
Foreign born	0.25 (0.44)	0.17 (0.38)	0.13 (0.34)	0.93
Number of pensions currently receiving	0.23 (0.53)	0.34 (0.61)	0.32 (0.56)	−0.91
Pension income (indicator)	0.19 (0.39)	0.28 (0.45)	0.28 (0.45)	−0.90
Total wealth (including 2nd Residence)	1255.51 (3697.66)	847.3 (1214.88)	632.35 (1354.35)	0.86
Per wealth	606.69 (1850.96)	406.81 (612.78)	292.42 (660.85)	0.84
Years at longest reported job	16.25 (19.29)	18.86 (12.47)	20.22 (15.23)	−0.82
Social Security income	1.29 (4.52)	2.1 (5.4)	6.95 (8.99)	−0.72
Self employed	0.09 (0.29)	0.14 (0.35)	0.14 (0.35)	−0.68
Hispanic	0.23 (0.42)	0.17 (0.38)	0.12 (0.33)	0.65
Pension income	5.18 (18.09)	7.4 (16.12)	4.77 (13.51)	−0.61
Education (years)	13.68 (3.74)	13.21 (3.64)	13.2 (2.89)	0.60
spouse log earnings	6.74 (5.23)	6.04 (5.58)	4.41 (5.13)	0.58
Earnings/prev.	61.37 (66.48)	54.98 (44.61)	34.34 (44.36)	0.57
Non-housing financial wealth (net)	280.73 (978.9)	181.43 (725.41)	142.9 (487.22)	0.57
Female	0.51 (0.5)	0.45 (0.51)	0.49 (0.5)	0.53
white	0.75 (0.44)	0.79 (0.41)	0.83 (0.38)	−0.51
Homeowner	0.82 (0.38)	0.86 (0.35)	0.91 (0.29)	−0.50
Household business assets	24.94 (187.98)	16.38 (67.91)	23.33 (213.77)	0.35
Replacement	0.36 (0.23)	0.38 (0.2)	0.44 (0.26)	−0.32
Value of other debt	9.38 (20.51)	11.05 (27.21)	4.04 (12.6)	−0.30
Age in months	754.51 (32.61)	752.14 (41.42)	781.12 (80.13)	0.28
college	0.27 (0.44)	0.24 (0.44)	0.25 (0.43)	0.26
Household total income	136.4 (162.4)	128.86 (120.53)	91.08 (130.2)	0.26
Medicare	0.15 (0.36)	0.17 (0.38)	0.44 (0.5)	−0.25
Household capital income	34.3 (116.54)	30.1 (57.51)	22.77 (85.67)	0.25
black	0.09 (0.29)	0.1 (0.31)	0.1 (0.3)	−0.23
Years worked	31.29 (12.8)	31.66 (14.66)	38.12 (13.49)	−0.12
Pension/annuity income	7.94 (29.99)	7.45 (16.23)	5.15 (14.21)	0.11
Number of jobs	2.8 (1.78)	2.76 (1.84)	2.81 (1.66)	0.10
Health limit work	0.16 (0.37)	0.17 (0.38)	0.21 (0.41)	−0.09
DB	0.22 (0.41)	0.21 (0.41)	0.18 (0.39)	0.09
Born 1938–1942	0.0 (0.0)	0.0 (0.0)	0.21 (0.41)	

Notes: The first column presents the mean and standard deviations (in parentheses) of each variable where the gradient boosted trees model outperforms the MINT model. The second column presents those where the MINT models outperform the gradient boosted trees model. The third column reports those where the GB and MINT predict the same outcome. "t stat" reports the t-statistic for a test of equality of means between the "GB better" and "MINT better" groups. The following variables are in thousands: Imputed wage rate/spouse, Value of other debt, Household business assets, Household total income, Net wealth, per capita wealth, Earnings/prev. wave, Earnings/spouse, Pension income, Pension/Annuity income, Non-housing financial wealth (net), Household capital income, and Social Security income.

Table 7
Average annual cost of misclassification by expected benefit quintile.

Quintile	GB ($)	MINT ($)
Q1	109.75	144.98
Q2	378.90	475.77
Q3	678.22	839.61
Q4	1107.20	1483.85
Q5	2876.26	4551.81

Notes: The table reports the *average annual cost of misclassification*—the dollar-valued loss incurred when the model incorrectly predicts an individual's Social Security claiming status—by quintiles of *expected annual benefit*. Q1 contains the lowest and Q5 the highest expected benefits. "GB" refers to the gradient-boosted tree model estimated on 18 selected covariates; "MINT" is the Social Security Administration's benchmark sequential Probit/Logit model.

5.1. Retirement

Figs. 1 and 2 illustrate how each variable contributes to prediction. If a data point yields a positive Shapley value for a certain variable, it indicates that the value of that variable's data point contributes to an increase in the predicted probability of retirement. Fig. 1 shows a summary of Shapley values, while Fig. 2 plots the Shapley values for variables with more than two discrete values, including continuous variables.

In Fig. 1, the horizontal axis represents the Shapley value of each given variable listed in the left part of the panel. The red dots represent a higher value of the variables. Thus, if the red dots are distributed on the positive side of the Shapley values, while the blue dots are distributed on the negative side, the higher values of the given variable are associated with a higher predicted probability of retirement.

Fig. 1 shows that the most important variable for retirement prediction is whether the respondent already receives Social Security.

Table 8
Average annual signed cost by expected benefit quintile.

Quintile	GB ($)	MINT ($)
Q1	26.52	61.75
Q2	54.86	81.75
Q3	−120.92	−338.41
Q4	−344.03	−1041.16
Q5	2.91	−2507.16

Notes: The figures report the mean signed cost, in US-dollars per person per year, within quintiles of expected annual Social-Security benefit (Q1 = lowest, Q5 = highest). The signed cost is defined as (true beneficiary status − predicted beneficiary status) × expected annual benefit. A positive value indicates the model failed to identify a true beneficiary status (potential *underpayment*); a negative value indicates it predicted benefits for someone not yet entitled (potential *overpayment*). "GB" denotes the gradient-boosted tree; "MINT" is the Social Security Administration's benchmark sequential Probit/Logit model.

Respondents who receive Social Security tend to have positive Shapley values, while those who do not receive Social Security tend to have negative Shapley values. This means that the current Social Security receipt strongly contributes to a higher predicted probability of retirement. This pattern is intuitive because workers receiving Social Security benefits are more likely to be retired, although the Shapley values do not imply a causal effect.

Fig. 1 also reveals several additional patterns for binary variables. Positive pension income (*Pension income (indicator)*), having a spouse who is retired and satisfied (*Spouse retired: satisfied*), and having health limitations on work (*Health limits work*) are all associated with positive Shapley values, suggesting a higher predicted probability of retirement. Pension income may reflect both an additional source of retirement income and the fact that pension receipt is often closely tied to retirement status, so it is naturally associated with a higher predicted probability of retirement. The result for health limitations is intuitive because health limitations reduce the ability to continue working and therefore increase the predicted probability of retirement. The result for spouse retirement satisfaction is particularly interesting. Having a spouse who is already retired and satisfied is associated with a higher predicted probability of retirement, which is consistent with the idea of joint retirement within the household.

By contrast, respondents with *Plan not covers retirees* tend to have negative Shapley values, implying a lower predicted probability of retirement. In other words, not having retiree health coverage is associated with a lower predicted probability of retirement. This suggests that retiree health coverage remains an important predictor of retirement, which is consistent with previous work emphasizing the role of employer-provided health insurance in retirement decisions (Rust and Phelan, 1997; French and Jones, 2011). We further explain the mechanism in Section 7.

Fig. 1 also shows systematic differences across birth-cohort indicators. *CODA cohort*, *HRS cohort*, and *War Babies cohort* correspond to respondents born between 1924 and 1947, *Early Baby Boomer cohort* corresponds to those born between 1948 and 1953, and *Mid Baby Boomer cohort* and *Late Baby Boomer cohort* correspond to those born between 1954 and 1965. Detailed definitions of these cohorts are given in Table G.8 of Appendix G. Fig. 1 suggests that respondents born between 1954 and 1965 are less likely to be predicted as retired, since the Shapley values for *Mid Baby Boomer cohort* and especially *Late Baby Boomer cohort* are more concentrated on the negative side. By contrast, respondents born between 1924 and 1947, represented by *CODA cohort*, *HRS cohort*, and *War Babies cohort*, are more likely to be predicted as retired because their Shapley values tend to be more positive. In turn, *Early Baby Boomer cohort*, which corresponds to those born between 1948 and 1953, shows a mixed pattern, with Shapley values scattered on both sides of zero.

While Fig. 1 is useful for identifying the overall direction and relative importance of each variable, Fig. 2 is useful for visualizing the nonlinear patterns in continuous and multi-valued variables. A general pattern we find from Fig. 2 is that the relationships between these variables and the predicted probability of retirement are highly nonlinear.

Below, we summarize some key findings from Fig. 2. First, the hours-worked variables provide suggestive evidence on transitions into retirement. For the main job, individuals who worked relatively few hours per week in the previous wave tend to have positive Shapley values, while those who worked more than roughly 26 h have mostly negative Shapley values. This means that working longer hours in the main job is associated with a lower predicted probability of retirement in the next wave. The pattern is even stronger for the secondary job (*Hours worked (job 2)*): once hours in the second job are positive, the Shapley values are almost always negative. Combining these patterns implies that individuals who work fewer hours per week in their primary job have a higher probability of being retired, while those who work more in their secondary job have a lower predicted probability of retirement. Although these patterns do not establish causation, they suggest that there are individuals in job transition who, as they age, move from full-time work to part-time work and then to full retirement. Maestas (2010) shows that a large portion of retirees follow a nontraditional retirement path involving partial retirement.

Second, the number of jobs a respondent has held over their employment history (*Number of jobs*) also emerges as an important variable in predicting retirement in Fig. 2. According to our model, having only a small number of jobs is associated with positive Shapley values, whereas having more jobs is associated with increasingly negative Shapley values. Although this pattern is not necessarily causal, a larger number of jobs may reflect either less stable employment histories, which may lead to lower retirement preparedness, or a greater willingness to switch jobs when needed rather than retire. Both interpretations are consistent with a lower predicted probability of retirement.

Third, Fig. 2 indicates that financial variables matter, but in different ways. Household capital income tends to have negative Shapley values, especially at relatively high levels of capital income. One possible explanation is that household capital income may partly reflect income from self-employment or privately owned businesses. If so, respondents with higher household capital income may be more likely to remain economically active and work longer, which is associated with a lower predicted probability of retirement. Non-housing financial wealth (net) shows a different pattern. Observations with very low or negative values tend to have negative Shapley values, while positive wealth is associated with increasingly positive Shapley values. This is consistent with a liquidity-constraint interpretation of retirement, in which individuals with more financial resources are more able to retire (Blundell et al., 2016; Gustman and Steinmeier, 2005). Finally, positive values of other debt are associated with negative Shapley values. This result is intuitive, as individuals with larger other debts would need to work more to repay them and delay retirement. Note that other debts include credit card balances, medical debts, life insurance policy loans, loans from relatives, and so forth.

Fourth, Fig. 2 shows that age in months contributes to prediction in a nonlinear way. The Shapley values are negative at younger ages, then move towards zero and eventually become positive at older ages. Thus, older respondents are more likely to be predicted as retired, but the relationship is clearly nonlinear rather than linear. In particular, the figure shows heterogeneity in the contribution of age around thresholds such as 62, 66, 68, and 70. These thresholds are important because they are around early or delayed retirement ages and the Full Retirement Age (FRA). The changes in Shapley values around these ages suggest that retirement behavior varies across individuals depending on their position in the retirement transition, especially when they are close to early retirement or near the FRA.

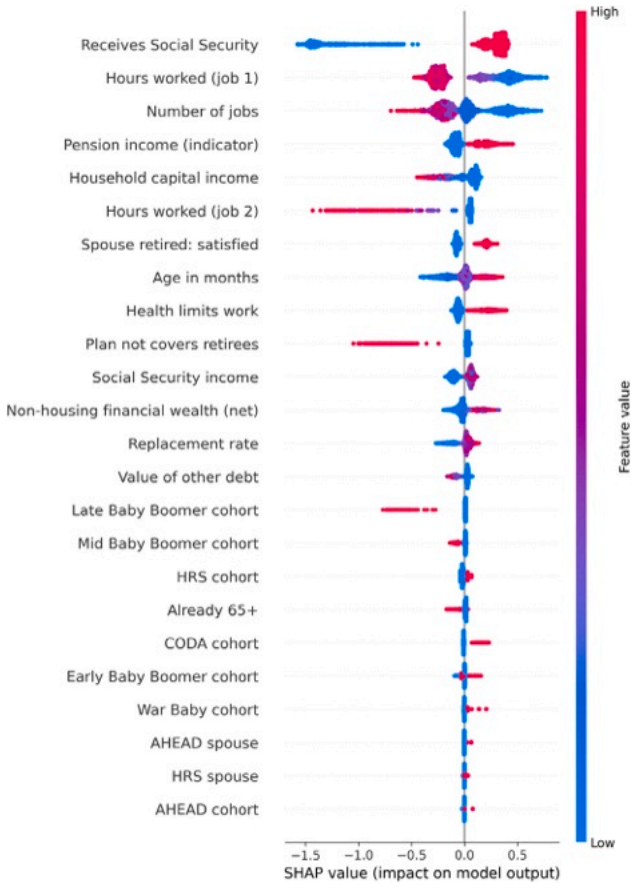

Fig. 1. Summary plot for the Shapley values for the retirement case.
Note: This figure presents a SHAP (SHapley Additive exPlanations) summary plot for the gradient boosted tree model predicting retirement. Each point on the plot represents the Shapley value for a specific variable and a single observation. The variables are ranked on the *y*-axis by their overall importance to the model's predictions. The *x*-axis indicates the Shapley value; positive values indicate a contribution towards a higher predicted probability of retirement, and negative values indicate a contribution towards a lower probability. The color of each point corresponds to the feature's value for that observation, with red indicating high values and blue indicating low values.

Finally, Fig. 2 also captures contributions from Social Security-related variables. Higher replacement rates are associated with more positive Shapley values once the replacement rate reaches around 0.3, whereas observations with zero or very low replacement rates tend to have negative Shapley values. This means that low replacement rates are associated with a lower predicted probability of retirement, which is consistent with the interpretation that a smaller benefit-to-earnings ratio may reduce the financial attractiveness of retirement. Similarly, positive Social Security income is associated with positive Shapley values. This may reflect both stronger access to retirement resources and the fact that retired individuals are more likely to claim Social Security benefits. These patterns suggest that, conditional on the other covariates, Social Security-related variables are associated with a higher predicted probability of retirement.

5.2. Social security

Figs. 3 and 4 show how the selected variables contribute to the prediction of Social Security beneficiary status. The summary plot in Fig. 3 indicates that birth cohort and age in months are among the most important predictors, along with retirement status, spouse-related claiming variables, and health insurance variables.

The cohort indicators show systematic differences across birth groups. Fig. 3 shows that *Mid Baby Boomer cohort*, *Late Baby Boomer cohort*, and *Early Baby Boomer cohort* have Shapley values concentrated in the negative domain when the dummy variable takes the value one. This means that respondents born between 1948 and 1965 are associated with a lower predicted probability of receiving Social Security benefits. By contrast, older cohorts, such as *HRS cohort* and *War Baby cohort*, tend to have more positive Shapley values, implying a higher predicted probability of receiving benefits. This result is intuitive because younger cohorts are less likely to have already claimed Social Security benefits at the time of observation.

The summary plot also shows several other important patterns. Respondents whose spouse already receives Social Security (*Receives Social Security (spouse)*) tend to have positive Shapley values, indicating a higher predicted probability that the respondent also receives benefits. Similarly, respondents who are already retired (*Retired*) tend to have positive Shapley values, which is intuitive because retirement and Social Security claiming are often closely related decisions. Medicare is also associated with positive Shapley values, which is intuitive because Medicare eligibility usually begins at age 65, near the Full Retirement Age for many individuals in our sample. In this sense, Medicare may capture that the respondent is close to or past an age at which Social Security claiming becomes more likely.

By contrast, respondents who are *Covered by employer HI* tend to have more negative Shapley values, implying a lower predicted probability of receiving Social Security benefits. One possible interpretation is that access to employer-provided health insurance reduces the incentives to claim benefits. More generally, individuals with current employer-provided coverage may be less likely to retire or claim Social Security if leaving work would mean losing health insurance before Medicare becomes available. Likewise, respondents whose employer health insurance plan does not cover retirees (*Plan not covers retirees*) tend to have negative Shapley values, suggesting that those without retiree health coverage may remain attached to work longer and therefore may be less likely to claim Social Security.

Fig. 3 also shows that some industry and birthplace variables contribute to prediction, although their effects are much smaller than those of age, cohort, retirement status, and spouse's Social Security receipt. In particular, *Birth: non-US* tends to be associated with negative Shapley values, implying a lower predicted probability of receiving Social Security benefits. One possible reason is that foreign-born respondents may have different work histories or shorter contribution histories in the United States. The transportation industry, *Ind: transport*, also tends to have negative Shapley values, suggesting a lower predicted probability of claiming. For other industries, variables such as *Ind: public admin*, *Ind: prof/rel*, *Ind: mfg dur*, and *Birth: WSC* also appear among the selected predictors and provide additional predictive information at the margin.

Fig. 4 shows that the relationship between several selected variables and Social Security prediction is highly nonlinear. Age in months provides the clearest example. The Shapley values are negative at younger ages, then rise sharply and become positive at older ages. In particular, the figure shows heterogeneity and nonlinear changes around ages 62, 66, 68, and 70. The pattern suggests that the contribution of age to Social Security prediction changes more sharply around these thresholds than within age intervals.

Education and earnings also show threshold-type patterns. The dependence plot for *Education (years)* suggests a break around 15 years: education levels below that point are associated with positive Shapley values, while education beyond that point is associated with lower predicted probabilities of receiving Social Security benefits. One possible interpretation is that individuals with higher education tend to remain in the labor force longer or have more financial flexibility to delay claiming. Log earnings show a related pattern. Values below roughly 10.8 are associated with positive Shapley values, while higher log earnings contribute negatively to the predicted probability of claiming.

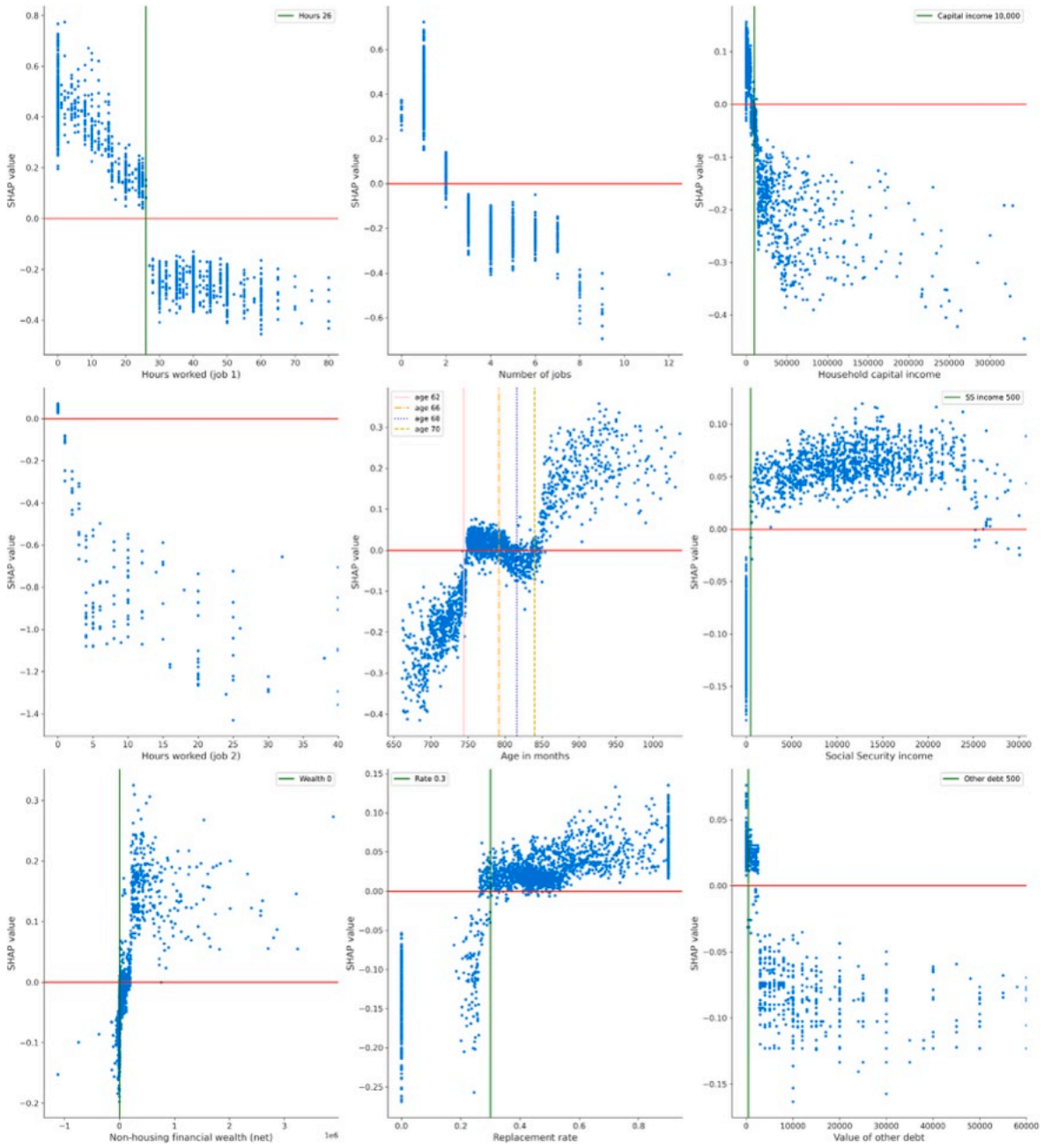

Fig. 2. Dependence plot for the retirement case.
Note: This figure displays SHAP dependence plots for selected important variables from the gradient boosted tree model for retirement prediction. Each subplot shows the relationship between the value of a single variable (x-axis) and its corresponding Shapley value (y-axis), which represents the variable's impact on the model's output for each observation. A positive Shapley value indicates that the variable's value for a given observation contributed to a higher predicted probability of retirement, while a negative value contributed to a lower probability. The horizontal line represents the zero point on the *y*-axis, and the vertical lines mark reference thresholds used to highlight where the patterns of Shapley values change. These plots reveal nonlinear associations between each variable and the model prediction.

This suggests that individuals with higher earnings in the previous interview period are less likely to be current beneficiaries.

The replacement rate and work-history variables also contain useful predictive information. When the replacement rate is low, the Shapley values are negative or close to zero, while higher replacement rates are associated with more positive Shapley values. This means that a lower replacement rate is associated with a lower predicted probability of receiving Social Security benefits, consistent with the idea that individuals have less financial incentive to claim when benefits replace only a smaller share of their earnings history.

The dependence plot for *Years worked* shows that the Shapley values become more positive after around 30 years of work history, suggesting that longer careers are associated with a higher predicted probability of receiving benefits. This pattern is consistent with the Social Security rules described in Appendix H. A worker must have sufficient work credits to qualify for Social Security retirement benefits, and benefit amounts depend in part on Average Indexed Monthly Earnings (AIME), which is based on the worker's highest 35 years of indexed earnings. In this sense, years worked is related both to eligibility and to the earnings history used to calculate benefits.

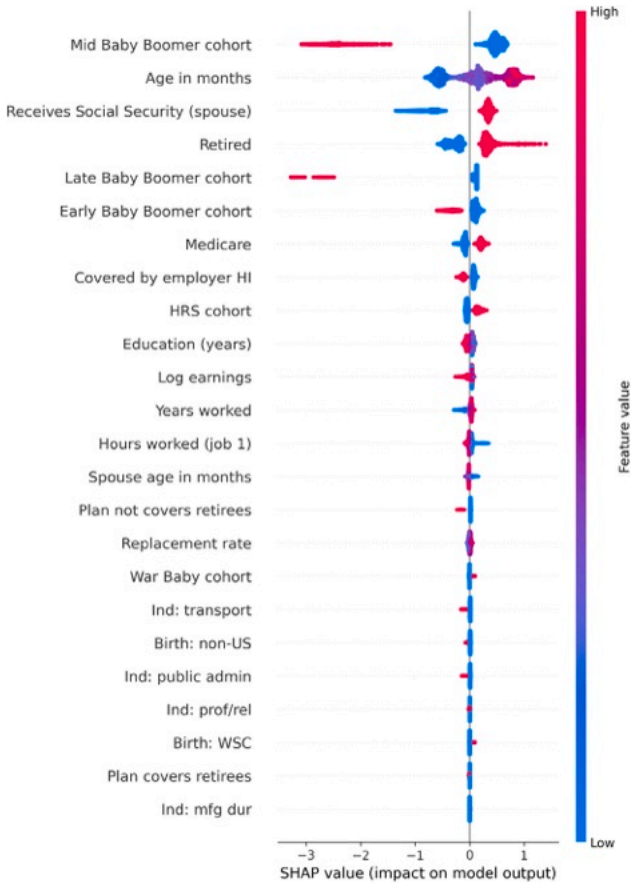

Fig. 3. Summary plot of the Shapley values for Social Security case.
Notes: This figure presents a SHAP (SHapley Additive exPlanations) summary plot for the gradient boosted tree model predicting Social Security beneficiary status. Each point on the plot represents the Shapley value for a specific variable and a single observation. The variables are ranked on the *y*-axis by their overall importance to the model's predictions. The *x*-axis indicates the Shapley value, where positive values signify a contribution towards a higher predicted probability of receiving Social Security benefits, and negative values indicate a contribution towards a lower probability. The color of each point corresponds to the variable's value for that observation, with red indicating high values and blue indicating low values.

By contrast, *Hours worked (job 1)* shows the opposite pattern: lower current hours are associated with positive Shapley values, while hours above around 30 are associated with negative Shapley values. This implies that stronger current labor-force attachment is associated with a lower predicted probability of already receiving benefits.

Finally, spouse age in months also contributes in a nonlinear way. The Shapley values are generally positive at relatively low spouse ages, then decline sharply around the Social Security eligibility range and remain close to zero or slightly negative at older spouse ages. One possible interpretation is that when the spouse is younger than the early eligibility age for Social Security, the household may be more likely to rely on the respondent's benefit as the first source of Social Security income. When the spouse is closer to claiming age, the household may have more scope to coordinate claiming decisions, which may alter the role of spouse age in predicting benefit receipt. Together with the strong positive pattern for *Receives Social Security (spouse)* in Fig. 3, this pattern is consistent with the idea that claiming behavior reflects joint household decision-making rather than only individual characteristics.

6. Robustness checks

In this section, we discuss three potential drawbacks of our results and provide suggestive evidence that our results are not affected by them.

First, one possibility would be that the important information in the non-selected variables of the MINT model is already reflected in the selected variables. For example, the respondent's race might already be influencing the number of years of work or non-housing financial wealth. Second, the common time effects, such as year-specific events or regimes, are not considered in our model, and they might be important. Third, the definitions of retirement in the MINT model and our dataset are different. In the MINT model, the individual who worked less than 20 h per week is defined as retired. However, retirement in our analysis is defined as those who consider themselves as completely retired. The difference in definitions might be responsible for the performance difference.

We assess whether our results change when we address the aforementioned three potential drawbacks. The first drawback is assessed by fitting each selected variable with the non-selected variables from the MINT model, using only the residuals for prediction. When fitting each selected variable, we employ polynomials up to the power of two for each non-selected MINT variable and include their interactions. Our goal is to determine whether the remaining variation in the selected variables contributes to prediction. The second drawback is checked by demeaning, i.e., by subtracting the mean of each variable across individuals in the same interview period. The third problem is checked by changing the definition of retirement to those who work under 20 h per week by following the MINT.

After considering different specifications, our method outperforms the MINT model. The results are summarized in Tables 9 and 10. The column named "Residuals" in Table 9 contains the result when we use the residuals of the selected variables after fitting with the non-selected MINT variables. The difference between the true and predicted beneficiary rates in the test data is −1.60 percentage points for our method, which is smaller than that of the MINT model (−3.39). This provides suggestive evidence that information in our selected variables, rather than in the non-selected variables, helps predict retirement status and Social Security status. The column "Demeaned" in Table 9 shows the result when we demean the variables to eliminate common time effects. When we demean the data, it generates a difference of −0.61 percentage points between the true and predicted beneficiary rates. A final specification, which combines this demeaning with the use of residuals, results in an error of −1.93 percentage points. The AUC values in Table 9 also suggest a similar story: all GB specifications continue to have higher AUCs than the MINT model, suggesting that the better performance of our model is not limited to matching the aggregate beneficiary rate, but also reflects better distinguishing beneficiaries from non-beneficiaries.

Table 10 shows the result when we redefine retirement following the MINT. We define an individual as retired if such an individual works less than 20 h per week. Our gradient boosted tree method yields a −0.57 percentage points difference between the true and predicted beneficiary rate, while the MINT model yields −3.44 percentage points difference. The last column of Table 10 shows the result when we use only the residuals and redefine retirement following the MINT. Our model still outperforms the MINT model even after redefining retirement. The AUC values reported in the two tables also point in the same direction: our GB analysis outperforms the MINT model.

7. Comparison of important variables with the previous literature

In this section, we provide a further discussion of the related literature. Table 11 summarizes selected studies, while Table 12 compares the important variables selected in our analysis with variables emphasized in the previous literature. Since retirement and Social Security

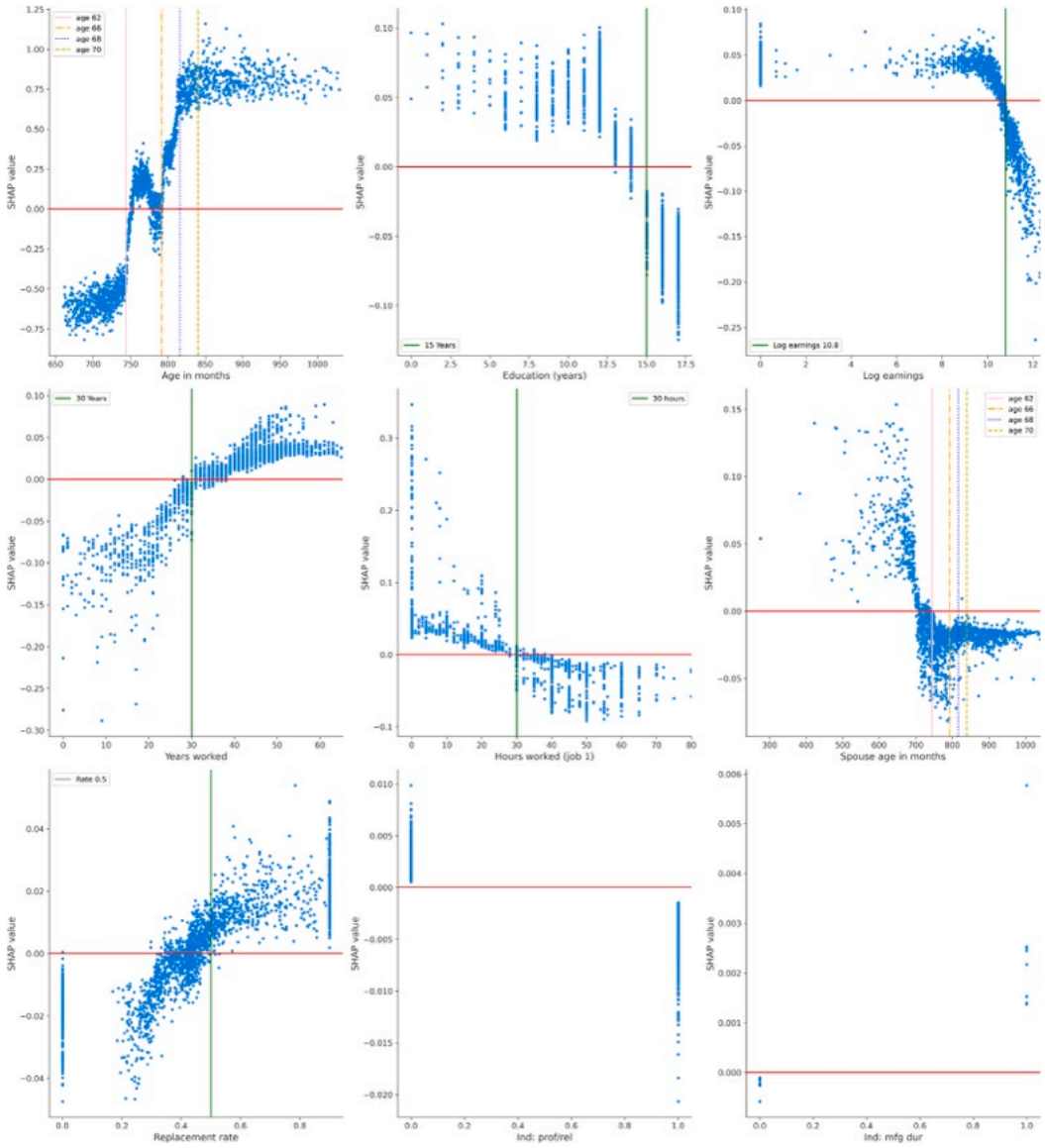

Fig. 4. Dependence plot for the Social Security case.

Notes: This figure displays SHAP dependence plots for selected important variables from the gradient boosted tree model for predicting Social Security beneficiary status. Each subplot shows the relationship between the value of a single variable (x-axis) and its corresponding Shapley value (y-axis), which represents the variable's impact on the model's output for each observation. A positive Shapley value indicates that the variable's value for a given observation contributed to a higher predicted probability of receiving Social Security benefits. The horizontal line represents a Shapley value of zero (the baseline prediction), and the vertical lines mark key thresholds where the variable's marginal effect on the prediction changes. These plots illustrate the non-linear relationships the model has captured.

Table 9
Robustness check with two-step prediction.

Observed beneficiary rate	75.14%				
	MINT	Original GB	Residuals	Demeaned	Demeaned and residuals
Predicted beneficiary rate	78.53%	75.61%	76.74%	75.75%	77.07%
Difference	−3.39	−0.47	−1.60	−0.61	−1.93
SS AUC	0.9423	0.9721	0.9608	0.9731	0.9584

Notes: This table compares the out-of-sample accuracy of the benchmark MINT model with our gradient boosted trees model in a two-step prediction of Social Security beneficiary rates under several robustness checks. The "Original GB" is our main model. The "Residuals" model tests whether the selected variables have predictive power beyond the information contained in the MINT variables. The "Demeaned" model controls for common time-specific effects by using variables demeaned by interview period. "Difference" is the observed beneficiary rate minus the predicted rate, in percentage points; negative values indicate overprediction of the beneficiary rate. "SS AUC" is the area under the ROC curve for the predicted Social Security beneficiary status and measures how well the model distinguishes beneficiaries from non-beneficiaries; a higher value indicates better discrimination.

Table 10
Two-step prediction after redefining retirement.

Observed beneficiary rate	75.14%		
	GB	MINT	Residuals
Predicted beneficiary rate	75.71%	78.58%	76.74%
Difference	−0.57	−3.44	−1.60
SS AUC	0.9709	0.9423	0.9598

Notes: This table presents a robustness check comparing model performance after redefining "retirement" to align with the benchmark MINT model's definition (working less than 20 h per week). The analysis uses a two-step procedure of first predicting retirement, then Social Security claiming. "GB" refers to our proposed gradient boosted model, and "Residuals" refers to the model trained on residual variation, both using the new retirement definition. "Difference" is the observed beneficiary rate minus the predicted rate, in percentage points. "SS AUC" is the area under the ROC curve for the predicted Social Security beneficiary status; a higher value indicates better discrimination between beneficiaries and non-beneficiaries.

claiming are closely related decisions, many variables discussed in the literature can matter for both outcomes. At the same time, our results show that some variables are more important for retirement, while others are more important for Social Security claiming.

As summarized in Table 11, previous research has emphasized the importance of health, health insurance, financial resources, and household decision-making for retirement and claiming behavior. Rust and Phelan (1997) show that health insurance and Medicare create important incentives for retirement, especially for individuals whose health insurance is tied to continued employment. Related work also finds that workers with employer-provided health insurance that does not continue after retirement are less likely to retire before Medicare eligibility (French and Jones, 2011). This is consistent with our results, where variables related to employer-provided health insurance, retiree health coverage, and Medicare are selected as important predictors. Previous work also shows that couples' retirement decisions tend to be closely linked (Hamermesh, 2000; Blau, 1997; Hurd, 1990), which is consistent with our findings that spouse-related variables are important in both the retirement and Social Security cases.

The literature on Social Security claiming identifies a broad set of determinants of early or delayed claiming. Coile et al. (2002) and Hurd et al. (2004) provide both theoretical discussion and empirical evidence on the timing of claiming. Coile et al. (2002) find that men with longer life expectancy delay claiming longer, that wealth has a non-monotone effect on delay, and that men with younger spouses delay more. Hurd et al. (2004) show that subjective survival probabilities matter for both retirement and claiming, although the effects are not large, and they also find that higher pension savings are associated with earlier claiming. Other work emphasizes the role of health limitations, work histories, wealth, debt, and job characteristics in claiming decisions (Li et al., 2008; Glickman and Hermes, 2015; Butrica and Karamcheva, 2018; Shoven et al., 2018; Huang et al., 2022; Haurin et al., 2022). These papers suggest that claiming behavior reflects a combination of health, work incentives, liquidity, household decision-making, and institutional rules.

When comparing our important variables to those discussed in the previous literature, age and birth cohort remain widely accepted predictors in both empirical and theoretical research on retirement and Social Security claiming. One reason is that age and cohort jointly convey information about eligibility and incentives under Social Security rules. For example, the Full Retirement Age differs by birth cohort, and this generates different incentives for retirement and claiming across cohorts. Our results are consistent with this. Age in months and respondent birth cohort are selected as important variables in both the retirement and Social Security cases, and the SHAP plots indicate highly nonlinear relationships with the prediction around threshold ages relevant to Social Security rules.

The health-related result in our analysis is whether a health problem limits work. Previous literature points to health as an important determinant of both retirement and claiming decisions, although the degree of importance varies across studies (Shoven et al., 2018; French, 2005; Blundell et al., 2016). For example, Li et al. (2008) show that health limitations at work increase the probability of early claiming. This is consistent with our retirement results, where observations with health limitations at work are associated with a higher predicted probability of retirement.

Among the variables related to income and wealth, previous literature emphasizes the role of liquidity, debt, earnings, and financial preparedness in shaping both retirement and claiming decisions (von Wachter, 2009; Glickman and Hermes, 2015; Rust and Phelan, 1997; Coile et al., 2002; Shoven et al., 2018; Blundell et al., 2016; Haurin et al., 2022). In our analysis, the most important variables in this category are non-housing financial wealth, household capital income, value of other debt, and previous-wave earnings. These findings are broadly consistent with the view that retirement and Social Security claiming depend on the household's ability to finance its post-retirement expenses. Our results suggest that positive non-housing financial wealth is associated with a higher predicted probability of retirement, while higher other debt is associated with a lower predicted probability of retirement. This is in line with the idea that liquidity constraints and debt can delay labor-force exit or claiming decisions.

Our analysis also shows that job history variables contain important predictive information. In the Social Security case, *Years worked* is selected as an important variable, and the SHAP values become more positive after roughly 30 years of work history. This result is also broadly consistent with Glickman and Hermes (2015), who find that individuals with longer work histories are more likely to claim early. By contrast, in the retirement case, variables such as *Hours worked (job 1)*, *Hours worked (job 2)*, and the *number of jobs over employment history* appear more important. Our results suggest that lower hours in the main job are associated with a higher predicted probability of retirement, which is consistent with French (2005)'s life-cycle model showing that labor supply declines sharply at older ages. The variables *Hours worked (job 2)* and *number of jobs over employment history* are less directly connected to the existing literature and require further research, but one possible interpretation is that they capture partial retirement or job-transition behavior before full retirement.

Our pension-related results are also broadly consistent with the previous literature, although the selected variables differ from the detailed pension-plan variables often emphasized in previous work. In our analysis, *Pension income (indicator)* and *number of pensions currently receiving* are selected as important predictors in the retirement case. This is broadly consistent with Hurd et al. (2004), who show that higher pension savings increase the likelihood of early claiming. More generally, our results suggest that the machine-learning model places more predictive weight on whether the respondent already has pension income than on detailed information about pension type.

Medicare, employer-provided health insurance, and retiree health coverage are also important predictors in our analysis. Medicare is associated with a higher predicted probability of Social Security claiming, while employer-provided health insurance is associated with a lower predicted probability of claiming. Retiree health coverage, by contrast, is associated with higher predicted retirement or claiming probabilities, while the absence of retiree health coverage is associated with continued work attachment. This is consistent with the previous literature and suggests that some respondents delay retirement or claiming because their health insurance remains tied to their current job and they are not yet fully protected by Medicare (Rust and Phelan, 1997; French and Jones, 2011; Blundell et al., 2016).

Finally, the bequest motive is widely recognized as an important factor in life-cycle models (Modigliani, 1988; Kotlikoff and Summers, 1981; Lockwood, 2018). In particular, Lockwood (2018) argues that bequest motives play a central role in understanding retirees' saving and

Table 11
Brief literature Review.

Retirement	
Paper	Summary
Blundell et al. (2016)	Summary of research on retirement incentives and labor supply. Couple of factors that affect retirement: Health, substitution effect, wealth effects, incentives from public pensions, incentives from private pensions, incentives from health insurance, behavioral aspects.
Rust and Phelan (1997), French and Jones (2011)	Social Security and Medicare provides a large incentive to retire for those who are not covered by employer-provided health insurance after retirement.
Hamermesh (2000), Blau (1997), Hurd (1990)	Discovers that couples' retirement roughly coincides.
Social Security	
Paper	Summary
Coile et al. (2002)	More likely to delay if longer life expectancy; and claiming delays follow a U-shape pattern in wealth (more wealth delays claiming at first but causes reduction in delays for further wealth due to bequest motive); having a pension result in shorter delay
Hurd et al. (2004)	High level of pension savings increases the likelihood of early claiming. Lower subjective probability of survival increases early claiming and retirement but not large effect.
von Wachter (2009)	Low-income households might claim early due to high replacement rate.
Li et al. (2008)	Health limitation at work increase the probability of early claiming.
Glickman and Hermes (2015)	Blue-collar jobs claim earlier. Those out of the workforce or who had longer work histories are more likely to claim early. Lower life expectancy leads to a higher probability of claiming early. Greater income and wealth lead to more delayed claiming
Butrica and Karamcheva (2018)	mortgage or credit card debt decreases the probability of claiming.
Shoven et al. (2018)	Ask the reason of early claiming: work, health, liquidity, and expectation-related reasons. Also, claiming was affected by behavioral aspects: normative retirement age, financial literacy
Huang et al. (2022)	Change in house value did not have an effect except during the 2002–2006 housing boom period: more housing wealth increased delayed claiming.
Haurin et al. (2022)	Financial stress increases the probability of delayed claiming at age 62.

Notes: This table provides a summary of selected research papers that are relevant to the study of retirement and Social Security claiming decisions. The table is organized into two sections: the top section focuses on literature related to retirement incentives, and the bottom section focuses on factors found to influence Social Security claiming behavior.

insurance decisions. However, unlike some previous studies, variables related to bequest motives are not selected as important predictors in our current model.[8]

8. Conclusion

In this paper, we show that machine learning, specifically gradient boosted trees, improves the prediction of individual retirement and Social Security claiming decisions using the Health and Retirement Study. The gains are especially large in the retirement case. In terms of the AUC, our model performs substantially better than the benchmark model used in the Social Security Administration's MINT. For Social Security claiming, the improvement is smaller in absolute terms, but still meaningful. In the two-step exercise that mimics the MINT procedure, our model predicts the share of Social Security beneficiaries much more accurately: the error is less than 1 percentage point in our model, while it is more than 3 percentage points in the benchmark. A conservative calculation implies that the gap in prediction translates into about $35.16 billion per year in aggregate benefit payments.

We further study where our model predicts better than MINT and where it does not. We find that the gradient-boosted trees model performs better in transitional cases, especially for individuals between the ages of 62 and the Full Retirement Age, and in cases where the decision problem may be more complicated or where the incentives are not captured with standard observable variables, such as working longer hours in a second job or having a DC pension. We also find that the dollar cost of prediction mistakes is lower under our model in every expected-benefit quintile, and that the difference becomes much larger at high benefit levels. In contrast, the MINT model performs better for respondents who are already receiving Social Security and in cases that more closely follow standard age-based or institutional patterns. In the 2020 external-validity exercise, both models continue to perform reasonably well despite the COVID shock, but the gradient boosted tree model performs better in that comparison.

Our results for the retirement case show that variables related to respondents' job histories and health insurance are important for prediction, even though they are not used in the benchmark MINT model. By contrast, the MINT includes demographic characteristics and education variables that our model finds less important for prediction. In the Social Security case, our analysis puts more weight on job history, health insurance, and birthplace, while the MINT puts more emphasis on self-employment status, health measures, household income and wealth, and private pension variables. These differences do not mean that the variables emphasized by the MINT are unimportant in a causal sense. Rather, they suggest that, for prediction, a different set of variables contains more useful information.

[8] In our working paper version that uses contemporaneous household asset variables, one bequest-related variable is selected as important for predicting Social Security beneficiary status.

Table 12
Comparison between the previous literature and selected variables.

Category	Selected variables	Retirement or SS	Previous literature
Age	Age in months	Retirement, SS	Mostly important
	Respondent birth cohort	Retirement, SS	Mostly important
Health	Health limits work	Retirement	Li et al. (2008)
Income/Wealth	Household total income	Retirement	
	Household capital income	Retirement	Blundell et al. (2016)
	Value of other debt	Retirement	Butrica and Karamcheva (2018)
Earnings	Earnings/prev. wave	SS	Blundell et al. (2016)
Retirement	Retired	SS	Glickman and Hermes (2015)
Job history	Hours worked (job 1)	Retirement, SS	French (2005)
	Hours worked (job 2)	Retirement	
	Number of jobs	Retirement	
	Industry code of longest job	SS	Glickman and Hermes (2015)
	Years worked	SS	Glickman and Hermes (2015)
Social Security benefits	Receives Social Security	Retirement	
	Pension income	Retirement	
Private pension	Pension income (indicator)	Retirement	Hurd et al. (2004)
	Number of pensions currently receiving	Retirement	Hurd et al. (2004)
Health insurance	Employer-provided health plan covers retirees	Retirement, SS	Rust and Phelan (1997), French and Jones (2011)
	Covered by employer HI	SS	Rust and Phelan (1997), French and Jones (2011)
	Medicare	SS	Rust and Phelan (1997), French and Jones (2011)
Household joint decision	Spouse birth cohort	Retirement	Hamermesh (2000), Blau (1997), Hurd (1990)
	Spouse retirement satisfaction	Retirement	
	Spouse age in months	SS	Hamermesh (2000), Blau (1997), Hurd (1990)
	Receives Social Security (spouse)	SS	Hamermesh (2000), Blau (1997), Hurd (1990)
Demographics	Birth place	SS	
Education	Education (years)	SS	
Related to Social Security rules	Replacement rate	Added for comparability	
Bequest	Not selected in current model	–	Lockwood (2018), Mukherjee (2022, 2018), Lee and Tan (2023)

Notes: This table connects the variables identified as important by our gradient-boosted trees model with findings from the existing economic literature. Variable names are harmonized to match the terminology used in the main text and Table 1 as closely as possible.

In addition, we use Shapley values to study how important variables contribute to the predictions. The results show that these relationships are often nonlinear. In that sense, the model is useful not only because it improves predictive accuracy, but also because it helps show where the data look more complicated than a standard linear or index-based specification would suggest. At the same time, these patterns should not be interpreted as causal effects. They are better viewed as descriptive evidence about which variables and interactions are most useful for prediction.

From a methodological perspective, a natural next step is to complement improved prediction with methods that allow a more formal analysis of heterogeneity. In particular, it would be useful to study whether the importance of key variables differs systematically across individuals. For example, causal machine-learning methods could be used to examine heterogeneous effects of the variables that emerge as most important in our analysis.

More broadly, our findings suggest that machine learning can be useful not only for improving forecasts, but also for uncovering patterns that may be useful for future economic work on retirement and Social Security claiming. We do not view the machine-learning model as a substitute for economic modeling. Instead, we see it as a complement. It can improve prediction, highlight where benchmark models make the biggest mistakes, and point to nonlinearities and interactions that may be worth building into richer economic models.

CRediT authorship contribution statement

Alexander Kwon: Writing – original draft, Software, Investigation, Data curation, Formal analysis. **Lilia Maliar:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT (OpenAI) and Claude (Anthropic) in order to improve the language and readability of the manuscript and to check for typographical errors. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendices

In this section, we present the supplementary results.

Appendix A. Comparison of different machine learning techniques

In this section, we describe the performance of other machine learning models, which include random forest, logistic Lasso, and neural networks, and we compare it with the performance of the gradient boosted tree—our main model.

A random forest is a bagging method applied to decision trees. The idea is to make many bootstrap samples and to train a decision tree on each bootstrap sample to decrease the variance of the overall model. If there are B bootstrap samples, we grow a random forest tree T_b for each $b =$ from $b = 1$ to $b = B$ with randomly selected m variables from

all the variables. For classification problems, the overall prediction is made with majority voting over B trees.

Let p be the number of variables, and let N be the sample size. The Lasso estimation method shrinks the coefficients and can set the coefficients of some variables to 0. Estimators from the logistic regression with Lasso penalty are obtained by solving the following problem:

$$\hat{\beta}^{\text{lasso}} = \arg\min_{\beta} \left\{ -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \ln \hat{p}_i + (1-y_i)\ln(1-\hat{p}_i)\right] + \lambda \sum_{j=1}^{p} |\beta_j| \right\}. \tag{A.1}$$

where the predicted probability is defined as

$$\hat{p}_i = \frac{1}{1+\exp(-\beta_0 - x_i^{\top}\beta)}. \tag{A.2}$$

To represent the neural network, we use the notations from Azinovic et al. (2022). Let K be the number of layers of the network, m_i be the number of neurons inside layer i, and τ_i be the activation function of layer i. Let $\rho = \{W_1, W_2, \ldots, W_K, b_1, b_2, \ldots, b_K\}$ be the trainable parameters that the neural network will update. Given hyper-parameters $\{K, \{m_i\}_{i=1}^{K}, \{\tau_i\}_{i=1}^{K}\}$, and trainable parameters

ρ, a neural network $\mathcal{N}_\rho$ is a map such that

$$x \to \mathcal{N}_\rho = \tau_K\left(W_k \ldots \tau_2(W_2\tau_1(W_1 x + b_1) + b_2)\cdots + b_K\right) \tag{A.3}$$

The objective function of the neural network we aim to minimize is given by the loss/cost function, $l(p)$.

The package `RandomForestClassifier` from scikit-learn is used to fit a random forest. The neural network is implemented using scikit-learn's `MLPClassifier`. We also estimate a logistic regression with an L1 penalty (logistic Lasso) using scikit-learn's `LogisticRegression`. All comparison models are tuned with 5-fold cross-validation using ROC–AUC as the scoring metric (`GridSearchCV`, `n_jobs=-1`) (see Table A.1). ***Gradient Boosting (XGBoost).*** The best-performing boosted-tree specification is as follows.

- `n_estimators=200`
- `max_depth=3`
- `learning_rate=0.1`
- `subsample=0.8`
- `colsample_bytree=1.0`
- `min_child_weight=3`
- `eval_metric=logloss`

Random Forest. The tuned random forest specification is as follows.

- `n_estimators=200`
- `max_depth=15`
- `criterion=entropy`
- `min_samples_split=5`
- `min_samples_leaf=2`

Neural network (MLPClassifier). The tuned MLP specification is as follows.

- `hidden_layer_sizes=(128,128)`
- `activation=logistic`
- `solver=adam`
- `alpha=10^-4`
- `learning_rate_init=0.001`
- `max_iter=500`

Logistic regression with L1 penalty (logistic Lasso). The tuned L1-penalized logistic regression in the notebook uses:

- `solver=saga`
- `penalty=l1`
- `C=10`
- `max_iter=100`

Appendix B. Data cleaning process

In this section, we summarize the data cleaning process. We perform the following steps:

1. Identify respondents who have retired for a specific interview period, called a wave. For instance, the interview period of wave 14 is from 2016 to 2018. We use waves from 4 to 14. The sample covers the period from 1996 to 2018.
2. Identify not-retired respondents from wave 14 (most recent interview period).
3. The variables like flu shot, cholesterol, mammogram, Pap smear, Prostate, and back variables are only interviewed on odd number waves. For even-number waves, the respondents can answer that the question was asked in the previous wave. If the respondents answered that they were asked in the previous wave, we use the previous wave answers for the variables. The occupation code for the longest job and the industry code for the longest job is not consistent. For example, the respondent's occupation code for the longest job can be in the 1980 census code, 2000 census code, or 2010 census code. Meanwhile, the AHEAD cohort is classified with another 9 categories. We try our best to match the occupation code to the 1980 census code with the provided data.
4. We replaced selected contemporaneous household financial variables with their previous-wave counterparts to avoid mechanical correlation with retirement and Social Security claiming outcomes. Specifically, we shifted the following variables from current wave to previous wave: *h14itot*, *h14absn*, *h14acd*, *h14aira*, *h14astck*, *h14achck*, *h14abond*, *h14aothr*, *h14adebt*, *h14ahous*, *h14amort*, *h14ahmln*, *h14atoth*, *h14atotf*, *h14atota*, *h14atotw*, *h14atotn*, *h14atotac*, and *h14icap* if we focus on wave 14.
5. We keep the longest job occupation code and the longest job industry code recorded in the 1980 census code.
6. If the respondent answers ".b", which means that the occupation or industry code is in other census code, we replace the ".b" with the other census code.
7. By using the detailed classification of the occupation code and industry code, we match similar names between the 1980 census code and other census codes. For instance, "Management occupations" in the 2000 census code is matched to "Managerial specialty operation" in the 1980 census code.
8. If there is no match for the name in other census codes, we categorize it in two ways. Let there be a category in another census code that does not have an exact match in names with the 1980 census code. For example, "financial specialists" have no exact match in the names of the 1980 census. Therefore, we find respondents who reported their occupation code as "financial specialists" and reported their occupation code in the 1980 census code in wave 14. We check if there is a category in the 1980 census code that most of those respondents are classified. For example, "financial specialists" are mostly "Managerial specialty operation" in the 1980 census code. We categorize "financial specialists" as "Managerial specialty operation". If there is no one category that most of those respondents are categorized, we split those respondents randomly into the 1980 census code as we observe from those who reported in both the 1980 census code and the other census code. For example, for those who reported "craftsmen/foremen/kindred workers", we split those respondents into "mechanics/repair", "construction trade/extractors", and "precision production" with the probability of 52/165, 50/165, and 63/165, respectively.

Table A.1
Performance comparison between different machine learning methods: retirement case.

Method	Gradient boosted tree	Random forest	Lasso (logit)	Neural network
AUC (Test)	0.8491	0.8243	0.5803	0.6593
AUC (Best CV)	0.8433	0.8129	0.5724	0.6400

Notes: This table compares the predictive performance of our main model (Gradient Boosted Tree) against three other common machine learning models for the retirement prediction task. Performance is measured by the Area Under the Receiver Operating Characteristic Curve (AUC). "AUC (Test)" is performance on the held-out test data, and "AUC (Best CV)" is the best score achieved during 5-fold cross-validation on the training data.

Table C.2
Gradient-boosted decision tree hyper-parameters used for variable selection (stable configuration).

Hyper-parameter	Retirement case	Social Security case
Maximum depth of a tree	5	3
Minimum sum of instance weight (hessian) needed in a child	1	1
Number of trees (estimators)	500	500
Subsample ratio of the training instances	0.5	1.0
Subsample ratio of columns when constructing each tree	1.0	1.0
Evaluation metric	logloss	logloss
Gamma (minimum loss reduction)	1.0	0.0
Lambda (L_2 regularization; `reg_lambda`)	3.0	1.0
Learning rate	0.01	0.3

Notes: This table reports the stable hyperparameter configuration used during the variable-selection procedure to ensure comparability across the contemporaneous and previous interview period household asset variables implementations.

9. When we merge waves from 4 to 14, some variables are not available among all 11 waves. We drop those variables.
10. We drop variables that contain information about the interview waves.
11. We drop variables that have a erroneous skip pattern specific to 2018 HRS survey.
12. We drop respondents who receive SSDI or SSI.
13. We only keep respondents who are married.
14. We keep respondents whose age are greater or equal to 55.
15. We change the missing reasons to a large negative value. We check and change the value if the negative value overlaps with the variables that can have negative values such as the net value of different assets.

Appendix C. Selected hyper-parameters

This section reports the hyper-parameters used for the gradient-boosted decision tree models.

During revision, we replaced contemporaneous household asset variables with their previous interview period analogs to avoid mechanical correlation with retirement and Social Security claiming outcomes. Re-optimizing hyper-parameters under the new covariate set turns out to matter little. For example, AUROC increases only from 0.9724 to 0.9740 relative to the previous wave-optimized (stable) configuration for the Social Security case—well within bootstrap uncertainty (SE $\approx$ 0.0032), and for the retirement case specifically, AUROC is marginally higher under the stable configuration. We therefore use a single stable configuration throughout the variable-selection procedure. The contemporaneous specification results are reported in the working paper version of this study.

Table C.2 reports the hyperparameter configuration used in the variable-selection procedure. Table C.3 reports the re-optimized hyper-parameters used in the main specification after variable selection. In both cases, hyperparameters are selected using 5-fold cross-validation with AUROC as the scoring metric.

Appendix D. Selected variables for early and delayed claiming prediction

In this section, we predict early and delayed claiming using our research method and the constructed dataset. Our objective is to understand the differences between those who claim early or late and others.

We keep individuals whose age is between 62 and 70 because 62 is the age when an individual can start to claim and 70 is the age after which there are no further delayed-retirement credits. We further exclude variables that contain answers that indicate age, since those variables would capture age information rather than the intended economic content of the variable. The excluded variables include some probability-related variables (namely *r14liv75*, *r14liv75r*, *r14liv75c*, *r14liv75f*, and *r14pnhm5y*) because, above certain ages, the answers are coded as missing values. In this appendix exercise, unlike the main specification, we use contemporaneous household asset variables.

Age related variables are not excluded when we select the important variables in the first stage because we want to control age and select variables that are important for prediction. However, we only keep the birth cohort of the individual and exclude other age related variables for the second stage since age is trivially related to early and delayed claiming.

The dependent variable for the early claiming case is a binary variable: it is 1 if the respondent received Social Security before the FRA, and 0 otherwise. Similarly, the dependent variable for the delayed claiming case is a binary variable: it is 1 if the respondent received Social Security after the FRA, and 0 otherwise. It is important to note that we do not exclude individuals who did not claim Social Security, meaning that our classification problem is 1 vs. rest. Out of 5918 individuals between the age 62 and 70, 3610 claimed Social Security earlier than the FRA, while 619 claimed Social Security later than the FRA.

In the second step, we clean the selected variables and train the gradient boosted tree with those variables. The AUC for the early claiming case is around 0.8358 and is around 0.7502 for the delayed claiming case. When we do not further clean the selected variables,

Table C.3
Gradient-boosted decision tree hyper-parameters used in the main specification (Household assets shifted to previous interview period).

Hyper-parameter	Retirement case	Social Security case
Maximum depth of a tree	3	5
Minimum sum of instance weight (hessian) needed in a child	3	3
Number of trees (estimators)	200	500
Subsample ratio of the training instances	1.0	0.5
Subsample ratio of columns when constructing each tree	1.0	1.0
Evaluation metric	logloss	logloss
Gamma (minimum loss reduction)	0.5	0.5
Lambda (L_2 regularization; `reg_lambda`)	1.0	1.0
Learning rate	0.1	0.01

Notes: This table reports the hyperparameters used for the main specification after the variable-selection step. Hyperparameters are selected using 5-fold cross-validation with AUROC as the scoring metric.

the AUC is 0.87 and 0.88 for early and delayed claiming, respectively. We find that our variable selection process decreases the performance of the prediction for the delayed claiming case more than the early claiming case. Therefore, when we interpret the Shapley values, we refrain from strong claims about the contribution patterns for the delayed claiming case since the performance is relatively low.

Table D.4 compares the selected variables for each early and delayed claiming to those selected when we predicted retirement and Social Security claiming decisions. Interestingly, there are variables that are newly selected as important for early claiming: the type of employer-provided pension (namely *Private pension type*), whether financial or family respondent (namely *Financial, family respondent*), and years of education (namely *Years of education*).

Meanwhile, a variable related to subjective life probability is newly introduced as important for predicting delayed claiming, namely *Chg live 80-100: R/LfTab ratio*. This variable, according to the data description, measures change in the respondent's self-reported probability of living to an age from 80 to 100, relative to the probabilities implied by the Vital Statistics life table. However, the data description warns users that this variable might be inconsistent due to differences in wording across various interview periods. Therefore, while we describe how subjective life probability affects predicting delayed claiming, we refrain from making strong claims about our findings regarding this variable. In an attempt to decrease the inconsistency from wording differences, we make dummy variables that indicate if the change in subjective life probability is negative, positive, or zero. We include these three dummy variables when we perform the second step.

Figs. D.1 and D.3 show the Shapley values for selected variables when predicting early claiming. We find that low earnings in previous waves are associated with a higher probability of early Social Security claiming, while high earnings are linked to a lower probability of early claiming. The same pattern holds for household total income: lower household total income is associated with a higher probability of early claiming.

Moreover, if we look at the last panel in the second row of Fig. D.3, high levels of education help the model predict a lower probability of early Social Security claiming. Specifically, having 15 or more years of education is associated with a lower probability of early claiming. The third panel in the first row of Fig. D.3 shows that working over approximately 30 h at the primary job in the previous wave is linked to a lower probability of early claiming.

If we look at Fig. D.1, we find that if the respondent is a financial or family representative, this role is associated with a higher probability of early Social Security claiming. Additionally, if the respondent's spouse was born outside of the US (specifically *spouse_BP_11*), this factor is linked to a lower probability of early claiming.

Figs. D.2 and D.4 present the Shapley values for the delayed claiming case. We observe an opposite pattern compared to the early claiming case: higher earnings in previous waves and higher household total income are associated with a higher probability of delayed Social Security claiming. However, if the respondent is classified as a financial or family representative, there is also a higher probability of delayed claiming, which reflects a similar pattern observed in the early claiming case. These results suggest that financial or family representatives are more likely to receive Social Security benefits rather than not. Additionally, if we look at the last row of Fig. D.4, we find that individuals experiencing a negative change in life expectancy are more likely to delay claiming, whereas those with a positive change in life expectancy are more likely to claim early. These results are counterintuitive because if someone's perceived life expectancy decreases, you would expect them to claim Social Security sooner rather than later. Further analysis is needed, but as mentioned earlier, the inconsistency in the variable could be the issue.

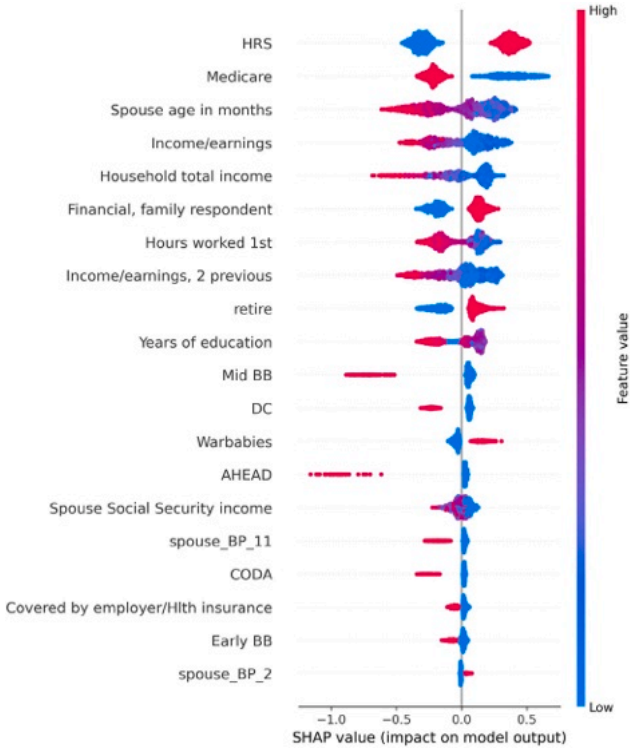

Fig. D.1. Summary plot for the Shapley values for the early claiming case.
*Notes:*This SHAP summary plot corresponds to the model predicting early Social Security claiming (i.e., claiming benefits before FRA). The analysis is limited to individuals aged 62 to 70. The *y*-axis ranks variables by importance. The *x*-axis shows the Shapley value, where a positive value indicates a contribution towards a higher predicted probability of early claiming. The color of each point shows the variable's value (red for high, blue for low).

Table D.4
Selected variables for early and delayed claiming.

Category	Originally selected variables	Early	Delayed
Age	Age in months Respondent birth cohort	Age in months Respondent birth cohort Respondent birth year	Age in months Respondent birth cohort Respondent birth year
Health	Health limits work Doctor visit/ dummy		
Income/Wealth	Household total income Net wealth/exclude secondary residence Value of other debt	Household total income	Household total income
Earnings	Earnings/prev. wave	Earnings/prev. wave Earnings, 2 previous	Earnings/prev. wave Earnings, 2 previous
Retirement	Retired	Retired	Retired
Job history	Years at longest reported job Hours worked per week in -1st job in previous wave Hours worked per week in -2nd job in previous wave Number of jobs over -employment history Industry code of longest job	Hours worked per week in -1st job in previous wave	Hours worked per week in -1st job in previous wave
Self-employment	Household business assets		
Social Security benefits	Receives Social Security Social Security income		
Private pension	Pension/Annuity income Pension income (indicator) Number of pensions currently receiving	Private pension type	
Health insurance	Medicare Covered by employer HI Employer-provided health plan covers retirees	Medicare Covered by employer HI	
Household joint decision	Receives Social Security (spouse)	Spouse age in months Spouse birth place Spouse Social Security income	Spouse birth date Spouse years worked at longest job
Demographics	Birth place	Years of education Financial, family respondent	Financial, family respondent
Bequest	Probability to leave bequest over 100K		
Life probability			Chg live 80-100: R/LfTab ratio

Notes: This table compares the sets of important variables selected for different prediction models. The "Originally selected variables" column lists the variables used for the main retirement and Social Security predictions in the paper. The "Early" and "Delayed" columns show variables selected for specialized models that predict Social Security claiming before or after an individual's FRA, respectively. This analysis is limited to individuals between the ages of 62 and 70.

Appendix E. Variable selection process

In this section, we describe the variable-selection process in detail. The three important-variable lists are shown in Tables E.5 and E.6.

A variable is selected as important if it appears in at least two of the three importance lists. The resulting lists are shown in Table E.7.

Among these selected variables, we then selectively drop variables whose pairwise correlation exceeds 0.6. The idea is that highly correlated variables tend to contain similar information about retirement or Social Security claiming status, so keeping all of them would add redundancy without contributing much additional predictive content.

For the retirement case, we make the following choices when two or more variables are highly correlated or convey very similar information:

Birth year is highly correlated with *Birth cohort*. We keep *Birth cohort* and drop *Birth year*.
Age when started to receive SS is highly correlated with *Receives Social Security*. We keep *Receives Social Security* and drop *Age when started to receive SS*.
Weeks worked (job 1) is highly correlated with *Hours worked (job 1)*. We keep *Hours worked (job 1)* and drop *Weeks worked (job 1)*.
Weeks worked (job 2) is highly correlated with *Hours worked (job 2)*. We keep *Hours worked (job 2)* and drop *Weeks worked (job 2)*.
Self-employed is highly correlated with *Hours worked (job 1)*. We therefore drop *Self-employed*.
Current pension type #1 is highly correlated with *Number of pensions*. We keep *Number of pensions* and drop *Current pension type #1*.
Spouse birth year and *Spouse birth date* contain closely related age information. We replace them with *Spouse birth cohort*.
Spouse retired/employed is also highly correlated with spouse age measures. After replacing *Spouse birth year* and *Spouse birth date* with *Spouse birth cohort*, we drop *Spouse retired/employed*.
Birth year and *Birth cohort* capture similar life-cycle information. After dropping *Birth year*, we additionally include *Age in months* in the final retirement-variable set to retain a continuous measure of age.

The final retirement-variable set therefore contains:

Respondent birth cohort
Age in months
Receives Social Security
Employer-provided health plan covers retirees
Household capital income
Value of other debt
Hours worked (job 1)
Hours worked (job 2)
Health limits work
Retired
Number of jobs over employment history

Fig. D.2. Summary plot for the Shapley values for the delayed claiming case. *Notes:* This SHAP summary plot corresponds to the model predicting delayed Social Security claiming (i.e., claiming benefits after Full Retirement Age). The analysis is limited to individuals aged 62 to 70. The y-axis ranks variables by importance. The x-axis shows the Shapley value, where a positive value indicates a contribution towards a higher predicted probability of delayed claiming. The color of each point shows the variable's value (red for high, blue for low).

Number of pensions currently receiving
Pension income
Pension income (indicator)
Spouse birth cohort
Spouse retirement satisfaction

For the Social Security case, we make the following choices when two or more variables are highly correlated or convey very similar information:

Birth year is highly correlated with both *Age in months* and *Birth cohort*. We keep *Birth cohort* and *Age in months* and drop *Birth year*.
Covered by government health insurance and *Medicare* are highly correlated. We therefore keep *Medicare* and drop *Covered by government health insurance*.
Spouse birth date and *Spouse birth year* contain closely related age information. We replace them with *Spouse age in months*.
Chance of living to 75 is highly correlated with *Age in months*. We drop *Chance of living to 75*.
We keep both *Spouse age in months* and *Receives Social Security (spouse)* even though they have a correlation of 0.66, because they capture conceptually distinct information: spouse age and spouse claiming status.

The final Social Security-variable set therefore contains:

Age in months
Birth place
Medicare
Birth cohort
Employer-provided health plan covers retirees
Hours worked (job 1)
Receives Social Security (spouse)
Years worked
Covered by employer health insurance
Retired
Industry code_1980
Years of education
Earnings/ prev. wave
Spouse age in months

Appendix F. Contributions when only using selected variables

In this section, we report the Shapley values of the gradient boosted trees model using only the selected variables for each retirement and Social Security case. The vertical lines in Figs. F.2 and F.4 are the thresholds discussed in the main text. The overall patterns of the Shapley values are similar to those obtained when using all available variables. While the patterns for the Social Security case appear somewhat less clear, they remain in line with the key findings from the main model that uses all available variables.

Appendix G. Definition of categorical variables

This section describes the categorical variables used in the analysis. In the code, we generate dummy variables for the respondent's birth cohort, the spouse's birth cohort, spouse retirement satisfaction, birthplace, industry, and whether the respondent's employer-provided health insurance plan covers retirees. These categories are constructed from HRS variables and then converted into indicator variables for estimation.

The respondent birth cohort indicators are created from *racohbyr*. The spouse birth cohort indicators are created from *s14cohbyr*. Birthplace indicators are created from *rabplace*. Industry indicators are created from the harmonized 1980 Census industry code variable *ind_code_1980*. In addition, the code creates categorical indicators for spouse retirement satisfaction using *s14retsat* and for retiree health insurance coverage using *r14covrt*.

Respondent birth cohorts are summarized in Table G.8, spouse birth cohorts in Table G.9, spouse retirement satisfaction in Table G.10, birthplace in Table G.11, industry in Table G.12, and retiree health insurance coverage in Table G.13.

Appendix H. Social security benefits

In this section, we summarize the rules that determine the eligibility and the amount of Social Security retirement benefits.

You may be eligible for monthly Social Security benefits if you are: a retired insured worker age 62 or over, a disabled insured worker who has not reached full retirement age, a spouse of a retired or disabled worker entitled to benefits, a divorced spouse of a retired or disabled worker entitled to benefits, the dependent of a retired, disabled, or deceased insured worker. To receive benefits based on your earnings, you should be 62 or older and worked and paid Social Security taxes for 10 years or more. The detailed requirements for other categories are in the SSA handbook.[9] One example of a detailed category is for spousal benefits given to divorced spouses. To receive benefits based on your ex-spouse's earnings, you need to be married to the worker for at least 10 years.

The Social Security benefits are based on "average indexed monthly earnings" (AIME), which is the monthly average over 35 years of a

[9] https://www.ssa.gov/OP_Home/handbook/handbook.html

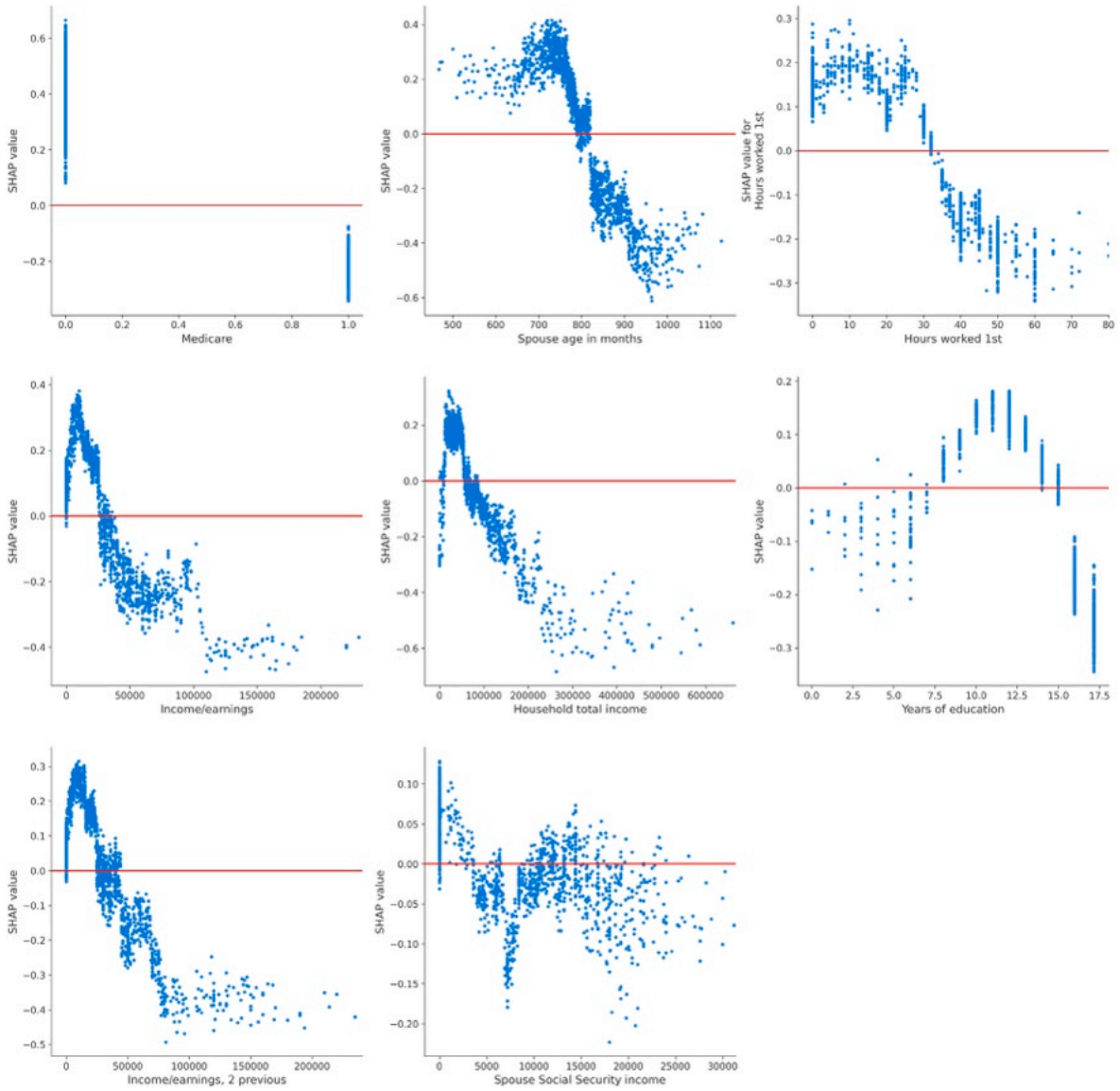

Fig. D.3. Dependence plot for the early claiming case.
Notes: This figure displays SHAP dependence plots for important variables from the model predicting early Social Security claiming. Each subplot shows a variable's value (x-axis) versus its Shapley value (y-axis). A positive Shapley value means the variable's value contributed to a higher predicted probability of early claiming. The horizontal line represents a Shapley value of zero.

worker's indexed earnings. The earnings are indexed to reflect the general rise in the standard of living over the years. The indexing is based on the national wage indexing series. Prior to the year that an individual reaches 60, the earnings are indexed to the year that the individual reaches 60. On and after the year that the individual reaches 60, the earnings are taken at face value. For example, if the individual is 60 in 2022, the earnings in 2021 are multiplied by the ratio of 63,795.13, which is the average wage index for 2022, and divided by 60,575.07, which is the average wage index for 2021.

AIME averages a maximum of 35 years of earnings. Among the years of earnings, SSA chooses those years with the highest indexed earnings and divides the total amount by the number of months in those years. This is the so-called AIME of the individual. The benefits that are paid to an individual are based on the primary insurance amount (PIA), which is a function of AIME. The portion that an individual can receive from the AIME is based on two "bend points". The "bend points" are based on the year the individual first becomes eligible. For example, if an individual first becomes eligible in 2024, the PIA is the sum of 90 percent of the first $1174 of his/her AIME and 32 percent of his/her AIME over $1174 and through $7078 and 15 percent of his/her AIME over $7078. There is a maximum for the monthly benefit and a cost-of-living adjustment over the years.

The actual amount of benefit may differ from PIA since the individual can retire early or later than the normal retirement age. The normal retirement age is increasing from 65 to 67 gradually and differs by birth cohort. For example, those born prior to or in 1937 are subject to the normal retirement age of 65 while those born after or in 1960 are subject to the normal retirement age of 67. If the individual chooses to retire early and receive benefits, the benefit is reduced by each month before the normal retirement age. For example, an individual can choose to retire at the age of 62 in 2024 and receive monthly benefits that are reduced by as much as 30%. On the other hand, the

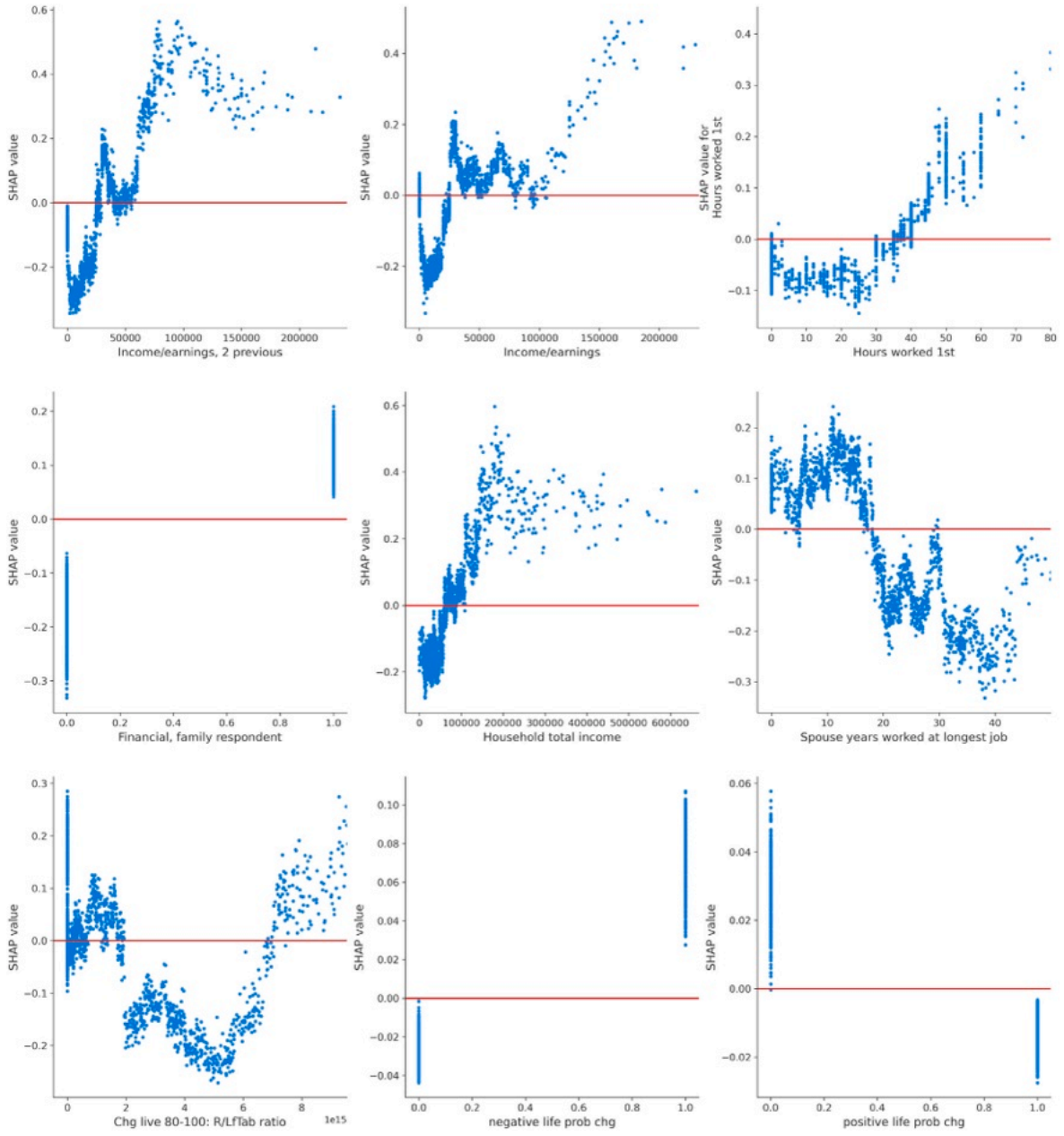

Fig. D.4. Dependence plot for the delayed claiming case.

Notes: This figure displays SHAP dependence plots for important variables from the model predicting delayed Social Security claiming. Each subplot shows a variable's value (x-axis) versus its Shapley value (y-axis). A positive Shapley value means the variable's value contributed to a higher predicted probability of delayed claiming. The horizontal line represents a Shapley value of zero.

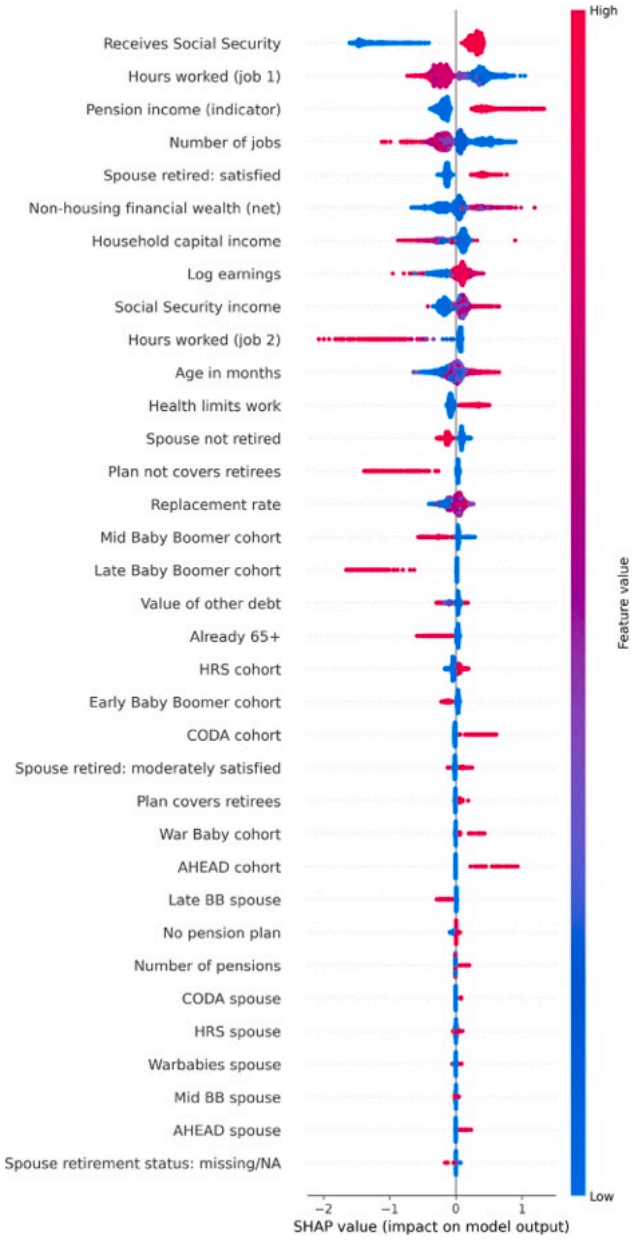

Fig. F.1. Summary plot for the Shapley values for the retirement case when only selected variables are used.
Notes: This plot corresponds to a version of the retirement prediction model that was trained using only the subset of important variables identified by our selection procedure. Fig. F.1 is the summary plot ranking variables by importance.

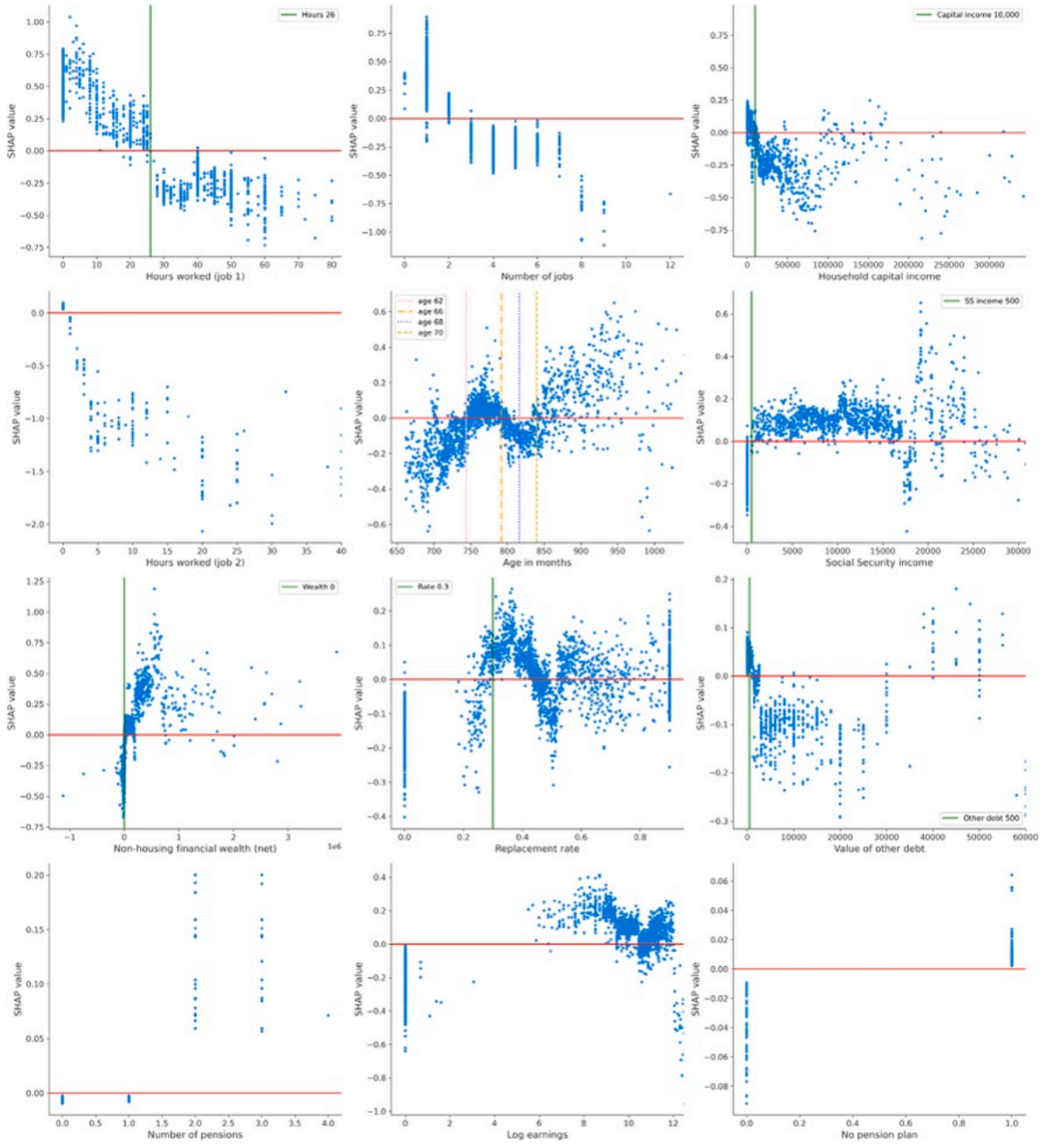

Fig. F.2. Dependence plot for the Shapley values for the retirement case when only selected variables are used.
Notes: This plot corresponds to a version of the retirement prediction model that was trained using only the subset of important variables identified by our selection procedure. Fig. F.2 shows the dependence plots, illustrating the marginal effect of each variable on the prediction. A positive Shapley value is associated with a higher predicted probability of retirement. The vertical lines represent the thresholds presented in the main results and are kept in the figure to show that the patterns of the Shapley values are similar.

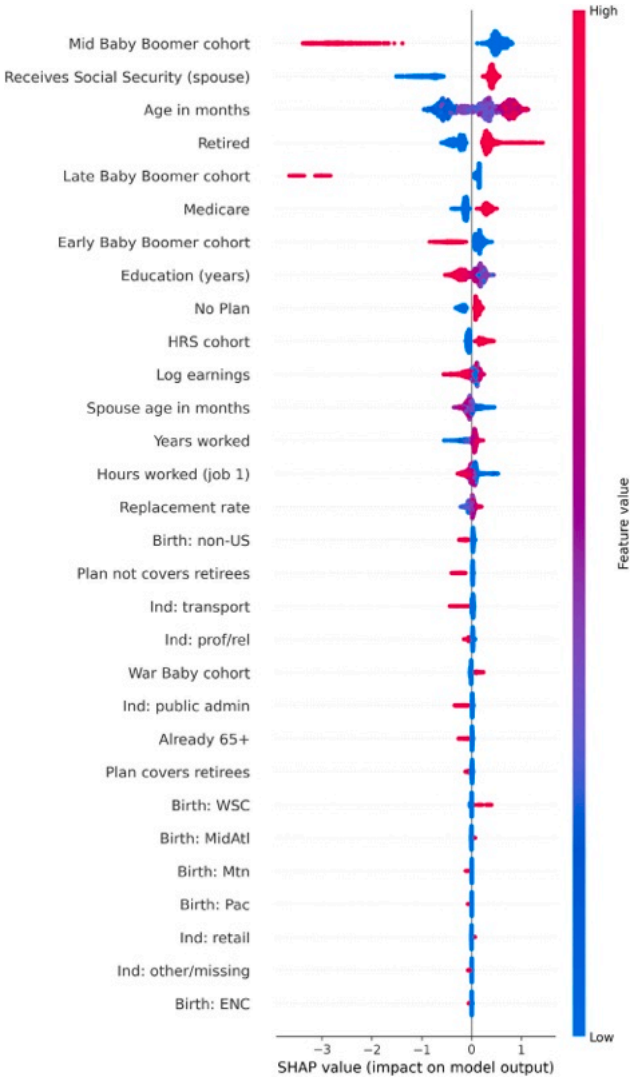

Fig. F.3. Summary plot of the Shapley values for the Social Security case using only selected variables.
Notes: This plot corresponds to a version of the Social Security prediction model that was trained using only the subset of important variables identified by our selection procedure. Fig. F.3 is the summary plot ranking variables by importance.

Table E.5
Variable importance lists for retirement: descending importance order.

Impurity base	Permutation base	Shapley values
Receives Social Security	Number of jobs	Birth year
Spouse unemployed	Weeks worked (job 2)	Receives Social Security
Spouse DC plan #3 mode	Household capital income	Hours worked (job 1)
Birth cohort	Hours worked (job 1)	Number of jobs
Birth year	Hours worked (job 2)	Birth cohort
Spouse DC contribution #3	Health limits work	Age when started to receive SS
Spouse labor-force history	Retired	Retired
Spouse DC balance #3	Self-employed	Spouse birth date
Spouse birth cohort	Value of other debt	Weeks worked (job 1)
Number of pensions	Weeks worked (job 1)	Health limits work
Spouse labor-force status	Spouse retirement satisfaction	Pension income (indicator)
Age when started to receive SS	Job-history count (alt.)	Household capital income
Spouse birth year	Total wealth excl. IRA	Number of pensions
Pension income	Household total income	Pension income
Spouse birth date	Employer-provided health plan covers retirees	Spouse retirement satisfaction
Spouse in labor force	Earnings/prev. wave	Pension/annuity income
Hours worked (job 2)	Receives Social Security	Self-employed
Weeks worked (job 2)	Spouse retirement year	Current pension type #1
Current pension type #1	Dividend income	Weeks worked (job 2)
Spouse retirement satisfaction	Spouse retired/employed	Household total income
Retired	Spouse last-worked year	Spouse birth year
Health limits work	Spouse occupation code	Age in months
Self-employed	Drinking frequency	Hours worked (job 2)
Spouse hours worked (job 1)	Spouse race	Value of other debt
Spouse retired/employed	Spouse father age	Employer-provided health plan covers retirees

Notes: This table reports the top 25 variables from the retirement notebook outputs, relabeled using the terminology in the working paper whenever available. When a direct working-paper label was not clearly available, the label follows the RAND HRS documentation or a cautious descriptive rendering of the variable name. Impurity base uses the model's built-in feature importance. Permutation base uses the mean decrease in predictive performance when the variable is permuted. Shapley values are ranked by mean absolute SHAP values.

Table E.6
Variable importance lists for Social Security: descending importance order.

Impurity base	Permutation base	Shapley values
Birth cohort	Birth year	Birth year
Birth year	Retired	Birth cohort
Spouse psychiatric problems	Birth cohort	Retired
Government health insurance premium	Receives Social Security (spouse)	Age in months
Retired	Hours worked (job 1)	Receives Social Security (spouse)
Spouse birth year	Government health insurance premium	Government health insurance premium
Receives Social Security (spouse)	Covered by employer health insurance	Covered by government health insurance
Age in months	Pension/annuity income	Spouse birth year
Covered by government health insurance	Household total income	Spouse birth date
Spouse birth date	Doctor visits	Covered by employer health insurance
Spouse birth cohort	Birth place	Hours worked (job 1)
HRS sample indicator	Value of other debt	Years at job
Spouse same job as previous wave	Industry code (1980)	Earnings (wave 12)
Spouse arthritis	Chance of living to 75 (relative)	Years of education
Spouse occupation code	Earnings/prev. wave	Earnings/prev. wave
Spouse gets help bathing	Weeks worked (job 1)	Spouse age when started receiving Social Security
Spouse industry code	Years of education	Spouse age in months
Spouse number of jobs	Chance of living to 75	Industry code (1980)
Probability of nursing home in 5 years	Spouse age in months	Birth place
Covered by employer health insurance	Years at job	Pension/annuity income
Employer-provided health plan covers retirees	Health-insurance count (r14henum)	Education degree
Spouse probability of nursing home in 5 years	Number of times widowed	Chance of living to 75
Spouse age in months	Spouse birth date	Ever had heart trouble
Spouse data source	Total non-housing wealth	MLEN measure (r14mlen)
Chance of living to 75	Age in months	Employer-provided health plan covers retirees

Notes: This table reports the top 25 variables from the Social Security notebook outputs, relabeled using the terminology in the working paper whenever available. When a direct working-paper label was not clearly available, the label follows the RAND HRS documentation or a cautious descriptive rendering of the variable name. Impurity base uses the model's built-in feature importance. Permutation base uses the mean decrease in predictive performance when the variable is permuted. Shapley values are ranked by mean absolute SHAP values.

Table E.7
Variables that enter at least two out of three importance lists.

Retirement case	Social Security case
Birth cohort	Birth cohort
Birth year	Birth year
Employer-provided health plan covers retirees	Employer-provided health plan covers retirees
Hours worked (job 1)	Hours worked (job 1)
Retired	Retired
Spouse birth date	Spouse birth date
Spouse birth year	Spouse birth year
Age when started to receive SS	Age in months
Current pension type #1	Birth place
Health limits work	Chance of living to 75
Household capital income	Covered by employer health insurance
Household total income	Covered by government health insurance
Number of jobs	Earnings/prev. wave
Number of pensions	Industry code_1980
Pension income	Medicare
Receives Social Security	Receives Social Security (spouse)
Spouse retired/employed	Spouse age in months
Spouse retirement satisfaction	Whether self-employed
Value of other debt	Years of education
Weeks worked (job 1)	Years worked
Weeks worked (job 2)	
Whether self-employed	
Hours worked (job 2)	

Notes: This table reports the raw union of the pairwise intersections from the three top-25 importance lists (impurity-based, permutation-based, and Shapley-based), using the corrected notebook outputs for the retirement and Social Security cases. A variable is included if it appears in at least two of the three lists. Labels follow the terminology used in the working paper whenever available.

Table G.8
Respondent birth cohort.

Dummy variable name	Birth years
AHEAD	before 1924
CODA	1924–1930
HRS	1931–1941
Warbabies	1942–1947
Early BB	1948–1953
Mid BB	1954–1959
Late BB	1960–1965

Table G.9
Spouse birth cohort.

Dummy variable name	Birth years
AHEAD spouse	before 1924
CODA spouse	1924–1930
HRS spouse	1931–1941
Warbabies spouse	1942–1947
Early BB spouse	1948–1953
Mid BB spouse	1954–1959
Late BB spouse	1960–1965
Spouse not in any cohort	not assigned or 1966 or later

Table G.10
Spouse retirement satisfaction.

Dummy variable label	Category
Spouse retired: satisfied	Spouse retired and satisfied
Spouse retired: moderately satisfied	Spouse retired and moderately satisfied
Spouse retired: not satisfied	Spouse retired and not satisfied
Spouse not retired	Spouse not retired
Spouse retirement status: missing/NA	Missing/not applicable

Table G.11
Birth place.

Dummy variable name	Birth place
bp_NE	New England
bp_MA	Mid Atlantic
bp_ENC	East North Central
bp_WNC	West North Central
bp_SA	South Atlantic
bp_ESC	East South Central
bp_WSC	West South Central
bp_Mtn	Mountain
bp_Pac	Pacific
bp_USNA	U.S./North America division not assigned
bp_nonUS	Not U.S. (including U.S. territories)

Table G.12
Industry categories based on 1980 Census code.

Dummy variable label	Industry category
Ind: other/missing	Other/missing
Ind: ag/for/fish	Agriculture, forestry, and fishing
Ind: mining/const	Mining and construction
Ind: mfg non-dur	Manufacturing: nondurable goods
Ind: mfg dur	Manufacturing: durable goods
Ind: transport	Transportation
Ind: wholesale	Wholesale trade
Ind: retail	Retail trade
Ind: FIRE	Finance, insurance, and real estate
Ind: bus/repair	Business and repair services
Ind: personal	Personal services
Ind: ent/rec	Entertainment and recreation
Ind: prof/rel	Professional and related services
Ind: public admin	Public administration

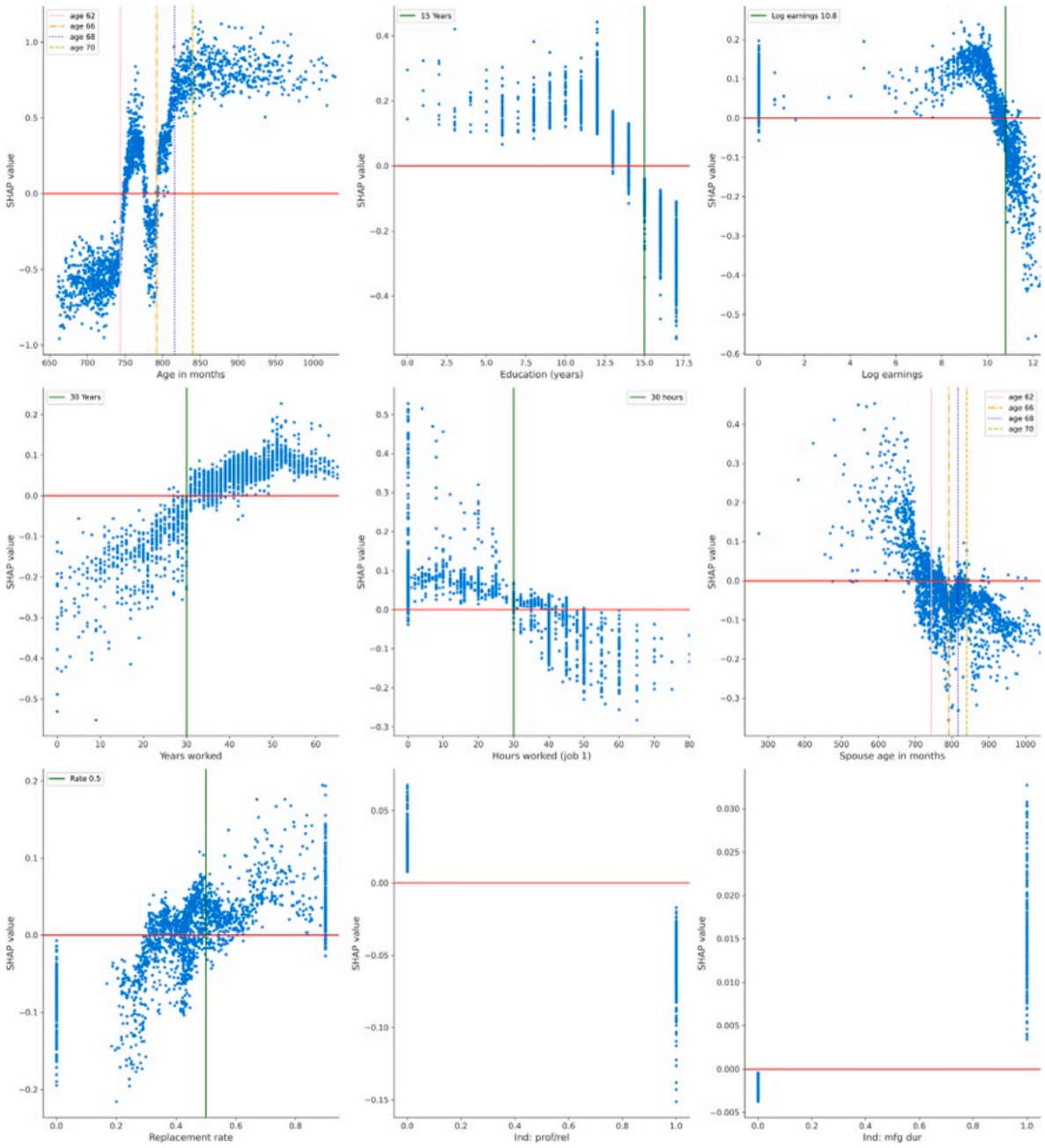

Fig. F.4. Dependence plot of the Shapley values for the Social Security case using only selected variables.
Notes: This plot corresponds to a version of the Social Security prediction model that was trained using only the subset of important variables identified by our selection procedure. The figure shows the dependence plots, illustrating the marginal effect of each variable on the prediction. A positive Shapley value contributes to a higher predicted probability of receiving Social Security. The vertical lines represent the thresholds presented in the main results and are kept in the figure to show that the patterns of the Shapley values are similar.

Table G.13
Employer-provided health insurance coverage for retirees.

Dummy variable name	Category
Plan covers retirees	Employer plan covers retirees
Plan not covers retirees	Employer plan does not cover retirees
No Plan	No employer-provided plan
Already 65+	Already age 65 or older

individual can choose to delay receiving Social Security over the normal retirement age and receive a credit that increases the monthly benefit by 8 percent per year for those born after 1942. There is no delayed credit given after the age 69.

Data availability

Data will be made available on request.

References

Albanesi, S., Vamossy, D.F., 2019. Predicting Consumer Default: A Deep Learning Approach. Working Paper 26165, National Bureau of Economic Research, http://dx.doi.org/10.3386/w26165, URL: http://www.nber.org/papers/w26165.

Ash, E., Galletta, S., Giommoni, T., 2025. A machine learning approach to analyze and support anticorruption policy. Am. Econ. J.: Econ. Policy 17 (2), 162–193. http://dx.doi.org/10.1257/pol.20210618, URL: https://www.aeaweb.org/articles?id=10.1257/pol.20210618.

Athey, S., 2017. Beyond prediction: Using big data for policy problems. Science 355 (6324), 483–485. http://dx.doi.org/10.1126/science.aal4321, URL: https://www.science.org/doi/abs/10.1126/science.aal4321, arXiv:https://www.science.org/doi/pdf/10.1126/science.aal4321.

Azinovic, M., Gaegauf, L., Scheidegger, S., 2022. Deep equilibrium nets. Internat. Econom. Rev. http://dx.doi.org/10.1111/iere.12575.

Barbaglia, L., Manzan, S., Tosetti, E., 2021. Forecasting loan default in europe with machine learning*. J. Financ. Econ. 21 (2), 569–596. http://dx.doi.org/10.1093/jjfinec/nbab010, arXiv:https://academic.oup.com/jfec/article-pdf/21/2/569/49676594/nbab010.pdf.

Blau, D.M., 1997. Social security and the labor supply of older married couples. Labour Econ. 4 (4), 373–418. http://dx.doi.org/10.1016/S0927-5371(97)00020-1, URL: https://www.sciencedirect.com/science/article/pii/S0927537197000201.

Blundell, R., French, E., Tetlow, G., 2016. Chapter 8 - Retirement incentives and labor supply. In: Piggott, J., Woodland, A. (Eds.), In: Handbook of the Economics of Population Aging, vol. 1, North-Holland, pp. 457–566. http://dx.doi.org/10.1016/bs.hespa.2016.10.001, URL: https://www.sciencedirect.com/science/article/pii/S2212007616300190.

Bluwstein, K., Buckmann, M., Joseph, A., Kapadia, S., Şimşek, Ö., 2023. Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. J. Int. Econ. 145, 103773. http://dx.doi.org/10.1016/j.jinteco.2023.103773, URL: https://www.sciencedirect.com/science/article/pii/S0022199623000594.

Butrica, B.A., Karamcheva, N.S., 2018. In debt and approaching retirement: Claim social security or work longer? AEA Pap. Proc. 108, 401–406. http://dx.doi.org/10.1257/pandp.20181116, URL: https://www.aeaweb.org/articles?id=10.1257/pandp.20181116.

Coile, C., Diamond, P., Gruber, J., Jousten, A., 2002. Delays in claiming social security benefits. J. Public Econ. 84 (3), 357–385. http://dx.doi.org/10.1016/S0047-2727(01)00129-3, URL: https://www.sciencedirect.com/science/article/pii/S0047272701001293.

Fouliard, J., Howell, M., Rey, H., 2020. Answering the Queen: Machine Learning and Financial Crises. Working Paper 28302, National Bureau of Economic Research, http://dx.doi.org/10.3386/w28302, URL: http://www.nber.org/papers/w28302.

French, E., 2005. The effects of health, wealth, and wages on labour supply and retirement behaviour. Rev. Econ. Stud. 72 (2), 395–427, URL: https://EconPapers.repec.org/RePEc:oup:restud:v:72:y:2005:i:2:p:395-427.

French, E., Jones, J.B., 2011. The effects of health insurance and self-insurance on retirement behavior. Econometrica 79 (3), 693–732.

Glickman, M.M., Hermes, S., 2015. Why retirees claim social security at 62 and how it affects their retirement income: Evidence from the health and retirement study. J. Retire. 2 (3), 25–39. http://dx.doi.org/10.3905/jor.2015.2.3.025, URL: https://jor.pm-research.com/content/2/3/25, arXiv:https://jor.pm-research.com/content/2/3/25.full.pdf.

Gustman, A.L., Steinmeier, T.L., 2005. The social security early entitlement age in a structural model of retirement and wealth. J. Public Econ. 89 (2), 441–463. http://dx.doi.org/10.1016/j.jpubeco.2004.03.007, URL: https://www.sciencedirect.com/science/article/pii/S0047272704000453.

Hamermesh, D.S., 2000. Togetherness: Spouses' Synchronous Leisure, and the Impact of Children. Working Paper 7455, National Bureau of Economic Research, http://dx.doi.org/10.3386/w7455, URL: http://www.nber.org/papers/w7455.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY, USA.

Haurin, D., Moulton, S., Loibl, C., 2022. The relationship of financial stress with the timing of the initial claim of U.S. Social Security retirement income. J. Econ. Ageing 21, 100362. http://dx.doi.org/10.1016/j.jeoa.2021.100362, URL: https://www.sciencedirect.com/science/article/pii/S2212828X21000554.

Huang, N., Li, J., Ross, A., 2022. Housing wealth shocks, home equity withdrawal, and the claiming of Social Security retirement benefits. Econ. Inq. 60 (2), 620–644. http://dx.doi.org/10.1111/ecin.13058, URL: https://ideas.repec.org/a/bla/ecinqu/v60y2022i2p620-644.html.

Hurd, M.D., 1990. The joint retirement decision of husbands and wives. In: Wise, D.A. (Ed.), Issues in the Economics of Aging. University of Chicago Press, Chicago, pp. 231–258.

Hurd, M.D., Smith, J.P., Zissimopoulos, J.M., 2004. The effects of subjective survival on retirement and Social Security claiming. J. Appl. Econometrics 19 (6), 761–775, URL: http://www.jstor.org/stable/25146321.

Kiefer, H., Mayock, T., 2025. Why do models that predict failure fail? J. Real Estate Financ. Econ. http://dx.doi.org/10.1007/s11146-025-10021-y, Advance online publication.

Kotlikoff, L.J., Summers, L.H., 1981. The role of intergenerational transfers in aggregate capital accumulation. J. Political Econ. 89 (4), 706–732, URL: http://www.jstor.org/stable/1833031.

Kumar, I.E., Venkatasubramanian, S., Scheidegger, C., Friedler, S., 2020. Problems with Shapley-value-based explanations as feature importance measures. In: III, H.D., Singh, A. (Eds.), Proceedings of the 37th International Conference on Machine Learning. In: Proceedings of Machine Learning Research, vol. 119, PMLR, pp. 5491–5500, URL: https://proceedings.mlr.press/v119/kumar20e.html.

Lee, S., Tan, K.T., 2023. Bequest motives and the social security notch. Rev. Econ. Dyn. 51, 888–914. http://dx.doi.org/10.1016/j.red.2023.09.001, URL: https://www.sciencedirect.com/science/article/pii/S1094202523000558.

Li, X., Hurd, M., Loughran, D., 2008. The characteristics of social security beneficiaries who claim benefits at the early entitlement age.

Lockwood, L.M., 2018. Incidental bequests and the choice to self-insure late-life risks. Am. Econ. Rev. 108 (9), 2513–2550. http://dx.doi.org/10.1257/aer.20141651, URL: https://www.aeaweb.org/articles?id=10.1257/aer.20141651.

Lucas, R.E., 1976. Econometric policy evaluation: A critique. Carnegie-Rochester Conf. Ser. Public Policy 1, 19–46. http://dx.doi.org/10.1016/S0167-2231(76)80003-6, URL: https://www.sciencedirect.com/science/article/pii/S0167223176800036.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.

Maestas, N., 2010. Back to work: Expectations and realizations of work after retirement. J. Hum. Resour. 45, 718–748. http://dx.doi.org/10.1353/jhr.2010.0011.

Modigliani, F., 1988. The role of intergenerational transfers and life cycle saving in the accumulation of wealth. J. Econ. Perspect. 2 (2), 15–40. http://dx.doi.org/10.1257/jep.2.2.15, URL: https://www.aeaweb.org/articles?id=10.1257/jep.2.2.15.

Molnar, C., 2022. Interpretable Machine Learning, second ed. URL: https://christophm.github.io/interpretable-ml-book.

Mukherjee, A., 2018. Time and money: Social security benefits and intergenerational transfers. AEA Pap. Proc. 108, 396–400. http://dx.doi.org/10.1257/pandp.20181115, URL: https://www.aeaweb.org/articles?id=10.1257/pandp.20181115.

Mukherjee, A., 2022. Intergenerational altruism and retirement transfers. J. Hum. Resour. 57 (5), 1466–1497. http://dx.doi.org/10.3368/jhr.58.1.0419-10140R3, URL: https://jhr.uwpress.org/content/57/5/1466, arXiv:https://jhr.uwpress.org/content/57/5/1466.full.pdf.

OECD, 2025. OECD Employment Outlook 2025: Can We Get Through the Demographic Crunch? OECD Publishing, Paris, http://dx.doi.org/10.1787/194a947b-en.

Qin, L., Zhu, Y., Liu, S., Zhang, X., Zhao, Y., 2025. The Shapley value in data science: Advances in computation, extensions, and applications. Mathematics 13 (10), http://dx.doi.org/10.3390/math13101581, URL: https://www.mdpi.com/2227-7390/13/10/1581.

Rust, J., Phelan, C., 1997. How social security and medicare affect retirement behavior in a world of incomplete markets. Econometrica 65 (4), 781–831, URL: http://www.jstor.org/stable/2171940.

Shoven, J., Slavov, S., Wise, D., 2018. Understanding social security claiming decisions using survey evidence.. J. Financ. Plan. 31 (11), 35–47.

Smith, K.E., Favreault, M., 2019. Modeling Income in the Near Term 8 and 2014 Primer. Research Report, Urban Institute, Washington, DC, URL: https://www.urban.org/research/publication/modeling-income-near-term-8-and-2014-primer, Primer for the Social Security Administration's MINT 8 model.

Smith, K.E., Williams, A., Murdrazija, S., 2021. Modeling Income in the Near Term: Version 8. Final Report, Urban Institute, Washington, DC, URL: https://www.urban.org/research/publication/modeling-income-near-term-version-8, Contract report for SSA's Office of Research, Evaluation, and Statistics.

Sundararajan, M., Najmi, A., 2020. The many Shapley values for model explanation. In: Daumé, H., Singh, A. (Eds.), Proceedings of the 37th International Conference on Machine Learning. In: Proceedings of Machine Learning Research, vol. 119, PMLR, pp. 9269–9278, URL: https://proceedings.mlr.press/v119/sundararajan20b.html.

von Wachter, T., 2009. The Effect of Labor Market Trends on the Incentives to Incidence for Claiming Social Security Benefits Early. Working Paper RRC NB09-06, National Bureau of Economic Research, http://dx.doi.org/10.3386/w8229, URL: https://www.nber.org/sites/default/files/2020-08/orrc09-06.pdf.

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., and, S.W., 2025. Evaluating treatment prioritization rules via rank-weighted average treatment effects. J. Amer. Statist. Assoc. 120 (549), 38–51. http://dx.doi.org/10.1080/01621459.2024.2393466.